

## Sources of variance in current sense of humor inventories: How much substance, how much method variance?

GABRIELE KÖHLER AND WILLIBALD RUCH

Heinrich-Heine-University of Düsseldorf, Germany  
Institute of General Psychology

### Authors address:

Gabriele Köhler and Willibald Ruch, Department of General Psychology, Heinrich-Heine-University of Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany.

Phone: 0049-211-311-3067; Fax: 0049-211-311-2856;  
e-mail: koehlerg@uni-duesseldorf.de and wruch@uni-duesseldorf.de

### **Abstract**

*The present study investigates the relationship of self-report inventories of "sense of humor" and behavioral measures of humor as well as their location in the Eysenckian PEN system. 110 male and female adults in the ages from 17 to 83 years answered the following inventories: SHRQ (Martin and Lefcourt 1984), SHQZ (Ziv 1981), SHQ-3 revised (Svebak 1993), CHS (Martin and Lefcourt 1983), MSHS (Thorson and Powell 1993), HIS (Bell, McGhee, and Duffey 1986), 3 WD-K (Ruch 1983), CPPT (Köhler and Ruch 1993), TDS (Murgatroyd, Rushton, Apter, and Ray 1978), STCI-T (Ruch, Freiss, and Köhler 1993), and EPQ-R (Eysenck, Eysenck, and Barrett 1985). Reliability of the humor scales is examined and convergent and discriminant validity of homologous scales of humor appreciation and humor creation is determined. Behavioral measures and self-report instruments yield only meager correlations. While humor appreciation and humor creation form distinct traits in the behavioral measures, they can not be validly discriminated in the self-reports. Factor analysis of self-report inventories yields that the sense of humor is composed of the two orthogonal dimensions of cheerfulness and seriousness. Extraversion is predictive of cheerfulness, low seriousness, and quantity of humor production. Psychoticism is associated with low seriousness, wit and quality of humor production. Finally, emotional stability correlates with cheerfulness. All in all, the general state of the art in the assessment of the sense of humor and its components appears to be far from being satisfactory.*

## Introduction

In recent years, there has been an increasing interest in the personality construct of "sense of humor". As a consequence, a deluge of "sense of humor"-inventories has been developed, mostly relating to habitual forms of the comprehension, enjoyment, creation, and initiation of humor or of the ability to entertain others. When administering such instruments to a sample of individuals, one implicitly assumes that differences in the observed test scores only reflect individual differences in the trait the respective scale — according to its label — is promising to measure. However, this would be the case only if psychological instruments were perfectly reliable and valid.

## Sources of variance in humor instruments

Campbell and Fiske (1959) have argued that principally the total variance of an instrument is composed of three parts: content variance, error variance, and method variance. Content variance reflects interindividual differences in the trait to be measured and should be very high; ideally, it should be the only source of variance. Error variance (that is, lack of reliability) and method variance (that is, variance produced by the methodological approach chosen to assess the trait) should be low, ideally zero. Thus, regarding the latter, it is assumed that different methods (such as self-rating, peer-rating, objective test, or questionnaire) assessing the same trait will yield different results due to the fact that there are specific components of variance associated with the different methods. These may be, for example, different format of the scales (ratings vs. yes/no answer format), different evaluation perspective (self vs. peer), answer distortions (for example, social desirability) in self- and peer-reports, and peculiarities (for example, familiarity) of the material used in the objective tests. If no method variance is present, the correlation between, for example, self- and peer-reports of humor creation would approach unity (only restricted by the lack of reliability). If there is method variance associated with one of the methods (or both of them), the size of the correlation would be diminished to the extent of how much method variance is inherent. Generally speaking, a high convergent validity (correlation between different instruments measuring the same trait) is required to assure that the instruments are relatively free of method variance. Likewise, conceptually unrelated traits may correlate if they are measured by the same method, again due to method variance. Hence, it is necessary to demonstrate the discriminant validity of instruments as well.

Campbell and Fiske (1959) proposed the multitrait-multimethod matrix (MTMM) analysis as a methodology for estimating the presence and amount of method variance. A MTMM analysis requires different traits (for example, humor creation, humor appreciation), each measured by different methods (for example, questionnaire, self-report, behavioral measures). Subsequently, all scales are intercorrelated and the pattern of correlation is studied. Convergent validity is investigated by the convergence of multiple indicators of a specific construct, while discriminant validity requires the demonstration of the indicators' independence for different constructs.

## Multitrait-multimethod matrix (MTMM) analysis of sense of humor

In the only real MTMM study in the domain of sense of humor, Koppel and Sechrest intended to "determine the degree to which appreciation of humor, humor creation, intelligence and introversion-extraversion can be distinguished as traits" (1970: 78). They assessed these four traits by three methods (self-rating, peer-rating, and objective measures) each. Definitions of the traits were provided for self- and peer-evaluation (5-point scales) and the aggregated funniness ratings of 10 cartoons and (peer-rated) funniness of the captions produced for 10 captionless

cartoons served as objective measures for humor appreciation and humor creation, respectively. The six coefficients representing the convergent validities for the two humor measures ranged from .20 to .62, with four of them being significant (self-reports of humor appreciation did not correlate with peer-ratings and the objective measure of humor appreciation). Further evidence for the presence of method variance can be estimated from the fact that the two humor traits correlated differently for the three methods: .30, .50, and .86 for objective measure, self- and peer-ratings, respectively. Thus, peers did not distinguish between humor creation and humor appreciation, while these two abilities were relatively independent in performance. The study by Babad (1974) provides some additional hints for convergent validity without being a MTMM study in its proper sense. Peers and Ss themselves agreed in their nomination of which individuals in the total sample of 77 students were appreciators and producers of humor.

The generalizability of the Koppel and Sechrest study is limited by the fact that no established measures of humor appreciation and creation were studied; instead, ad hoc measures were created for this study. Unfortunately, no MTMM study has been undertaken to date with any of the current "sense of humor"-inventories. Therefore, we do not know how much these questionnaires and performance tests are affected by method variance.

### Questionnaire and behavioral measures of humor appreciation and humor creation

Taking current sense of humor instruments as a criterion, humor appreciation and humor creation are still among the most prominent elements in the domain of both questionnaires and self-reports. It was argued that for a comprehensive assessment of humor appreciation via behavioral measures (or performance tests), representative samples of both humor stimuli and humor responses need to be considered (Ruch 1992). Factor analytic studies yielded such a two mode taxonomy of humor appreciation which, on the part of the humor stimuli, distinguishes three broad humor categories (incongruity-resolution, nonsense, and sexual humor), while in the response mode humor appreciation is defined by two nearly orthogonal components of positive (that is, funniness) and negative (that is, aversiveness) responses (Ruch 1992)<sup>3</sup>. The 3 WD (Witz-Dimensionen) humor test (Ruch 1983) was developed for the reliable and valid assessment of this model of humor appreciation. As regards performance tests, humor creation behavior typically was assessed by means of caption removed cartoons (Babad 1974; Brodzinsky and Rubien 1976; Clabby 1980; Koppel and Sechrest 1970) or still pictures (Derks 1987; Derks and Hervas 1988; Nevo 1984; Nevo, Aharonson, and Klingman 1993; for a review, see O'Quin and Derks in press). Subjects were asked to produce funny captions. Their number and/or peer-rated funniness served as an index of humor creation ability. No such widely used and standardized instrument of humor creation ability exists and therefore it was necessary to design the Cartoon Punch line Production Test (CPPT; Köhler and Ruch 1993) for use in the present study.

As regards questionnaire measures of humor appreciation and creation we have to distinguish three classes of instruments, namely (1) broader scales of the sense of humor explicitly incorporating the respective element as a definitional component (but not having it as a separate subscale), (2) the set of scales or subscales measuring the respective construct (usually the scale is labeled identical to the trait, that is, is homologous), and (3) scales of concepts sufficiently similar to the construct of interest.

The Situational Humor Response Questionnaire (SHRQ; Martin and Lefcourt 1984) and the Multidimensional Sense of Humor Scale (MSHS; Thorson and Powell 1993) can be subsumed under the first category; they cover both aspects of humor appreciation and creation<sup>4</sup>. As regards the second category, the SHQZ (Ziv 1981) and the MSHS

---

<sup>3</sup> Thus, humor appreciation is maximal when funniness is high and aversiveness is low, while humor appreciation is minimal when funniness is low and aversiveness is high.

<sup>4</sup> One definitional component mentioned by Martin and Lefcourt is that the SHRQ "... may also be used to measure the sense of humor in terms of the productive definition. Individuals who report smiling and laughter in situations that are not obviously or necessarily humor-arousing might tend actively to produce humor rather than to respond to it passively.

contain respective homologous scales or factors for both humor appreciation and humor creation; and the Liking of Humor scale of the SHQ-3 (Svebak 1993) may be counted as a humor appreciation scale as well. For humor creation, there are two instruments falling into the third category, namely the HIS (Humor Initiation Scale; Bell, McGhee, and Duffey 1986) and the Metamessage Sensitivity scale of the SHQ-3 (Svebak 1993). With this arrangement taken, one can try to check the convergent validity of instruments measuring two traits (humor appreciation; humor creation) via two methods (questionnaire, behavioral measures).

While the present study attempts to address this question, a MTMM analysis in its proper sense will not be employed for the following four reasons. (1) While there are humor measures of humor appreciation and creation based on at least two different methods, they do not explicitly refer to the identical construct. In other words, while some may share the same label (that is, are homologous scales), they do not share an identical construct definition. (2) Whereas gaps could be overcome by constructing instruments for the missing trait-method combinations, the prime aim of the present study is to get information about the existing and currently used instruments, rather than studying the traits per se. Any such newly constructed instruments would contribute to the already existing vast and confusing mass of humor instruments. (3) At least in the realm of questionnaire measures of humor appreciation and creation, there is not one — but several — instruments or subscales available. Instead of arbitrarily choosing one, we prefer to assess all of them, because their intercorrelations provide information akin to convergent validity, too. Although they share the same method, their correlations might be low because of differences in the item content. Thus, the evaluation of convergent validity will comprise two sorts of intercorrelations: monomethod-monotrait correlations (that is, the several questionnaire versions of humor appreciation and humor creation) and heteromethod-monotrait correlations (that is, self-report and behavioral measures of humor appreciation and humor creation). While the former allows an estimation of the convergence of the different operationalisations using the same methodological approach, only the latter allows to estimate the proportion of method variance involved. (4) A prerequisite for a proper MTMM analysis is the existence of explicit theoretical constructs with clear reference to behavioral and experiential indicators for their assessment. Unfortunately, in humor research — with rare exceptions — we are still in want of such a theoretical framework.

## **Dimensions underlying the sense of humor**

The mathematical/statistical tool of factor analysis is a further methodology for the examination of construct validity. It allows to determine the number and nature of dimensions involved in the sense of humor. Furthermore, it provides information about which of the instruments are the best markers of the different components of humor.

The attempt to investigate the number and nature of the dimensions involved in the variance of some "sense of humor"-inventories was undertaken by Ruch (1994a). This factor analytic study of five humor inventories (CHS, SHQ [precursor of the SHQ-3, Svebak 1974], SHQZ, SHRQ, and TDS) with 10 scales yielded two factors of "surgency" and "restraint vs. expressive" (p. 231). Surgency is a component of the Eysenckian superfactor of Extraversion, and this term was chosen to refer to the fact that items of sense of humor often relate to definitional components of Extraversion, such as susceptibility for positive affect, smiling and laughter, enjoyment of entertaining others, carefreeness, and (a low degree of) seriousness. "Cheerfulness" was suggested to be an alternative label, which does represent the context of humor more appropriately. The "restraint vs. expressive"-factor was loaded positively by the three TDS-scales and negatively by the SHRQ and Emotional Expressiveness (SHQ EE). The suggested alternative label for this factor was "seriousness". Thus, the hypothesis put forward was that only two

---

... it is likely, that such individuals have developed the sort of mental facility for playing with ideas and perceptions in novel ways that is necessary for humor production ..." (Lefcourt and Martin 1986: 22).

dimensions underlie the current sense of humor questionnaires and individual instruments differ in their composition of "cheerfulness" and (lack of) "seriousness".

The present study will investigate the validity of this two-dimensional framework in three different ways: It will examine (1) whether these two factors can be replicated, (2) whether inventories that were subsequently constructed (MSHS; Thorson and Powell 1993), revised (SHQ-3; Svebak 1993) or not considered in the prior study (HIS; Bell et al. 1986) neatly fit into that space rather than requiring additional dimensions, and (3) whether the hypotheses regarding the nature of the factors can be substantiated by using measures of cheerfulness and seriousness.

## **Sense of humor and personality**

It has long been acknowledged that the "sense of humor"-construct is a node in a net of personality traits, not an isolated phenomenon. Thus, the study of the relationships between different conceptualizations of sense of humor and personality has a long history. In the present study only the domain of temperament will be considered. More precisely, as in the prior study (Ruch 1994a), the Eysenckian PEN model, comprising the superfactors of Psychoticism, Extraversion, and Neuroticism, will serve as a model for locating the various aspects of the sense of humor (for a description of the PEN model see Eysenck and Eysenck 1985; or Ruch 1994a).

### Self-report dimensions and the PEN system of temperament

Recently, Ruch (1994a) argued that certain facets of the sense of humor can be located in the realm of temperament. His inspection of the items of self-report sense of humor inventories showed that they often deal with facets of the superfactor of Extraversion, such as susceptibility for positive affect, smiling and laughter (for example, SHRQ, SHQZ-HA, SHQ EE), enjoyment of entertaining others (for example, SHQZ-HC), and carefreeness (for example, CHS). Additionally, he hypothesized that the ability to not lose one's sense of humor in trying situations (as measured by the CHS) might also relate to the disposition of emotional stability, that is, low Neuroticism. Moreover he supposed that a low sense of humor (for example, a more serious frame of mind) might be related to Introversion, but also to (low) Psychoticism.

To confirm these hypotheses, correlations were computed between the two factor scores (that is, the dimensions underlying the analyzed sense of humor inventories; see above) and the EPQ-R. Results clearly showed the surgency factor to be related to Extraversion only, while the restraint vs. expressive factor correlated mainly with (low) Psychoticism, but also with Introversion. Thus, Ruch (1994a: 232) concluded that "... most of the reliable variance of the humor inventories studied could be accounted for by the two general temperament dimensions of Extraversion and Psychoticism."

### Humor appreciation behavior and the PEN system of temperament

Several previous studies have been conducted to evaluate the relationship between humor appreciation and Extraversion (see, for example, Ruch 1992). Summarizing the findings for the 3 WD humor test, all computations of correlations consistently yielded positive coefficients, albeit only some of them were significant. The assumption which all those studies were based on — that extraverts respond to humor with more positive affect than do introverts — can not be rejected on the basis of the current evidence.

Since Neuroticism is known to be a predictor of negative affectivity, a positive relationship between N and the rejection of jokes and cartoons (that is, aversiveness) was hypothesized. The respective data analyses (see Ruch 1992) yielded that nearly all correlation coefficients were positive, but only one fourth of them were statistically significant. Furthermore, Introversion tended also to be positively related to aversiveness.

In these studies, however, no total scores for funniness and aversiveness were employed. Since E and N can be expected to be related to general funniness and aversiveness, respectively, aggregation across the three humor categories seems to be necessary for a more appropriate testing of the hypotheses.<sup>5</sup>

## Humor creation behavior and the PEN system of temperament

Predictors of "wit" or "humor creativity" might be found in the domain of general ability or creativity (Feingold and Mazzella 1991; O'Quin and Derks in press). However, there are also grounds to assume that Extraversion and Psychoticism are associated with quantity and quality of humor creativity, respectively. Extraversion is associated with the fluency component of creativity (Eysenck 1995), and might thus be hypothesized to predict the quantity of humor production. In support of this, Koppel and Sechrest (1970) found a positive relationship between humor creativity and extraversion. Furthermore, factor analyses of trait-descriptive adjectives often yielded that "witty" loads on the Extraversion factor. The mere reproductive entertainment aspect has been shown to be an element of Extraversion, too (Ruch 1994a).

Eysenck recently suggested that "... psychoticism is closely related to creativity, that underlying both is ... a tendency to have a flat associative gradient which allows the individual a wider interpretation of 'relevance' as far as responses to stimuli are concerned" (Eysenck 1995: 248). One might speculate that this richer source of associations allows the high P-scorer to write more uncommon, diverse, and incongruous, and hence wittier and more original punch lines; that is the high P-scorer have more "wit". This hypothesis gets direct support from the fact that P is higher among professional British cartoonists (Pearson 1983), and is compatible with the prior findings relating to the predictors of humor creativity among preschoolers (McGhee 1980) and the comics' familial environment during upbringing (Fisher and Fisher 1983).

The aim of the present study, then, is fourfold. (1) Psychometric properties of all commonly used inventories (especially of the German adaptations of the 'new' ones) will be evaluated drawing special attention to the proportion of error variance. (2) Convergent and discriminant validities will be examined for the homologous dimensions relative to humor appreciation and humor creation. (3) The number and nature of the dimensions of sense of humor will be examined by factor-analyzing the self-report instruments and thus attempting to replicate the findings of Ruch (1994a). Finally, (4) the relationship of sense of humor and the PEN system will be evaluated by correlating the emerging factors (as representatives of the self-report measures) and the scales of the behavioral measures (3 WD-K and CPPT) with the four scales of the EPQ-R.

## **Methods**

### Subjects

---

<sup>5</sup> Zuckerman (1994: 220) suggested labels of "general sense of humor" and "general aversiveness".

110 German adults, 51 males and 57 females, from 17 to 83 years ( $M = 45.6$ ;  $SD = 15.8$  years) were recruited by newspaper announcements in the Düsseldorf area. Their personal status can be described as follows: 37.6% unmarried, 44% married, 5.5% widowed, and 12.8% separated or divorced. 58.7% were white collar workers, 18.6% public-servants, 9.3% laborers, and 13.3% specified another trade group.

## Inventories

### Self-report inventories

CHS. The Coping Humor Scale (Martin and Lefcourt 1983; translation by Ruch 1994a) assesses the degree to which individuals make use of humor in coping with the stressful events that they encounter in their lives. The seven items have to be answered with a 4-point Likert scale ("1 = strongly disagree" to "4 = strongly agree").

HIS. The Humor Initiation Scale (Bell et al. 1986) contains 6 items in a five-point Likert-type format measuring the frequency of humor initiation. HIS was defined as the total score for the six items. The scale was translated into German by the authors.

MSHS. The Multidimensional Sense of Humor Scale (Thorson and Powell 1993) contains 24 items in a 5-point Likert format ("0 = strongly disagree" to "4 = strongly agree") investigating the "sense of humor" in its four components of humor generation or creativity, uses of humor as a coping mechanism, appreciation of humor, and attitudes toward humor and humorous persons. While Thorson and Powell intended only the evaluation of one total score, the present authors will also distinguish the four factor analytically derived components (Thorson and Powell 1993) by computing four scores: (1) Humor Creativity (MSHS-HC, 11 items); (2) Coping with Humor (MSHS-CH, 7 items); (3) Humor Appreciation (MSHS-HA, 2 items); and (4) Attitudes toward Humor (MSHS-AH, 4 items) supplementarily to the total score (MSHS) of sense of humor. This inventory was translated into German for this study by three persons independently, while agreement of the best translation of each item was found after discussion.

SHQ-3. The revised Sense of Humor Questionnaire (Svebak 1993) contains 21 items in a 4-point Likert-type format assessing generalized individual differences in humor production and appreciation. There are three scales, Metamessage Sensitivity (SHQ-3 M; the ability to recognize humor in situations), personal Liking of Humor (SHQ-3 L; the enjoyment of humor and acceptance of the humorous role), and Emotional Expressiveness (SHQ-3 E; the tendency to freely express one's emotions). 14 of the 21 items overlap with the previous version of the SHQ (Svebak 1974; translated by Ruch 1994a). The 7 new items were translated by the authors.

SHQZ. The Sense of Humor Questionnaire (Ziv 1981; translation by Ruch 1994a) contains 14 items in a 7-point Likert format assessing two components of the sense of humor: Humor Appreciation (SHQZ-HA) and Humor Creativity (SHQZ-HC). These two scores are combined to form a total Sense of Humor score (SHQZ-tot).

SHRQ. The Situational Humor Response Questionnaire (Martin and Lefcourt 1984) is a 21 item assessment of the individual's capacity to respond to a variety of situations (18 items) with amusement, smiling or laughter even if they are unexpected or demanding. Furthermore, three items are included which relate to self perception of humor.

STCI-T. The State-Trait-Cheerfulness Inventory (Ruch, Freiss, and Köhler 1993) is a 122-item questionnaire (pilot form) in a 4-point answer format and assesses the temperamental basis of the sense of humor. It provides scores for the three domains of Cheerfulness (STCI-T CH; 38 items scored), Seriousness (STCI-T SE; 37 items scored), and

Bad Mood (STCI-T BM; 31 items scored) and the 5, 6, and 5 definitional components, respectively, are scored as subscales. The 106 items scoring key was applied (see Ruch 1994b).

TDS. The Telic Dominance Scale (Murgatroyd, Rushton, Apter, and Ray 1978; translation by Ruch 1994a) is a 42-item questionnaire in a 3-point answer format measuring the three components of telic dominance: Seriousmindedness (TDS-SM), Planning Orientation (TDS-PO), and Arousal Avoidance (TDS-AA). Scores for each subscale are summed to give a total score of Telic Dominance (TDS-tot).

EPQ-R. The German version of the Eysenck Personality Questionnaire - Revised (Eysenck, Eysenck, and Barrett 1985) is a 102-item questionnaire in a yes-no answer format investigating the PEN personality system, namely Psychoticism (P; 32 items), Extraversion (E; 23 items), and Neuroticism (N; 25 items), and an additional Lie scale (L; 22 items).

## Behavioral measures

3 WD-K. The 3 WD ("3 Witz-Dimensionen") humor test (Ruch 1983) was designed to assess appreciation of jokes and cartoons of the three humor categories of incongruity-resolution humor, nonsense humor, and sexual humor. Form K contains 50 jokes and cartoons which are rated for "funniness" and "aversiveness" using two 7-point scales. Usually, six scores may be derived, three for funniness and three for aversiveness of incongruity-resolution (INC-RES), nonsense (NON), and sexual (SEX) humor. Additionally, two total scores of Funniness and Aversiveness were computed. Furthermore, a total Humor Appreciation index (3 WD-HA) was derived by subtracting total aversiveness from total funniness. Positive scores indicate high appreciation (Funniness high, Aversiveness low), a negative (or small positive) score indicates low appreciation (Funniness low, Aversiveness high).

CPPT. The Cartoon Punch line Production Test (Köhler and Ruch 1993) was developed to assess the individual's quantitative and qualitative humor creation abilities. It contains 15 caption-removed cartoons of the three humor categories INC-RES, NON and SEX (5 each), and Ss are asked to create as many funny punch lines as they are able to within a period of 30 minutes. The total number of punch lines created forms the CPPT NP score. To get some further information about the quality of the created punch lines, all 1650 written punch lines were rated independently by 12 students<sup>6</sup> along two dimensions on 9-point scales: (a) wittiness of the punch line (or — in case of more than one punch line per cartoon — of the punch line the rater considered the best) and (b) originality of the (best) punch line. The two scores (CPPT WP and OP) were derived by summing the ipsativized 6 ratings and dividing by the respective number of cartoons for each S individually (for example, if a S wrote punch lines for 7 of all 15 cartoons, the denominator was 7) to eliminate the frequency component. Additionally, two ratings dealing with the punch lines' "author" were assessed: (c) how marked is the wit of that person (9-point Likert scale; CPPT WI) and (d) how poor versus rich is this person's fantasy (9-point bipolar rating-scale; CPPT FA).

## Procedure

Testing took place at the University of Düsseldorf. In small groups of 1 to 5 persons (with different starting times) all subjects filled in the inventories in a fixed order. Halfway through the procedure, they were given a break of 15 minutes. They were paid DM 21 for their participation.

---

<sup>6</sup> Each student rated half the punch lines at their own pace and without time restrictions.



## Results

Means, standard deviations, and reliabilities (coefficient Alpha) of all instruments and their correlations with age and sex are presented in Table 1.

Table 1. Means, Standard Deviations and Reliability (Coefficient Alpha) of all instruments used and their correlations with age and sex

Scales	Mean	SD		Age	Sex <sup>a</sup>
<u>Behavioral measures</u>					
3 WD-K INC-RES <sub>f</sub>	50.05	19.76	.91	.33***	-.04
3 WD-K NON <sub>f</sub>	37.02	18.34	.87	-.11	-.03
3 WD-K SEX <sub>f</sub>	34.65	20.12	.90	.01	-.28**
3 WD-K Total Funniness	121.72	46.62	.94	.10	-.15
3 WD-K INC-RES <sub>a</sub>	9.72	11.93	.86	-.01	.11
3 WD-K NON <sub>a</sub>	16.72	16.31	.87	.13	.11
3 WD-K SEX <sub>a</sub>	30.25	23.97	.93	.18	.30**
3 WD-K Total Aversiveness	56.68	44.82	.94	.14	.23*
CPPT Number of punch lines	15.00	9.30	.94	-.18	.05
CPPT Wittiness of punch lines <sup>b</sup>	-.52	2.44	.63	-.41***	.18
CPPT Originality of punch lines <sup>b</sup>	-.62	2.46	.52	-.37***	.16
CPPT Ss' Wit <sup>b</sup>	-.58	4.24	.71	-.41***	.13
CPPT Ss' Fantasy <sup>b</sup>	-.75	4.50	.72	-.35***	.04
<u>Self-report inventories</u>					
CHS Coping Humor	20.53	3.84	.66	.15	.04
HIS Humor Initiation	16.25	5.87	.88	-.28**	-.11
MSHS Sense of Humor	61.31	12.89	.91	-.08	-.09
SHQ-3 Metamessage Sensitivity	19.75	2.56	.47	-.15	-.01
SHQ-3 Liking of Humor	18.91	3.07	.64	-.18	-.01
SHQ-3 Emotional Expressiveness	21.75	2.93	.62	-.18	.18
SHQZ Humor Appreciation	29.66	7.23	.72	-.07	.05
SHQZ Humor Creativity	31.15	7.49	.69	.05	-.01
SHQZ Total Sense of humor	60.81	13.24	.81	-.01	.02
SHRQ	57.85	10.14	.81	-.05	.02
STCI-T Cheerfulness	119.28	17.85	.96	-.13	.12
STCI-T Seriousness	95.91	16.27	.92	.46***	-.03
STCI-T Bad Mood	63.96	15.91	.95	.04	-.06
TDS Seriousmindedness	9.85	4.09	.47	.26**	-.15
TDS Planning Orientation	11.35	4.66	.54	.18	-.08
TDS Arousal Avoidance	11.82	5.36	.72	.29**	.07
TDS Total Telic Dominance	33.01	10.27	.74	.34***	-.06
EPQ-R Psychoticism	8.32	4.50	.76	-.50***	-.01
EPQ-R Extraversion	14.39	5.51	.87	-.22*	.05
EPQ-R Neuroticism	10.67	5.60	.86	-.04	.15
EPQ-R Lying	8.20	3.97	.77	.40***	-.02

Notes. <sup>a</sup> Male sex coded as 1, female sex as 2; <sup>b</sup> ipsativized and aggregated evaluation of 6 raters.

Abbreviations. 3 WD-K = 3 "Witz-Dimensionen" humor test, Form K, INC-RES = incongruity-resolution humor, NON = nonsense humor, SEX = sexual humor, <sub>f</sub> = funniness, <sub>a</sub> = aversiveness; CPPT = Cartoon Punch line

Production Test; CHS = Coping Humor Scale; HIS = Humor Initiation Scale; MSHS = Multidimensional Sense of Humor Scale; SHQ-3 = revised Sense of Humor Questionnaire by Svebak; SHQZ = Sense of Humor Questionnaire by Ziv; SHRQ = Situational Humor Response Questionnaire; STCI-T = State-Trait-Cheerfulness-Inventory; TDS = Telic Dominance Scale; EPQ-R = Eysenck Personality Questionnaire - Revised.

\* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

Table 1 shows that most of the scales yielded sufficiently high reliabilities. The Alpha coefficients ranged from .47 to .96 (median .86). As in the prior study (Ruch 1994a), TDS-SM yielded an unacceptably low coefficient (.47). The revision of Svebak's SHQ (SHQ-3) was partly successful; reliabilities did increase for SHQ-3 L (.64) and SHQ-3 E (.62) to a satisfactory level, but were unacceptable for SHQ-3 M (.47).

Scores on the following scales decreased with age: among the humor scales only HIS, among the behavioral measures all quality scores of the CPPT, and among the EPQ-R the scales P and E. Scores for INC-RES<sub>f</sub>, STCI-T Seriousness, all TDS scales and EPQ-L increased with age. Sex differences were rare; females showed lower funniness scores and higher aversiveness scores for sexual humor and for total Aversiveness than males.

### Convergent and discriminant validity: Monomethod correlations

#### Self-report scales

The self-report scales of humor appreciation (MSHS-HA, SHQZ-HA, SHQ-3 L) and humor creation (MSHS-HC, SHQZ-HC, HIS, SHQ-3 M) and the related broader scales (MSHS, SHRQ) were intercorrelated (see Table 2).

All coefficients of Table 2 are positive, and with one exception, they are all statistically significant. The comparison of the coefficients' size of homologous scales (monotrait correlations) yields that the humor appreciation scales (MSHS-HA, SHQZ-HA, SHQ-3 L, and their related scales MSHS and SHRQ) did not form a homogeneous cluster (coefficients between .16 and .55, median .34). A more consistent range could be observed for the humor creation scales (MSHS-HC, SHQZ-HC, HIS, SHQ-3 M, MSHS, and SHRQ): minimum .42, maximum .76, and median .56. Thus, while the humor creation scales (within this monomethod comparison) do measure a similar dimension, the humor appreciation scales seem to differ substantially with regard to their content; that is, they lack convergent validity.

As regards discriminant validity of the self-report scales, Table 2 shows that the coefficients between humor appreciation scales and humor creation scales (heterotrait correlations) range between .26 and .62 (median .40). Almost all humor appreciation scales correlated higher with scales measuring a different characteristic (humor creation) than with homologous scales. Hence, within the self-report scales, it is doubtful that humor appreciation and humor creation form two distinct traits.<sup>7</sup>

#### Behavioral measures

Next, the monomethod correlations were computed for the behavioral measures, that is, the 3 WD-K and the CPPT. The resulting coefficients are presented in Table 3.

---

<sup>7</sup> Likewise, while the two scales assessing the element of coping humor (CHS; MSHS-CH) do indeed correlate highly positively ( $r = .58$ ), there are again problems with discriminant validity when compared with humor appreciation and humor creation.

Table 2. Intercorrelations of the self-report scales of humor appreciation, humor creation, and related scales

	Humor appreciation scales			Humor creation scales			Related scales	
	MSHS-HA	SHQZ-HA	SHQ-3 L	MSHS-HC	SHQZ-HC	HIS	SHQ-3 M	MSHS
Humor appreciation scales								
SHQZ-HA	.38***							
SHQ-3 L	.16	.28**						
Humor creation scales								
MSHS-HC	.39***	.49***	.34***					
SHQZ-HC	.33***	.62***	.38***	.71***				
HIS	.40***	.56***	.41***	.76***	.67***			
Related scales								
SHQ-3 M	.26**	.52***	.28**	.56***	.49***	.50***		
MSHS	.58 <sup>a</sup>	.55***	.37***	.90 <sup>a</sup>	.74***	.74***	.60***	
SHRQ	.22*	.50***	.22*	.42***	.53***	.53***	.37***	.46***

Note. <sup>a</sup> Correlations of subscales with total score were not considered for test of significance. \*P<0.05, \*\*P<0.01, \*\*\*P<0.001.

Table 3. Intercorrelations of the behavioral measures of humor appreciation (3 WD-K) and humor creation (CPPT)

	3 WD-K total scales			CPPT			
	3 WD-HA	Funniness	Aversiveness	NP	WP	OP	WI
3 WD-K							
Funniness	.79 <sup>a</sup>						
Aversiveness	-.77 <sup>a</sup>	-.23*					
CPPT							
NP	.18	.18	-.09				
WP	.12	.06	-.13	.54***			
OP	.10	.10	-.06	.65***	.93***		
WI	.13	.07	-.15	.63***	.92***	.91***	
FA	.19	.17	-.13	.68***	.80***	.88***	.88***

Note. <sup>a</sup> Correlations of subscales with total score were not considered for test of significance. \*P<0.05, \*\*\*P<0.001.

Table 3 shows a clear pattern of coefficients. While there are significant relationships among the indices of humor appreciation and humor creation (monotrait), respectively, the indices are not correlated between the two tests (heterotrait). Hence, the monomethod comparison of the behavioral measures yields that the two tests do measure two distinct dimensions, while the different indices of each test are related.

### Convergent and discriminant validity: Heteromethod correlations

#### Humor appreciation

The correlations of the three total scores of the 3 WD-K (3 WD-HA, Funniness, and Aversiveness) and the corresponding self-report scales are presented in Table 4.

Table 4. Correlations between self-report and behavioral measures of humor appreciation

	3 WD-K total scales		
	3 WD-HA	Funniness	Aversiveness
Self-report scales			
Humor Appreciation (MSHS)	.21*	.12	-.21*
Humor Appreciation (SHQZ)	.31**	.26*	-.22*
Liking of Humor (SHQ-3)	.20*	.13	-.18
MSHS	.27**	.21*	-.20*
SHRQ	.13	.17	-.04

\* $P < 0.05$ , \*\* $P < 0.01$ .

Table 4 shows some evidence of convergent validity for self-report and behavioral measures of humor appreciation. In detail, 3 WD-HA correlated positively with all self-report scales except the SHRQ, while total Funniness showed significant relations only to SHQZ-HA and MSHS (total score). For Aversiveness, all coefficients are negative, but significantly so only for MSHS-HA, SHQZ-HA, and MSHS (total score).

However, the convergence between the self-report measures and performance is not very strong at all. Even the highest coefficient does explain only 9.6% of the variance of the 3 WD, indicating a considerable amount of inherent method variance. A multiple regression analysis with the self-report measures as predictors and the 3 WD scales as criteria yielded multiple correlation coefficients of .35 ( $P < .05$ ), .28 (n.s.), and .30 (n.s.; all  $df = 5, 104$ ) for 3 WD-HA, Funniness, and Aversiveness, respectively. Thus, even all questionnaires combined can not predict a substantial portion of the variance in humor appreciation behavior.

#### Humor creation

The correlations between the humor creation self-report scales and the CPPT scores were computed next (see Table 5).

Table 5. Correlation between self-report and performance tests of humor creation

	CPPT				
	NP	WP	OP	WI	FA
Self-report scales					
Humor Creativity (MSHS)	.23*	.08	.08	.09	.13
Humor Creativity (SHQZ)	.10	.04	.01	.01	.00
Humor Initiation (HIS)	.22*	.14	.11	.13	.14
Metamessage Sensitivity (SHQ-3)	.30**	.38***	.34***	.34***	.35***
MSHS	.24*	.17	.13	.15	.16
SHRQ	.23*	.12	.12	.14	.14

\*P<0.05, \*\*P<0.01, \*\*\*P<0.001.

As expected, Table 5 shows positive relationships only. Regarding fluency, that is the number of written punch lines (CPPT NP), the coefficients range from .10 for the SHQZ-HC scale (not significant) to .30 (P<.01) for the SHQ-3 M scale. The multiple regression analysis between the self-report measures as predictors and CPPT NP as criterion yielded a coefficient of .38 (P<.01;  $df=6, 103$ ). Thus, all self-report measures combined are able to explain only 14.4% of the variance of the CPPT fluency score.

Regarding the quality of created punch lines, the coefficients range from zero (SHQZ-HC) to .38 (P<.001; SHQ-3 M). They are significant only for the SHQ-3 M scale, and they are especially low for the scales labeled "humor creativity". The multiple regression analysis between the self-report measures as predictor and CPPT WP, OP, WI, and FA as criterion yielded coefficients of .43, .40, .40, and .42 (all P<.01;  $df=6, 103$ ), respectively. The proportion of explained variance comes to 18.5%, 16%, 16%, and 17.6%.

Generally, the examination of discriminant validity is useful only if convergent validity has been demonstrated. As regards discriminant validity of humor appreciation and creation scales, this requirement was given only for some of them. However, inspection of the respective correlations yielded that the self-report scales of humor appreciation were positively related to the performance measures of humor creation ( $r$ 's between .08 and .24), and that self-reported humor creation went along with humor appreciation ( $r$ 's between .13 and .31 for 3 WD-HA and Funniness, and between -.04 and -.24 for Aversiveness) of the behavioral measure; for both, the coefficients' size was very similar to the heterotrait-monomethod comparisons. Thus, self-reported humor appreciation and creation can also not be discriminated when regarding their relationship to heterologous behavioral measures.

### Dimensionality of the humor-related self-report inventories

A principal components analysis was performed for the 23 humor-related self-report scales, namely the nine sense of humor scales, the definitional components of Cheerfulness and Seriousness<sup>8</sup> of the STCI, and the three TDS-scales.

Whereas four Eigenvalues were greater than 1, only two of them were markedly different from the others (Eigenvalues 9.49, 3.37, 1.35, 1.23, .98, .85). Varimax-rotation of 2, 3, and 4 factors led to the conclusion that data are well represented by a 2-factor solution, accounting for 56% of the variance. Since the Eigenvalue of the first

<sup>8</sup> Bad Mood as the other component of humorlessness was not considered because this aspect is not explicitly an element of the current sense of humor inventories. Thus, its inclusion would only have produced a separate factor of Bad Mood loaded mainly by its five facets.

factor was much higher than the others, a general-factor-solution needed to be examined as well. The two unrotated factors and the two Varimax-rotated factors, and communalities are given in Table 6.

Table 6. The two factors (unrotated and Varimax-rotated) underlying the 23 humor-related self-report scales

	unrotated		Varimax rotated		h <sup>2</sup>
	Factor 1	Factor 2	Factor 1	Factor 2	
CHS	.59	.37	.70	.03	.49
HIS	.74	.12	.71	-.26	.57
MSHS	.79	.17	.77	-.24	.66
SHQ-3 M	.74	-.06	.62	-.42	.56
SHQ-3 L	.42	-.01	.36	-.21	.18
SHQ-3 E	.75	.14	.73	-.25	.59
SHQZ-HA	.77	.13	.74	-.26	.61
SHQZ-HC	.77	.35	.84	-.07	.71
SHRQ	.60	.29	.67	-.04	.44
STCI-T CH1	.79	.33	.85	-.10	.73
STCI-T CH2	.77	.29	.82	-.12	.69
STCI-T CH3	.74	.23	.75	-.17	.60
STCI-T CH4	.82	.17	.80	-.25	.70
STCI-T CH5	.82	.30	.86	-.14	.77
STCI-T SE1	-.51	.44	-.23	.63	.45
STCI-T SE2	-.36	.70	.03	.79	.63
STCI-T SE3	-.45	.68	-.06	.82	.67
STCI-T SE4	-.47	.73	-.05	.87	.76
STCI-T SE5	-.49	.61	-.13	.77	.61
STCI-T SE6	-.72	.39	-.44	.70	.68
TDS-SM	-.44	.36	-.21	.52	.32
TDS-PO	-.38	.45	-.11	.58	.35
TDS-AA	-.28	.18	-.16	.29	.11

Table 6 shows that all scales loaded on a first unrotated component, suggesting a general bipolar factor of low vs. high sense of humor. This factor accounted for 41.3% of the variance and combined all humor scales and the Cheerfulness facets with positive loadings, and the Seriousness facets and the TDS scales with negative loadings. However, a second non-instrument-specific factor emerged, accounting for 14.7% of the variance. It was loaded positively by all STCI-T SE facets, TDS-SM, TDS-PO, and some of the humor scales (CHS, SHQZ-HC, SHRQ, STCI-T CH1, and STCI-T CH5). The highest positive loadings could be found for the Seriousness facets 2 to 5 (STCI-T).

The Varimax rotation yielded the expected pattern. The first factor ("cheerfulness") accounted for 34.8% of the variance and was loaded positively by all sense of humor scales and all Cheerfulness facets of the STCI-T and negatively only by STCI-T SE6. Thus, according to these results, "cheerfulness" is a broad factor composed of elements, such as a prevalent cheerful mood (CH1), the tendency to smile or laugh and to be merry (CH2 – low threshold for smiling and laughter; SHQZ Humor Appreciation; SHQ-3 Emotional Expressiveness; SHRQ), coping humor/cheerful composedness (MSHS-CH<sup>9</sup>; CHS; CH3 – composed view of adverse life circumstances), initiating humor/liking to entertaining others (CH5 – cheerful interaction style; SHQ-Z Humor Creativity; HIS; MSHS-HC),

<sup>9</sup> The four subscores of the MSHS, although not included in the present factor analysis, were projected in the emerging factor space by correlating the Varimax-rotated factor scores with these four subscales.

liking of humor stimuli (CH4 – broad range of active elicitors of cheerfulness; SHQ-3 Liking of Humor; MSHS-HA), and a positive attitude about things being related to cheerfulness and playfulness (MSHS-AH; negatively SE6 – humorless attitude toward cheerful events).

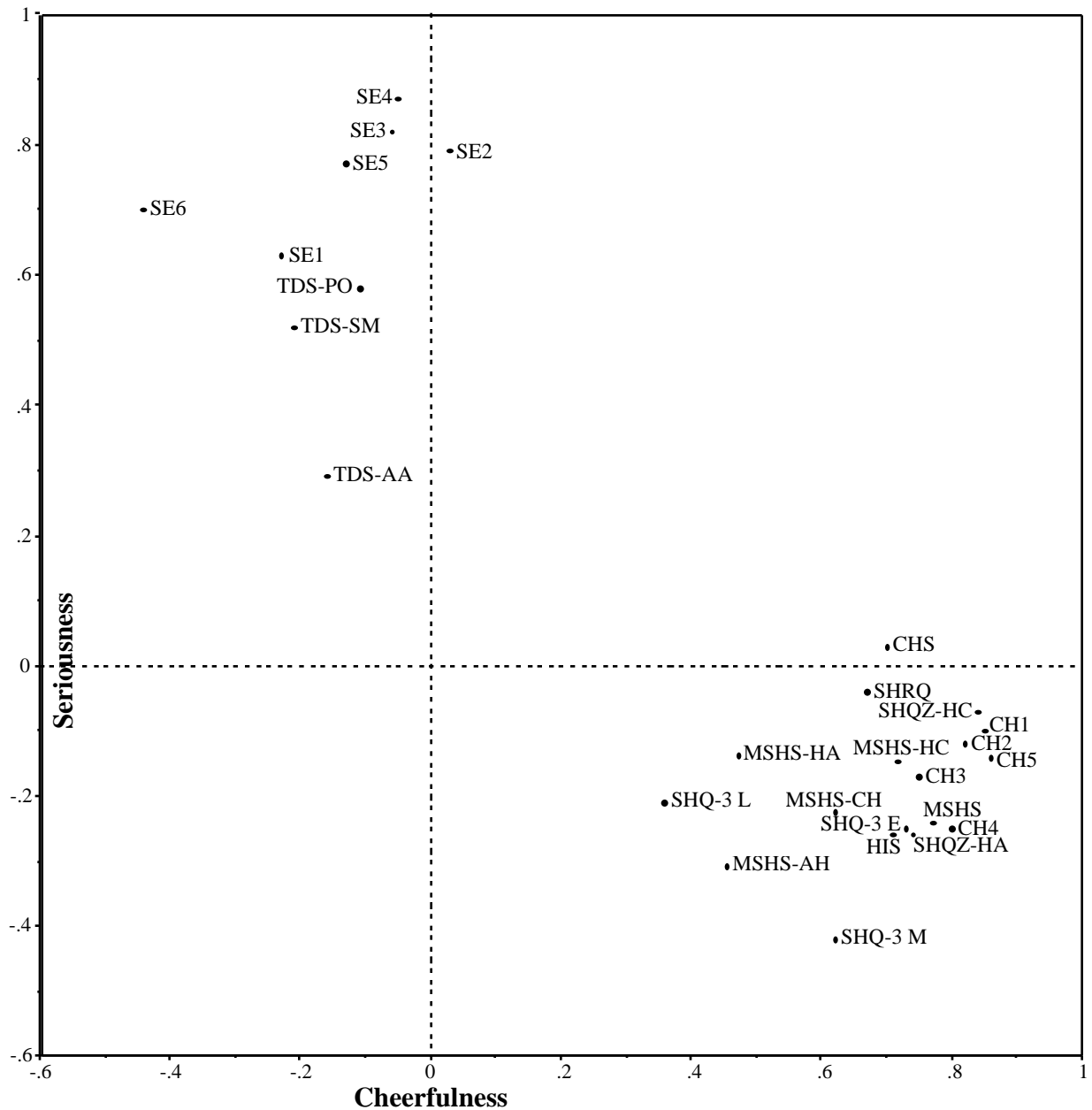


Figure 1. Location of the self-report scales in the factor space of cheerfulness and seriousness

The second factor ("seriousness") explained 21.1% of the variance and was loaded positively by all Seriousness facets of the STCI-T and the 3 TDS scales, and negatively only by SHQ-3 M. Thus, the present study broadens the contours of this factor by suggesting that seriousness emerges from the intercorrelations of several components, such

as the need for making sense rather than nonsense (SE4 – preferring activities for which concrete, rational reasons can be produced), taking even every-day matters as important (SE2), preferring a sober and object-oriented communication style (SE5), a humorless attitude toward cheerful events (SE6; MSHS-AH), planning far in advance (SE3 – tendency to set far reaching goals; TDS Planning Orientation), a generally serious frame of mind (SE1 – prevalence of serious states, TDS Seriousmindedness), and — to a lesser extent — the (self-reported) inability to recognize humor in situations (SHQ-3 Metamessage Sensitivity)<sup>10</sup>.

In order to allow for a closer inspection of the location of the scales in this two-dimensional space, the loadings of the two Varimax-rotated factors are presented in Figure 1.

Figure 1 shows that (with the exception of the CHS) all of the sense of humor scales are located in the cheerfulness/low seriousness quadrant. Thus, while they share a common loading on the cheerfulness-factor, they differ with respect to whether they are also loaded by (low) seriousness and how marked this loading is. The more affect-related scales (for example, CHS, SHRQ, SHQZ-HC, CH1, CH2, and CH5) are close to the axis and can be considered to be relatively pure markers of that factor. The sense of humor-scales involving mentality or attitudes (reflecting a lower degree of seriousness; SHQ-3 M, MSHS-AH) are loaded negatively by seriousness as well. Thus, seriousness is not only important for the location of the scales reflecting a low sense of humor; it also does account for differences among genuine sense of humor scales (the span of loadings being .45).

Nearly all communalities exceeded .50. However, not all of the variance of SHQ-3 L and the TDS scales, especially TDS-AA (which, however, essentially does not constitute a humor-related trait in the more narrow sense), was explained by the two common factors. The extraction of one or further factors, however, would simply introduce specific factors and thus would not improve this result.

### Relationship between sense of humor and the PEN system of temperament

The last question to be addressed is the one regarding the location of sense of humor in a descriptive model of personality. For this reason, the representatives of self-reported sense of humor (that is, the factor scores of both, unrotated and Varimax-rotated solutions) and the scales of the behavioral measures were correlated with the scales of the EPQ-R. The results are given in Table 7.

Table 7. Correlations between the EPQ-R and self-report dimensions of sense of humor, and behavioral measures of humor appreciation (3 WD-K), and humor creation (CPPT)

	EPQ-R			
	P	E	N	L
<u>Self-report dimensions</u>				
"Sense of humor" unrotated factor 2	.23*	.67***	-.35***	-.08
"cheerfulness"	.47***	-.11	-.01	-.38***
"seriousness"	-.03	.64***	-.31**	.11
"seriousness"	-.53***	-.23*	.18	.37***
<u>Humor appreciation</u>				
3 WD-K INC-RES <sub>f</sub>	-.23*	.12	-.06	.04
3 WD-K NON <sub>f</sub>	.11	.04	-.13	-.17

<sup>10</sup> The tendency to avoid arousal (TDS AA) is not listed as a marker of seriousness since its zero-order correlations with the seriousness-scales are low to non-significant.



3 WD-K SEX <sub>f</sub>	-.06	.17	-.01	-.08
3 WD-K Funniness	-.08	.14	-.08	-.08
3 WD-K INC-RES <sub>a</sub>	.02	-.16	.16	-.01
3 WD-K NON <sub>a</sub>	-.09	-.05	.09	.12
3 WD-K SEX <sub>a</sub>	-.00	-.15	.01	.16
3 WD-K Aversiveness	-.03	-.14	.08	.13
3 WD-HA	-.03	.18	-.11	-.13
<u>Humor creation</u>				
CPPT NP	.13	.25**	-.01	-.11
CPPT WP	.20*	.19	-.11	-.21*
CPPT OP	.21*	.18	-.10	-.19
CPPT WI	.26**	.16	-.09	-.25*
CPPT FA	.23*	.20*	-.09	-.25*

\*P<0.05, \*\*P<0.01, \*\*\*P<0.001.

Table 7 shows that the first unrotated factor ("sense of humor") correlated highly positively with E and to a lower extent also with P, and negatively with N. The multiple correlation between the three PEN variables and "sense of humor" is .71 (P<.001;  $df=3, 106$ ); that is, temperament and sense of humor share 50% of the observed (and much more of the reliable) variance. The second unrotated factor correlated positively with P, and negatively with L. The correlations of the Varimax-rotated factor scores show the expected pattern. The "cheerfulness"-factor correlated positively only with E (and quite highly so, underscoring that surgency is a component of Extraversion), and negatively with N, but not with P. The "seriousness"-factor showed a strong negative relationship with P and a positive with L. Furthermore, a weaker relationship with Introversion could be observed.

As expected, Table 7 shows no significant relationship between 3 WD-K humor appreciation and E and N. Again, Extraversion correlated consistently positively to all funniness scores (and negatively to all aversiveness scores) and Neuroticism correlated consistently positively with aversiveness and negatively with funniness; however, the coefficients lack practical and statistical significance. While the positive correlation between global humor appreciation (that is, the 3 WD-HA score) and Extraversion just failed to be significant (P=.06), funniness of incongruity-resolution humor correlated significantly with (low) Psychoticism.

As regards the CPPT, Table 7 provides support for the hypothesized relationships between humor creation and P and E. The quantity of humor production, that is, the total number of created punch lines (CPPT NP) correlated significantly positively only with E. Hence, the sociable, lively, and surgent individual produced more captions within a given time limit. More importantly, the high P-scorer turned out to be the person of wit (or humor creation ability). All quality scores are significantly correlated positively with P, and most highly so the Ss' rated wit (CPPT WI). CPPT FA also correlated significantly positively with E. Furthermore, lower scores in the Lie-scale went along with higher values for wittiness, wit, and fantasy. Neuroticism was not related either to quantity or to quality of punch line production.

## Discussion

The aim of the present study was fourfold. Besides the evaluation of the reliability of all inventories used, convergent and discriminant validities were examined for the homologous dimensions relative to humor appreciation and humor creation. Moreover, the number and nature of the dimensions underlying the sense of humor self-report instruments were determined putting the findings of the prior study (Ruch 1994a) to a first test. Finally, the relationship of sense of humor and the PEN system was evaluated by correlating the emerging self-report factors of the sense of humor and the scales of the 3 WD-K and the CPPT (as representatives of the behavioral measures) with the EPQ-R.

### How much of the variance of humor instruments is reliable?

Inspection of the reliabilities of the instruments included in this study yielded that the proportion of reliable variance for most of the scales is sufficiently high. Hence, the proportion of error variance for the most part is low. However, some scales (SHQ-3 M, TDS-SM) only showed coefficient-alphas smaller than .50, indicating that a great part of their total variance has to be considered as error variance. However, it should be recalled that reliability was based on internal consistency only (and not on stability), and that the results might be representative of the German versions of the instruments only. Nevertheless, researchers need to begin to draw the distinction whether an instrument is suitable for research purposes (that is, when gross distinction of groups is sufficient) or also for differential diagnostics (that is, when a reliable discrimination among individuals needs to be achieved). While for the former reliability needs to exceed .60, coefficients of .80 or higher are required for the latter.

### Convergent and discriminant validity of instruments measuring humor appreciation and humor creation

Can humor appreciation and humor creation be validly distinguished as different components of the sense of humor? The answer is a clear "yes" for the behavioral measures. As in the Koppel and Sechrest study, there is only a very low positive correlation between humor appreciation and humor creation performance. In other words, those who rate jokes and cartoons as funny are not necessarily the ones who are able to produce many or funny punch lines; and vice versa, the wit may equally well be a person who appreciates humor or who dislikes the humor of others. Besides discriminant validity, there is evidence for convergent validity as well; the indices of quality and quantity of humor creation ability correlate highly positively, and total funniness and aversiveness correlate negatively.

For the questionnaire measures, the outcome is highly different. The monomethod correlations suggest that while convergence was clearly given for self-reported humor creation, the self-report scales of humor appreciation did not form a homogeneous cluster (that is, the correlations were lower and partly even insignificantly so)<sup>11</sup>. Moreover, there seems to be a lack of discriminant validity. The self-report scales of humor appreciation and humor creation were highly intercorrelated; in fact, the heterotrait coefficients were even higher than the monotrait ones.

A more powerful examination of convergent and discriminant validity is provided by the heteromethod-correlations; that is, the coefficients concerning the relationship among the homologous and heterologous scales as assessed by the two different methodologies (performance and self-report). For both humor appreciation and humor creation the evidence for convergent validity was rather meager. While all of the correlations had the expected sign, the coefficients were low and not always significant. The prediction was not much better once all the self-report scales were combined in multiple regression analyses. Consequently, there was also a failure to demonstrate discriminant validity.

Apparently, even homologous scales of different methodological approaches do arrive at only slightly overlapping conclusions regarding the individual's ability to appreciate or create humor. How can this be explained? (1) This might be an indicator of the presence of larger portions of method variance. If so, one would expect that the self-report instruments are more heavily loaded by it; this method variance makes the (almost) unrelated components of humor appreciation and humor creation correlate indistinguishably highly in self-reports, and reduces their

---

<sup>11</sup> There might be two reasons for this heterogeneity. (1) While the instruments by Ziv and Martin and Lefcourt emphasize the affective nature of humor appreciation, the scales by Svebak and Thorson and Powell seem to locate humor appreciation in the domain of attitudes and convictions. (2) Some instruments (SHQZ, MSHS-HA, SHRQ) overwhelmingly or exclusively use positively worded items while the SHQ-3 Liking of Humor scale, however, contains exclusively negatively views and judgments of others.

correlations with the behavioral measures. Such an effect was previously apparent in the MTMM study by Koppel and Sechrest (1970). (2) Irrespective of whether method variance is present or not, the low correlations may emerge from differences in the constructs being measured. In other words, it might be that the behavioral measures emphasize different aspects of humor appreciation and humor creation than the self-report instruments do (which, in turn, have little overlap as regards humor appreciation). In other words, not only are the measures different, but the constructs may also differ, or, of even more concern, the construct definition may not be adequately represented in the item formulations.

There is much evidence in favor of the second explanation. What factors make the humor behavioral measures representative or nonrepresentative? On the positive side, they seem to be reduced to the core of the component to be measured; that is, whether the person perceives the jokes and cartoons as humorous or not, and whether the person is able to create new and funny punch lines. Thus, the face or content validity of the CPPT and the 3 WD is very high. Furthermore, the objectivity of administration prevents other factors (for example, a social component) from becoming involved, which might alter the responses. For example, in a social setting jokes might be perceived differently (depending on whether others laugh or not) and people might differ in respect to whether they have confidence in the punch lines they produced. On the negative side, one might argue that, for example, the 3 WD measures a very narrow aspect of humor appreciation, namely the appreciation of jokes and cartoons. This is true if one considers the item content. However, there are hints for a generalizability of appreciation of jokes and cartoons to the appreciation of other humorous stimulus material. Köhler (1993) found a positive relationship between the 3 WD scores and the funniness ratings of a weight judging task (sensu Deckers 1993), that is, after having lifted an incongruous weight following a series of forced-choice weight-comparisons. Furthermore, Frost (1992) found that the 3 WD is predictive of facial and verbal reactions to humorous film-clips of the same humor categories (INC-RES, NON, and SEX). Hence, the humor appreciation as measured with the 3 WD humor test is not restricted to jokes and cartoons, but can be considered an indicator of appreciation of humor per se. As regards the CPPT, one might argue correctly that it is not yet well established. However, its psychometric properties are sufficiently good (see Table 1) and it replicates findings of similar studies of self-report and performance tests of humor creation. For example, Nevo, Aharonson, and Klingman (1993) found a significant, but weak correlation between a punch line productivity index and the SHRQ, while, as in the present study, the Humor Creativity scale of Ziv's SHQ was unrelated to that index.

With regard to the humor self-report scales, there is evidence for both arguments listed above, namely that the constructs are different, and that the construct definition is not stringently converted in the content of the items. As regards the first argument, three issues need to be considered. (1) Some construct definitions are broader, covering appreciation of more global or everyday situations. Only a few items refer to jokes and cartoons. (2) The constructs have a different emphasis. For example, humor appreciation is understood as an attitude concept or as an affective concept. Likewise, self-report scales of humor creation put emphasis on reproduction or entertainment aspects rather than the mere wit, or creative production element. (3) Finally, the dislike aspect is often missing in the construct definition of self-report scales, and only a few items (the reversed ones) contain statements about explicit dislike of humor. In other words, they are unipolar constructs (that is, low vs. high appreciation, but not disliking vs. liking of humor).

As regards the second argument, namely the valid translation of the construct definition into item content, one can state that several items apparently lack content validity (that is, they do not reflect what the scale purports to measure). It seems that it would be of great interest to conduct a study examining the prototypicality of the whole pool of items of the inventories of the present study, as well as other scales that have been constructed recently. Thus, a sample of experts (or trained lay persons) should be provided with definitions of the constructs the scales are aimed to measure (the mere label of the scale is not sufficient, however) and judge whether and to what extent the item content matches the respective definitions.

Irrespective of these problems of content validity, the question arises as to what is, then, the "right" measurement approach to the assessment of humor appreciation and humor creation? First of all, the present authors do not follow

Babad's (1974) conclusion that self- and peer-report should be preferred to performance tests. No-one doubts that asking a person "Are you intelligent?" is a less reliable and valid source of information than to run an objective test of intelligence with that person. So why should an objective assessment of an ability be replaced by a self-report of that same ability; especially when such highly social desirable constructs like humor appreciation or creation are involved? The position taken here is that behavioral measures can not be replaced by self-reports (which, of course, are more economical), unless it has been demonstrated for that particular measure that it is able to predict humor performance with sufficient accuracy.

### Dimensionality of self-reported sense of humor

The present factor analysis comprised the 10 self-report scales previously used by Ruch (1994a) as well as the HIS, MSHS, and STCI-T Cheerfulness and Seriousness components. The results show that (1) the dimensions of "surgency" and "restraint vs. expressive" can be recovered from the present study, (2) the additional scales (HIS, MSHS) fit neatly in the two-dimensional frame-work (rather than requiring new dimensions), and (3) "cheerfulness" and "seriousness" are viable alternative interpretations; although not intending to measure the sense of humor (but rather its temperamental basis), the facets of the two STCI-T-scales of cheerfulness and seriousness merged with the genuine sense of humor-scales and some of them even yielded the highest loadings; that is, they may serve as markers of the two factors.

Due to the larger number of scales involved in the present study the meaning of the factors of cheerfulness and, especially, of seriousness are illuminated more clearly (see the result section for elaborated definitions). Furthermore, while seriousness was conceptualized to form one element of humorlessness (Ruch 1994b), the results show that it contributes to the differentiation of genuine sense of humor scales as well. While some scales (for example, those emphasizing affect) load exclusively on cheerfulness, some humor scales (for example, SHQ-3 M, MSHS-AH) have an additional negative loading on this factor, too. This might be due to the mental element involved in the items, for example, when the content is dealing with attitudes toward humorous situations and people. In general, it appears fruitful to distinguish the affective and mental-attitudinal components in the sense of humor more clearly. Also, it might be of interest to explore the 'empty' quadrants; that is, investigate what sorts of humor phenomena might be located in the cheerfulness-seriousness and low cheerfulness-low seriousness areas.

Finally, it should be noted that, contrary to the findings of Ruch (1994a), the SHRQ was unrelated to seriousness in the present study. Also, the Emotional Expressiveness scale of the SHQ-3 lost its marked negative loading on the "restraint vs. expressive" factor; probably because the revision eliminated its impulsivity component.

### Relationship between sense of humor and temperament

Generally, as in the prior study (Ruch 1994a), the PEN system of temperament provides a useful framework for locating the different components of the sense of humor used in the present study. The different localization of the individual components provides knowledge about both the PEN dimensions and the different dimensions of self-report and humor behavior. Again, Extraversion is the superfactor most relevant for humor. In general, the extravert is characterized by traits, such as sociable, lively, active, assertive, sensation-seeking, carefree, dominant, surgent, and venturesome. In the field of humor, extraverts (as compared to introverts) are more cheerful, less serious, and able to produce more (but not necessarily funnier) punch lines.

Psychoticism is made up of traits like aggressive, cold, egocentric, impersonal, impulsive, antisocial, unemphatic, creative, and tough-minded. As regards their humor profile, the high P-scorers are less serious and are persons of wit (that is, their captions were of higher funniness and originality, and they were judged as being higher in wit and

fantasy). Thus, their generally higher creative potential (Eysenck 1995) extends to the realm of humor production<sup>12</sup>. Additionally, as in prior studies<sup>13</sup>, low Psychoticism went along with one of the humor categories, namely funniness of incongruity-resolution humor. Thus, individuals characterized by high super ego strength, conformity, and oversocialization tend to like jokes and cartoons which contain punch lines with fully resolvable incongruities. They share this tendency with people high in conservatism and low in sensation seeking (Zuckerman 1994).

Neuroticism is made up of traits like anxious, depressed, guilt feelings, low self-esteem, tense, irrational, shy, moody, and emotional. While the pattern of correlations with humor appreciation (funniness negative, aversiveness positive) emerged as expected, none of the coefficients was significant. Neuroticism correlated negatively with cheerfulness, indicating that cheerfulness goes along with emotional stability. Two reasons might account for this finding differing from the results of the prior study. (1) In the present sample Neuroticism correlated more highly negatively than usual with Extraversion ( $r = -.23$ ,  $P < .05$ ). When partialing out Extraversion from Neuroticism, the coefficient of the cheerfulness factor with N is largely reduced ( $r = -.16$ , *n.s.*). (2) The newly included scales might have shifted the cheerfulness-factor towards stable extraversion. For example, some of the scales (CHS, MSHS-CH, STCI-T CH3 - composed cheerfulness) emphasize the aspect of humor as an antidote to negative affect; they should indeed correlate negatively with N (which represents negative affectivity).

## Conclusions

It is evident from the results of the present paper that humor research must pay more attention to the methodological issues involved in the assessment of the sense of humor. Clearly, it is not sufficient to rely on implicit assumptions of or lay concepts about its nature, do some brain-storming and write up a list of items, label the resulting questionnaire a "sense of humor" inventory and go into validation studies using small sample sizes. Research should be focused on at least two goals: (1) theoretical and empirical work aimed at a more precise outline and definition of the construct; that is, identification of the number and nature of validly distinguishable components, (2) application of a more sophisticated technology of constructing instruments for the assessment of the sense of humor and/or the components of this construct (see, for example, Angleitner, John, and Löhrl 1986; Kline 1993). While we should, of course, continue to study the relevance of the sense of humor in different fields of application, we will be able to integrate and accumulate research findings sooner and more effectively if we agree upon the theoretical components of the sense of humor and apply well-constructed instruments for its assessment.

## Notes

The data of this study stem from a research project grant provided by the German Research Council (DFG) (Ru 480/5-1). The preparation of this manuscript was facilitated by a Heisenberg-grant (Ru 480/1-1) awarded to Willibald Ruch. Special thanks to the Department of Physiological Psychology at the University of Düsseldorf (F.-J. Hehl) for hosting this research. Correspondence and reprint requests should be addressed to either author at Department of General Psychology, Heinrich-Heine-University of Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany.

---

<sup>12</sup> While the coefficients were low, one has to consider that judgments of wit and fantasy of the punch lines' authors were based on the inspected material only. Furthermore, the raters' general humor preference affects the judgments of how witty or how original the punch lines are rated; the use of 6 raters does not completely average out this tendency. Thus, one could argue that the coefficients in fact underestimate the real strength of relationship. It is also of interest to study this hypothesis in a homogeneous sample (as regards age, education etc.). For example, a refined analysis of the present data revealed that the P-wit-relationship was more pronounced in a subsample of younger adults.

<sup>13</sup> P was negatively correlated with funniness of incongruity-resolution humor in all samples studied previously (Hehl and Ruch 1985; see review by Ruch and Hehl 1985), but only significantly so in the ones with a greater number of subjects.

## References

- Angleitner, Alois, Oliver P. John, and Franz J. Löhrl  
1986 It's what you ask and how you ask it: An itemmetric analysis of personality questionnaires. In Angleitner, Alois and Jerry S. Wiggins (eds.), Personality assessment via questionnaires. Current issues in theory and measurement. Berlin, Germany: Springer, 61-108.
- Babad, Elisha Y.  
1974 A multi-method approach to the assessment of humor: A critical look at humor tests. Journal of Personality 42, 618-631.
- Bell, Nancy J., Paul E. McGhee, and Nelda S. Duffey  
1986 Interpersonal competence, social assertiveness, and the development of humor. British Journal of Developmental Psychology 4, 51-55.
- Brodzinsky, David M., and Janet Rubien  
1976 Humor production as a function of sex subject, creativity, and cartoon content. Journal of Consulting and Clinical Psychology 44, 597-600.
- Campbell, D.T. and D.W. Fiske  
1959 Convergent and discriminant validation by the Multitrait-Multimethod Matrix. Psychological Bulletin 56, 81-105.
- Clabby, John F.  
1980 The wit: A personality analysis. Journal of Personality Assessment 44, 307-310.
- Deckers, Lambert  
1993 On the validity of a weight-judging paradigm for the study of humor. Humor 6, 43-56.
- Derks, Peter L.  
1987 Humor production: An examination of three models of creativity. Journal of Creative Behavior 21, 325-326.
- Derks, Peter, and Dedreck Hervas  
1988 Creativity in humor production: Quantity and quality in divergent thinking. Bulletin of the Psychonomic Society 26, 37-39.
- Eysenck, Hans-Jürgen  
1995 Genius. The natural history of creativity. Cambridge: Cambridge University Press.
- Eysenck, Hans-Jürgen, and Michael W. Eysenck  
1985 Personality and individual differences: A natural science approach. New York: Plenum Press.
- Eysenck, Sybil B.G., Hans-Jürgen Eysenck, and Paul Barrett  
1985 A revised version of the psychoticism scale. Personality and Individual Differences 6, 21-29.
- Feingold, Alan, and Ronald Mazzella  
1991 Psychometric intelligence and verbal humor ability. Personality and Individual Differences 12, 427-435.
- Fisher, Seymour and Rhoda L. Fisher  
1983 Personality and psychopathology in the comic. In: McGhee, Paul E. and Jeffrey H. Goldstein, Handbook of humor research. New York: Springer, 41-60.
- Frost, Renate  
1992 Einfluß von stimmungsförderndem Verhalten eines Versuchsleiters auf die Erheiterung während der Darbietung humoriger Filmausschnitte. Unpublished masters thesis, Department of Psychology, University of Düsseldorf, Germany.
- Hehl, Franz-Josef, and Willibald Ruch  
1985 The location of sense of humor within comprehensive personality spaces: An exploratory study. Personality and Individual Differences 6, 713-715.
- Kline, Paul  
1993 The handbook of psychological testing. London: Routledge.
- Köhler, Gabriele  
1993 Das Weight Judging Paradigma (WJP): Eine Replikations- und Validierungsstudie. Unpublished masters thesis, Department of Psychology, University of Düsseldorf, Germany.
- Köhler, Gabriele, and Willibald Ruch  
1993 The Cartoon Punch line Production Test - CPPT. Unpublished manuscript. University of Düsseldorf, Department of Psychology, Düsseldorf, Germany.

- Koppel, Mark A., and Lee Sechrest  
 1970 A multitrait-multimethod matrix analysis of sense of humor. Educational and Psychological Measurement 30, 77-85.
- Lefcourt, Herbert M., and Rod A. Martin  
 1986 Humor and Life Stress. New York: Springer-Verlag.
- Martin, Rod A., and Herbert M. Lefcourt  
 1983 Sense of humor as a moderator of the relation between stressors and moods. Journal of Personality and Social Psychology 45, 1313-1324.  
 1984 Situational Humor Response Questionnaire: Quantitative measure of sense of humor. Journal of Personality and Social Psychology 47, 145-155.
- McGhee, Paul E.  
 1980 Development of the sense of humour in childhood: A longitudinal study. In: McGhee, Paul E. and Antony J. Chapman (eds.), Children's humour. Chichester: John Wiley and Sons, 213-236.
- Murgatroyd, S., C. Rushton, Michael Apter, and C. Ray  
 1978 The development of the Telic Dominance Scale. Journal of Personality Assessment 42, 519-528.
- Nevo, Ofra  
 1984 Appreciation and production of humor as an expression of aggression. Journal of Cross-Cultural Psychology 15, 181-198.
- Nevo, Ofra, Haim Aharonson, and Avigdor Klingman  
 1993 Is it possible to improve sense of humor? The development and evaluation of a systematic program for improving sense of humor. Paper delivered at the International Society for Humor Studies Conference - ISHS '93 in Paris, France.
- O'Quin, Karen, and Peter Derks  
 i.p. Humor and creativity: A review of the empirical literature. In: Runco, Marc A. (ed.) Creativity Research Handbook (vol. 1). Cresskill, NJ: Hampton Press.
- Pearson, Paul  
 1983 Personality characteristics of cartoonists. Personality and Individual Differences 4, 227-228.
- Ruch, Willibald  
 1983 Humor-Test 3 WD (Form A, B and K). Unpublished manuscript. University of Düsseldorf, Department of Psychology, Düsseldorf, Germany.  
 1992 Assessment of appreciation of humor: Studies with the 3 WD humor test. In: Spielberger, Charles D. and James N. Butcher (eds.), Advances in personality assessment, vol. 9. Hillsdale, NJ: Lawrence Erlbaum Associates, 27-75.  
 1994a Temperament, Eysenck's PEN system, and humor-related traits. Humor 7, 209-244.  
 1994b A state-trait approach to cheerfulness, seriousness, and bad mood: a progress report. Paper presented for the symposium Approaches to the sense of humor: Concepts and measurement (Chair: W. Ruch) conducted at the International Society for Humor Studies Conference - ISHS '94, Ithaca College, NY, June 22-26.
- Ruch, Willibald, Michael Freiss, and Gabriele Köhler  
 1993 The State-Trait Cheerfulness Inventory (STCI). Preliminary manual. Unpublished manuscript, Department of General Psychology, Heinrich-Heine-University Düsseldorf, Germany.
- Ruch, Willibald, and Franz-Josef Hehl  
 1985 Diagnose des Humors - Humor als Diagnostikum. In Hehl, Franz-Josef, Volker Ebel, and Willibald Ruch (eds.), Diagnostik psychischer und psychophysiologischer Störungen. Bonn: Deutscher Psychologen Verlag, 253-325.
- Svebak, Sven  
 1974 A theory of sense of humor. Scandinavian Journal of Psychology 15, 99-107.  
 1993 Sense of humor questionnaire - Revised (SHQ-3). Unpublished manuscript. University of Bergen, Norway.
- Thorson, James A., and F.C. Powell  
 1993 Development and validation of a multidimensional sense of humor scale. Journal of Clinical Psychology 49, 13-23.
- Ziv, Avner  
 1981 The self concept of adolescent humorists. Journal of Adolescence 4, 187-197.
- Zuckerman, Marvin  
 1994 Behavioral expressions and biosocial bases of sensation seeking. New York: Cambridge University Press.