

Sonja Heintz\* and Willibald Ruch  
**Reply to Martin (2015):  
Why our conclusions hold**

DOI 10.1515/humor-2015-0132

**Abstract:** We (Heintz and Ruch 2015, 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. *Humor: International Journal of Humor Research* 28. 611–633) pointed to a lack of convergence between the conceptualization of humor styles and how they are measured with the HSQ (i. e. the Humor Styles Questionnaire and recommended adjusting the model of the humor styles or alternatively the HSQ. The reply (Martin 2015, On the challenges of measuring humor styles: Response to Heintz and Ruch. *Humor: International Journal of Humor Research* 28. 635–639) suggested that our study could be methodologically flawed, thereby limiting the conclusions that can be drawn. In the present reply, we discuss each of these criticisms and demonstrate that these are likely unfounded and do not influence our results and conclusions. Thus we still suggest that the gap between the conceptualization and the measurement of humor styles should be closed to improve the construct validity of the HSQ. This would allow interpreting the manifold findings with the HSQ and deriving hypotheses for future studies based on the humor style constructs.

**Keywords:** humor styles, humor styles questionnaire (HSQ), validity, conceptualization

We (Heintz and Ruch 2015) recently observed a discrepancy in the different construct stages of the HSQ (Humor Styles Questionnaire; Martin et al. 2003). Employing a multitrait-multimethod framework, we tested the convergence of three of these stages (i. e. definitions, construct descriptions, and HSQ scales as provided by Martin et al. 2003), which served as three indicators of the four humor styles. This resulted in a high convergence for the self-defeating humor style and a lower convergence and partial mismatches for the affiliative,

---

**\*Corresponding author: Sonja Heintz**, Department of Psychology, University of Zurich, Binzmuehlestr. 14/7, Zurich 8050, Switzerland, E-mail: s.heintz@psychologie.uzh.ch  
**Willibald Ruch**, Department of Psychology, University of Zurich, Binzmuehlestr. 14/7, Zurich 8050, Switzerland

self-enhancing, and aggressive humor styles. The discrimination among the humor styles was low only for affiliative and self-enhancing. Multiple regression analyses revealed that the HSQ represented several of its conceptual elements. To improve the construct validity of the HSQ, Heintz and Ruch (2015: 611) suggested “[...] that either the constructs and model of the humor styles need to be adjusted or newly developed, or the HSQ does.”

Martin (2015: 635) discusses the methodology of our study and the relevance of our results and concludes that “[...] their study has several weaknesses that limit the conclusions that can be drawn.” More specifically Martin (2015) points to four potential methodological flaws: (1) The label “multitrait-multimethod” (MTMM), (2) the psychometric properties of the definition and construct description statements, (3) the lack of quantifiers in these statements, and (4) the use of technical terms and explicit motivations and functions of humor in these statements. We refute these assertions and address them individually. We demonstrate (partly by providing additional analyses) that they are unlikely to limit the conclusions that can be drawn from our study.

(1) *Does the use of only one measurement approach justify referring to a “multitrait-multimethod” (MTMM) study?* We do think so for two reasons. First, the seminal article by Campbell and Fiske (1959), which introduced the MTMM methodology to the psychological literature, exemplified MTMM studies involving only one measurement approach (e. g. different raters, observations, contents, or subtests as different methods). Second, the term “method” could be easily substituted for something more suitable for the present study, but this would not change anything. It is justified to use the MTMM framework, as this approach allows one to study convergent and discriminant validity; that is, construct validity. Thus independent of the label of the methodology, the common denominator remains: Do the three indicators of the four traits converge or not?

(2) *Is the construct validity of the HSQ mainly impaired by the psychometric properties of the definition and construct description statements that we used?* We believe that neither the lower internal consistency nor the higher intercorrelations of the definitions and construct descriptions affect the findings we obtained in relation to discriminant and convergent validity. This is because mostly the (qualitative) patterns of intercorrelations are important rather than their size (which could be corrected for attenuation anyway), and lower reliability would only affect the latter. Along these lines, we also analyzed the statements individually to show where the overlap is and where no overlap with the HSQ scales exists.

Importantly we have to reject the view that we created “alternative measures” (Martin 2015: 637) of the HSQ (i. e. measures aimed at substituting the HSQ or

measures to be used in further studies). What we wanted to accomplish is to obtain the most valid indicators for the constructs measured by the HSQ by taking statements from the original article by Martin and colleagues (2003). Thus, these statements do not constitute externally derived criteria or invented items, but they are different indicators of humor styles using Martin et al.'s (2003) own words. To analyze these statements (and their aggregations) statistically, we reported their psychometric properties (like internal consistency, mean, and so on) and treated them like items and scales in the MTMM framework. This should by no means be mistaken as alternative (but insufficient) measures.

(3) *Are our results flawed because of the presence of quantifiers in the HSQ items and their absence in the definitions and construct descriptions?* Martin (2015) suggests that this might be a key for obtaining lower correlations between the conceptualization and the measurement of the humor styles. Although this argument is theoretically feasible, our data speak against it. The quantifiers did not influence the size of the correlations: Seven of the 23 construct descriptions entail quantifiers (e.g. tend to, likely, excessively), yet there is no significant difference in the size of the correlation with the four HSQ scales between these (mean  $r = 0.28$ ) and the other 16 statements (mean  $r = 0.22$ ;  $z = 0.92$ ,  $p = 0.36$ ). Thus the absence of quantifiers, which was due to our aim to stay as close as possible to the original source (which mostly did not include quantifiers), did not significantly lower the size of the obtained correlations.

(4) *Are the results (i.e. the partly low correlations) artificial, as they are caused by the use of technical terms in the definitions and construct descriptions and because participants are required to identify and analyze their motivations and functions of humor when performing their ratings?* Martin (2015) is certainly right in saying that good items of a scale should avoid technical jargon and not overstretch the skills of participants for self-analysis. However, it was not our aim to write perfect items but to capture the key elements of the constructs to be measured as close to the original source as possible. We admit that a strong presence of such factors might make answering more difficult (and hence reduce the correlations) – but would it also create qualitatively different correlational patterns? This is unlikely to be the case and thus cannot explain our findings in terms of discriminant and convergent validity. Importantly, no significant difference can be found for the size of the correlations with the four HSQ scales when comparing the 13 construct descriptions that incorporate functions, motivations, and technical terms (mean  $r = 0.22$ ) with the ten that contain only observable elements (mean  $r = 0.27$ ;  $z = 0.76$ ,  $p = 0.44$ ). Thus, directly asking participants about the functions and motivations and using technical terms (vs. observable aspects of humor styles only) are not relevant factors in explaining our findings.

Overall, we did not make “[...] use of measures of questionable reliability and validity to test the validity of the HSQ” (Martin 2015: 639). Instead, we directly derived two inherently valid indicators of the four humor styles from the construction article that should be underlying the HSQ. Differences in the wording or the psychometric properties (such as reliability) cannot explain the low convergence and mismatches between some of the definitions and construct descriptions of the humor styles and the HSQ scale they should theoretically belong to, as well as the low discriminant validity between the affiliative and self-enhancing humor style.

Besides these methodological issues, Martin (2015: 636–637) doubted the relevance of our results for the construct validity of the HSQ: “[...] if it turns out that a measure does not completely conform to its original conceptualization, this does not automatically invalidate it.” Surely, no perfect overlap is needed, and this does not “automatically invalidate” the HSQ altogether. Validity is an important criterion that any instrument needs to satisfy, and users of an instrument need to know whether what is intended to be measured is actually measured or not. Validity is not an all-or-nothing criterion, but there are various degrees of how construct valid an instrument is. If there is a low convergence between the conceptualization and the measurement of a construct or even mismatches (as we found for some HSQ scales in our study), then the construct validity of an instrument has to be impaired. Thus either the model or concept needs to be adjusted or the instrument does to close the gap between the conceptualization and the measurement of humor styles.

Finally, Martin (2015: 637) rightly points out that it is “[...] quite common in the history of personality psychology for our conceptualization of particular constructs to change over time as research advances our understanding of them.” While adjusting theories and concepts (or developing new ones) on the basis of empirical findings is an important part of scientific progress, it is still important to keep the description of the concepts and their measurement parallel. This has been done, for example, with various measures of the five-factor model of personality, including its flagship, the NEO-PI-R (Costa and McCrae 1992). One should add that it is highly desirable that – next to valid items and fitting construct descriptions – a theory about the traits exists. For example, there are biological explanations of general intelligence or extraversion. It is not clear where the HSQ stands in this respect, as the concepts were initially derived from a clustering of theories or statements by a variety of authors. This is by no means a particular flaw of the HSQ, but the various humor traits often do lack a theoretical underpinning. The HSQ mostly starts from theories, but it is not clear whether these theories can be used to derive hypotheses to be tested with the four scales of the HSQ. So we would not only strongly endorse keeping the item

pool and construct descriptions parallel but also to develop formal models and eventually a theory of the humor styles. This should be a goal for everyone, not only the investigators driving the HSQ research.

## References

- Campbell, Donald T. & Donald W. Fiske. 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin* 56. 81–105.
- Costa, Paul T. & Robert R. McCrae. 1992. *NEO Personality Inventory-Revised (NEO PI-R)*. Odessa, FL: Psychological Assessment Resources.
- Heintz, Sonja & Willibald Ruch. 2015. 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. *Humor: International Journal of Humor Research* 28. 611–633.
- Martin, Rod A. 2015. On the challenges of measuring humor styles: Response to Heintz and Ruch. *Humor: International Journal of Humor Research* 28. 635–639.
- 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.

## Bionotes

### Sonja Heintz

Sonja Heintz received her master's degree in Psychology at the Saarland University in 2012, and she has since worked in the Section of Personality and Assessment at the University of Zurich, Switzerland. She works as a research, administration, and teaching assistant and pursues her PhD. Her main research interests are humor behaviors, individual differences in humor and positive psychological variables, methodology, and measurement.

### Willibald Ruch

Willibald Ruch is a Full Professor of Psychology at the University of Zurich, Switzerland. His research interests are in the field of humor and laughter, cheerfulness, and smiling. In his doctoral dissertation at the University of Graz (Austria) in 1980, he developed a taxonomy of jokes and cartoons and studied their relation to personality. His more recent work, together with his research team at the University of Zurich, includes humor from a positive psychology perspective, the effectiveness of humor training programs and clown interventions, the ability to laugh at oneself, the fear of being laughed at (gelotophobia), and the measurement of humor.