Psychological Test and Assessment Modeling, Volume 64, 2022 (2), 186-209

The Big-Five Between The Lines

Approaching Quantitative Operationalization of Text Analytical Personality Assessment Using Linguistic Markers

Christina Glasauer[1] *& Rainer W. Alexandrowicz*[2]

Abstract

Personality assessment based on questionnaires can be distorted due to social desirability, deliberate faking, or unconscious factors like false memories. Using objective measures of personality can counteract these distortions as the intention of the assessment usually is not apparent. Present study aims to develop and examine an objective measure of personality based on quantitative text analysis. We use linguistic markers that have been shown to be related to the Big-Five personality traits in existing research and define a linguistic model of personality based on these findings. The linguistic model's usefulness in personality assessment is investigated by analyzing reliability and validity of the proposed procedure. After gathering personal texts and conventional measures of the Big-Five of $N = 124$ individuals, we processed text using LIWC and analyzed data in a structural equation modeling approach. Especially the linguistic scales of the dimensions Extraversion, Conscientiousness, and Agreeableness showed satisfying model fits. Neuroticism, Conscientiousness, and Openness showed promising convergence with traditional measures of personality. The results emphasize the potential of the quantitative text analytical approach to personality assessment and provide indications for future adaptions of the instrument.

Keywords: Personality Assessment, Speech, LIWC, Big-Five, Linguistic Markers

[1]University of Klagenfurt. *Correspondence concerning this article should be addressed to:* Christina Glasauer; Karl Popper Kolleg SEEROSE; Universität Klagenfurt; Universitätsstraße 65-67; 9020 Klagenfurt; Austria; email: christina.glasauer@aau.at

[2]University of Klagenfurt, Department of Psychology

Introduction

Since personality can be defined as the set of all temporally persistent patterns of habitual emotions, thinking, and behavior (Cloninger, 2009), personality traits are characterized by temporal stability and cross-situational nature. Because of this stability they offer promising possibilities for predicting, e. g., mental and physical health, academic or occupational success, or life satisfaction (e. g., Alizadeh et al., 2018; Alvarez-Moya et al., 2007; Rothstein & Goffin, 2006; Cardenal, Cerezo, Martínez, Ortiz-Tallo, & José Blanca, 2012; Grant, Langan-Fox, & Anglim, 2009). Such predictions require an appropriate assessment of the corresponding personality traits, which traditionally is often questionnaire-based and conducted by means of self-assessment. For the example of the so-called "Big-Five" (Goldberg, 1993; Fiske, 1949; McCrae & John, 1992), the BFI (Rammstedt & Danner, 2017) or the NEO-PI-R (Costa & McCrae, 1992) may be mentioned. Inherently problematic in personality assessment by questionnaires and interviews are social desirability, lack of self-awareness, bias due to deceptive memories, or willful distortion/faking (Ones, Viswesvaran, Dilchert, & Deller, 2006; Ziegler, 2015; Schmidt-Atzert, Krumm, & Amelang, 2021). The problem of social desirability can be attempted to be addressed with neutral item formulations (Bäckström, Björklund, & Larsson, 2009; Bäckström & Björklund, 2014) or scales developed specifically for this purpose (Crowne & Marlowe, 1960; Musch, Brockhaus, & Bröder, 2002; Stöber, 1999). Also, faking behavior and social desirability bias can be hampered with less transparent instruments to meet the quality criterion of tamper-resistance (Moosbrugger & Kelava, 2020). As a consequent continuation of this approach, the so-called "objective personality tests" (OPTs; Ortner, Proyer, & Kubinger, 2006) can be understood: In these operationalization and measurement intention are less obvious or even deliberately obscured and therefore complicate targeted behavioral control (Ortner, Proyer, & Proyer, 2015). In the present study, a new, "objective" procedure designed to capture the Big-Five personality traits through quantitative analysis of linguistic features is proposed, reviewed, and examined in terms of trends worth pursuing.

Background

Objective personality assessment

In the case of OPTs, the personality assessment is carried out without any apparent recognition of the measurement intention, which increases the objectivity of these tests and permits the evaluation of the collected data in a standardized form (Kubinger, 2006; Pawlik, 2006). These objective tests are mostly computer-based, which can also be perceived as beneficial in terms of the quality criterion of objectivity (especially

evaluation and implementation objectivity; cf. Moosbrugger & Kelava, 2020; Schmidt-Atzert et al., 2021). One possibility of objective testing consists of analyzing data from behavioral observations in which the trait to be measured is not made transparent, but rather certain behaviors in a situation are regarded as indicators of certain personality traits (Ortner et al., 2015, p. 141). In the approach examined here, language is understood as a specific form of behavior and is chosen as the basis of an OPT. The *objective* assessment of personality traits is based on text samples of the test subjects, which are analyzed by means of an automated scoring system and from which personality traits are inferred.

Quantitative text analysis

Since language plays a central role in conveying thoughts and emotions, linguistic utterances are full of clues for understanding people, as Tausczik and Pennebaker (2010, p. 25) explain the importance of linguistic analytic methods. Especially computer-based approaches to text analysis are of increasing interest: On the one hand the computational power of modern computers is constantly increasing. On the other hand, due to advancing digitization, more and more text material is available, e. g., in blogs, digitized books, or social media posts (Ireland & Mehl, 2014). In a differentiating summary of automated text analysis, Günther and Quandt (2016) distinguish between deductive methods, which are based on predefined sets of rules, and inductive methods, which explore text material for characteristics. Examples of text analytic methods and an analysis of their conceptual differences can be found in Mehl (2006). Nadkarni, Ohno-Machado, and Chapman (2011) provide an overview of existing approaches of artificially intelligent (AI) algorithms for language processing along with existing challenges for these methods. However, AI approaches to language processing that rely on the use of neural networks (NNs; such as IBM, 2020; PRECIRE Technologies GmbH, 2021) represent a "black box" from a psychological perspective, i. e., the process from input to output cannot be traced in detail because the nodes and weights of an NN defy a substantive interpretation. In contrast, non-AI-based methods such as the Linguistic Inquiry and Word Count (LIWC; Pennebaker, Boyd, Jordan, & Blackburn, 2015) software offer the advantage of higher transparency and better traceability of the evaluation, which is essential for understanding psychological contexts.

The LIWC software consists on the one hand of word category systems ("dictionaries") developed primarily for psychological questions, and on the other hand of a text processing module that analyzes the text word by word, compares it with word lists stored in the dictionary, and assigns it to corresponding categories. In addition to content-related criteria such as *Social Processes*, *Work*, or *Leisure*, these also include affective qualities such as *Positive Emotions*, *Anxiousness*, or *Anger* as well as linguistic-stylistic features

such as *Pronouns*, *1st Person Plural*, or *Negations*. The German dictionary DE-LIWC15 (Meier et al., 2018) comprises 18 711 words arranged in 77 categories. The result of a LIWC text analysis are relative category frequencies, which describe how often any words from the respective category are mentioned in the text relative to the total number of words.

Psychological correlates of linguistic features

Successful applications of quantitative language analytic tools such as the LIWC software to psychological-diagnostic questions have already shown inter- and intra-individual differences in language use as well as associations between linguistic patterns and person characteristics (Boyd & Pennebaker, 2015): Age groups and genders seem to differ in the use of filler words (Laserna, Seih, & Pennebaker, 2014), depression seems to be related to self-referential expression (Tackman et al., 2019), narcissism to the use of certain personal pronouns (Carey et al., 2015; Holtzman et al., 2019), and psychotic disorders could be predicted using algorithm-based and graph-theoretic approaches (Bedi et al., 2015; Elvevåg, Foltz, Rosenstein, & Delisi, 2010; Mota et al., 2012). Summary reports of inter- and intraindividual differences in language use are provided by Tausczik and Pennebaker (2010).

The Big-Five model of personality is also based on the analysis of language in its origins (Allport & Odbert, 1936; Goldberg, 1981; Galton, 1884). In addition to this historical relationship, personality and language also share the property of stability: Like personality, language use within a person appears to be relatively stable over time (Pennebaker, Mehl, & Niederhoffer, 2003). Studies have found associations between Big-Five factors and language patterns, especially with Hirsh and Peterson (2009) and Yarkoni (2010) reporting results of particularly extensive correlational studies.

Research question

In view of the advantages of more difficult tampering and reduced bias due to social desirability, the present study attempts to make an assessment in regard to the Big-Five according to the principle of an OPT on the basis of linguistic markers of written text. Given correlations between personality factors and linguistic markers the described in literature (Hirsh & Peterson, 2009; Yarkoni, 2010), these seem promising as a basis for personality assessment. Therefore, in this study the potential of a quantitative linguistic OPT is explored and subjected to a first test-theoretical examination.

Based on the linguistic markers of personality found in existing research, we formulate scales of linguistic markers characteristic for each factor of the Big-Five dimensions.

The scales are defined by selecting those linguistic markers that showed the strongest correlations with the Big-Five in the exploratory correlational studies of Hirsh and Peterson (2009) and Yarkoni (2010). Thus, the aim is to test whether such a text-based assessment of the Big-Five can represent a viable and valid approach to personality diagnostics.

Method

Study design

A cross-sectional study conducted online served to investigate the research question. The study was conducted in German. In addition to sociodemographic data, the survey included 1) a writing prompt to capture linguistic patterns and 2) a Big-Five personality questionnaire to validate the instrument. Due to the sequential presentation of the writing prompt and questionnaire, order effects may occur; hence, the order of these tasks was randomized.

Instruments

Writing instruction. Subjects were asked to write a *personal* text, as Hirsh and Peterson (2009) and Ireland and Mehl (2014) demonstrated that personal texts are more likely to reflect information about personality traits than factual texts. The writing instruction used is essentially the same as that of the "Expressive Writing Paradigm" (Kacewicz, Slatcher, & Pennebaker, 2007). The present version was only modified with respect to the writing frequency and the removal of the suggestion to write about friends, lovers, or family, since *Friends* and *Family* are language categories in the German LIWC dictionary and this suggestion could potentially bias the results. In addition, we changed the prompt used in the original to think about the "most traumatic experience of [one's] entire life" (Kacewicz et al., 2007, p. 272) to a prompt to think about an event that is "still bothering" the person to avoid excessive strain. Supplementary file 1 contains the writing prompt in German. Its English translation is:

> "Take 15–30 minutes and go to a quiet place where you are undisturbed. Think of an event from your past that is still bothering you and write about your thoughts and feelings related to that event. Let go of your expectations and explore your emotions and thoughts about the subject. You can relate the event to your past, present, or future. Who were you? Who would you like to be? Who are you right now? You can write about a single or multiple topics, any topic is possible. Just write down anything that

comes to mind, and try to write consistently if possible. Don't worry about spelling, sentence structure, and grammar. Just write as freely as possible. Everything you write is correct!"

IPIP40. The conventional personality questionnaire IPIP40 by Hartig, Jude, and Rauch (2003) is used to validate the text-based assessments regarding the Big-Five. This instrument is based on the International personality item pool (Goldberg, 1999). With 40 items, the IPIP40 incorporates the five Big-Five scales through eight items each, using a five-point response format. The IPIP40 scales show agreement with those of the NEO-FFI (Costa & McCrae, 1992; Borkenau & Ostendorf, 1993; Körner, Geyer, & Brähler, 2002) and satisfactory internal consistencies (Cronbach's alpha per scale: $\alpha_{Cr}^{(E)} = .87$, $\alpha_{Cr}^{(N)} = .90$, $\alpha_{Cr}^{(G)} = .84$, $\alpha_{Cr}^{(V)} = .72$, and $\alpha_{Cr}^{(O)} = .77$; Hartig et al., 2003).

Survey

The ethics committee of the University of Klagenfurt has approved the conduct of this study on 19.06.2020. Data collection was conducted online via the survey platform LimeSurvey (Limesurvey GmbH, 2013) in the period from 10.07.2020 to 23.10.2020. The subjects were recruited via the university-wide e-mail distribution list of the University of Klagenfurt as well as in the wider environment of the study managers. The duration of the study was between 30 and 60 minutes.

Sample

A total of 148 people completed the study in full. Texts that were not written in German or were less than 200 characters in length were excluded from further analysis, as from a methodological point of view no meaningful evaluation is possible without a sufficient amount of manifest observations. Of the remaining $N = 124$ subjects, 75 were female and the mean age of participants was $M = 26.97$ ($SD = 10.22$)[1]. About two-thirds (68.5 %) of the subjects reported studying psychology, 25 % were students of other disciplines, and 6.5 % were not studying. The two groups with different presentation order of subtasks comprised 48.4 % (experimental group 1: text/test) and 51.6 % (experimental group 2: test/text). The analyzed texts were on average $M = 422.42$ words long ($SD = 359.75$, $Mdn = 327.5$).

[1] We assessed age in categories (<20, 20–25, 25–30, 30–35, 35–45, 45–60, >60) and computed the mean age by replacing each category manifestation with the respective category's mean value, e. g., 22.5 years for the 20–25 bin. We replaced <20 with 19 years and >60 with 65 years.

Analysis

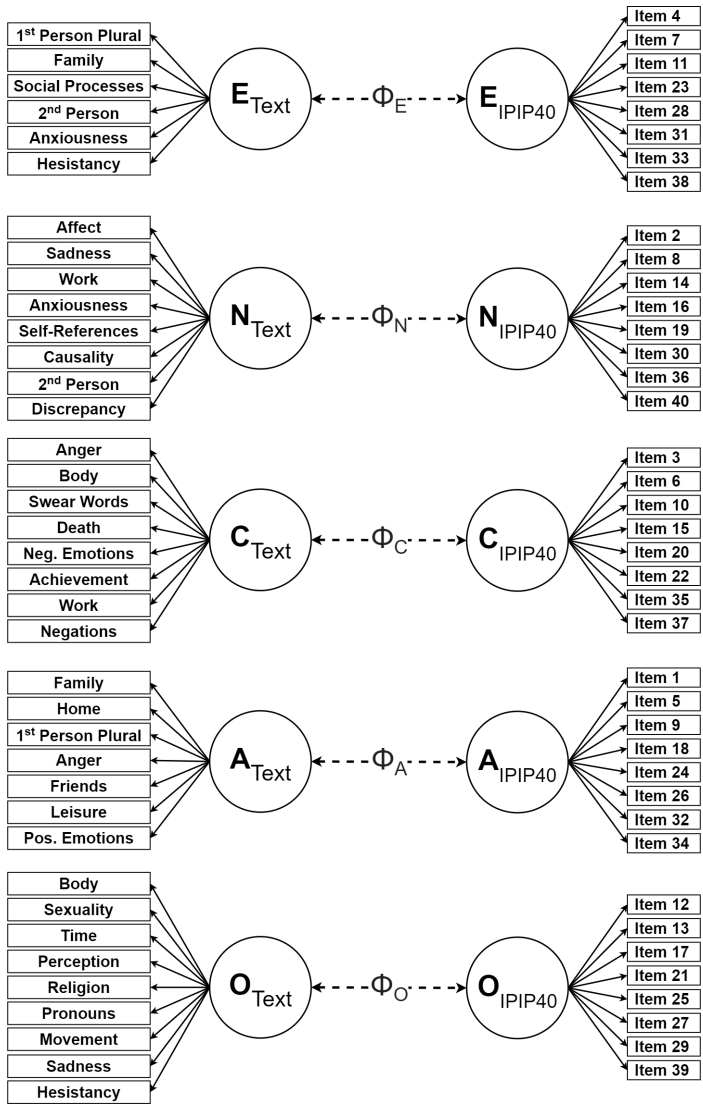
The IPIP40, which is used as a validation criterion, is also subjected to psychometric testing by means of confirmatory factor analysis (CFA). For each Big-Five scale of the questionnaire, the relevant personality construct is postulated as a latent factor in a one-dimensional model, on which the associated eight IPIP40 items are supposed to load as manifest indicators.

For the evaluation of the linguistic material, a categorization of the collected text material is performed by means of the LIWC software (Pennebaker, Booth, Boyd, & Francis, 2015; Pennebaker, Boyd, et al., 2015) using the German dictionary DE-LIWC15 (Meier et al., 2018) prior to further analysis. In the first step, the analysis of the individual scales is conducted by means of five one-factor confirmatory factor analyses (CFA) with the five personality factors as latent constructs and the respective linguistic markers as manifest variables. To assess model fit, we follow the conventions summarized by Ullman (2013). However, given the novelty of the approach, even lower fit measures are considered indicative of further research.

The text-based assessment of the Big-Five is also validated against the corresponding subscales of the IPIP40 as competing validity criteria using CFAs: For each Big-Five dimension, both the test-based and text-based personality factors are assumed to be latent constructs, and latent correlations (i. e., standardized latent covariances; Φ) between them are modeled to provide information about the validity of the proposed method. These five two-dimensional models with the latent correlations of text- and test-based factors are further referred to as BF_{corr}-models. In addition, we estimate one CFA per personality dimension, which also include the two corresponding factors of the text- and test-based measure, but do not allow for a correlation Φ between the latent constructs (BF_{ortho}-models). The χ^2-difference tests between the five BF_{corr}- and BF_{ortho}-models allow for the inferential statistical validation of the standardized latent covariance Φ, which are to be regarded as validity coefficients. Figure 1 depicts a graphical representation of the BF_{corr}-models, whereas the BF_{ortho}-models are identical except for the latent correlations being fixed to zero.

Finally, we investigate the text-based scales' equivalence to test-based personality assessment: For each Big-Five dimension, we estimate a one-factor model with one single *common* personality factor representing both the linguistic markers and the respective IPIP40 items (BF_{comm}-models). We compare the five BF_{comm}-models to the BF_{corr}-models by means of χ^2-difference tests. If both the linguistic assessment and the IPIP40 measure the same construct, the model fit of the BF_{comm}-models should be (statistically) equal to the fit of the BF_{corr}-models.

Figure 1: Visualization of the two-dimensional CFA models to be analyzed. For each Big-Five factor (E, N, C, A, O) a two-dimensional model *with* a latent correlation (BF_{corr}) as well as a two-dimensional model *without* a latent correlation (BF_{ortho}) is estimated.



Note: E: Extraversion, N: Neuroticism, C: Conscientiousness, A: Agreeableness, O: Openness. Circles: latent constructs, rectangles: manifest variables, continous arrows: loadings, dashed arrows: latent correlations Φ. The latent correlations between corresponding text- and test-based assessment of factors are present solely in the five BF_{corr}-models. In the five BF_{ortho}-models these correlations are set to zero.

Structural equation analysis is performed in R (R Core Team, 2017) using the `lavaan` package (Rosseel, 2012). Since the IPIP40 items are ordinally scaled and, in addition, some of the text variables are in dichotomous format, the Diagonally Weighted Least Squares (DWLS) method is used to estimate the CFAs. The reliability of the IPIP40 scales is determined using the reliability function from the `semTools` package (Jorgensen, Pornprasertmanit, Schoemann, & Rosseel, 2021), with the calculation based on polychoric correlations due to the ordinal scaling of the IPIP40 items (p. 154). The significance level for all analyses is assumed to be $\alpha = .05$.

Results

Possible order effect and text variables

Examination of the order effect revealed no significant difference between experimental group 1 (text/test; $M = 466.42$ words, $SD = 357.18$) and experimental group 2 (test/text; $M = 381.17$ words, $SD = 360.04$) for the word count of the composed texts with $t(122) = 1.32$, $p = .188$. A Mann-Whitney U-test, conducted supplementally due to the skewed frequency distribution of word count, similarly revealed no significant difference in word count between groups with $U(60,64) = 2295$, $p = .061$. With regard to responses in the IPIP40, a MANOVA also revealed no significant group difference in the five Big-Five dimensions with $F(1,122) = 0.867$, $p = .506$.

Since not all linguistic properties occurred in all texts, zero frequencies were present for some language categories. The frequency distributions of some text variables (especially *Swear words*, *Death*, *Sexuality*, *2nd Person Pronouns*, and *Anger*) showed an extremely skewed distribution due to the abundant occurrence of such zero frequencies. Therefore, a dichotomization of these variables into "used"/"not used" was performed for variables with a proportion of zero-frequencies greater than one third of the texts. At this point, we want to emphasize that this exploratory step represents an intervention in the data that was not specified in the study design and mark it as such.

Unidimensional scale analyses

In the first step of the analysis, the individual subscales of the text-based assessment and the IPIP40 were examined by means of unidimensional CFAs with respect to model fit and loadings.

Text-based assessment. The top section of Table 1 provides the results of the scale-wise CFAs for the five dimensions of the text-based assessment of the Big-Five. Table 2 depicts the markers used and their scale affiliations.

Only the CFAs of the scales Openness and Conscientiousness showed a (marginally) significant difference between the postulated model and the data. The remaining three scales did not show a significant χ^2-test. Examination of the magnitude of the loading coefficients of the individual linguistic markers in the unidimensional models provided information about which linguistic patterns are distinctive for each Big-Five personality dimension. Table 2 shows the estimated loading coefficients.

Significant loadings emerged for the extraversion scale for *1st Person Plural*, *Family*, *Social Processes*, and *2nd Person*. The Neuroticism factor predicted the linguistic markers *Affect*, *Sadness*, *Work*, *Anxiousness* and *Self-References* with loadings significantly differing from zero. For Conscientiousness, only the linguistic marker *Negations*, and for Agreeableness, only *Friends* were not significant. The latent factor Openness was only able to predict the text variables *Body*, *Sexuality*, and *Time*; no significant loadings were found for the remaining markers of this scale.

IPIP40. The IPIP40 scales (Hartig et al., 2003) Extraversion, Conscientiousness and Openness showed moderate to good fit indices. The internal consistencies of the IPIP40 scales ranged from .73 to .90. However, all scales of the IPIP40 except for Agreeableness had significant χ^2-tests (see Table 1). With the exception of a single item of the agreeableness scale (item "I believe others have good intentions" with a standardized loading coefficient $\lambda = .108, p = .089$), all items of the IPIP40 loaded significantly with $p = .000$ on the respective latent constructs.

The IPIP40 scale Extraversion was mainly characterized by inversely coded items: The items "I keep myself in the background" ($\lambda = -.868$), "I don't talk much" ($\lambda = -.826$), and "I don't like to draw attention to myself" ($\lambda = -.814$) loaded highest on the latent construct. The lowest loading was found with $\lambda = .550$ for the item "I know how to win people over." For the neuroticism scale, loadings were found from $\lambda = .401$ ("I panic easily") to $\lambda = -.799$ ("I am pleased with myself"). Conscientiousness was most strongly captured in the IPIP40 by the items "I carry out my plans most of the time" ($\lambda = .899$), "I make plans and stick to them" ($\lambda = .892$), and "I do unpleasant obligations immediately" ($\lambda = .751$). The item "I work only as much as I have to" loaded most weakly on this scale with ($\lambda = -.389$). The IPIP40 agreeableness scale was characterized by generally somewhat lower loadings with $\lambda = -.658$ for "I get rude quickly," $\lambda = -.639$ for "I insult people," and $\lambda = .575$ for "I respect others." The openness scale was most notable for the items "I don't like art" ($\lambda = -.957$), "I don't like going to art shows" ($\lambda = -.866$), and "I think art is important" ($\lambda = .791$). The

Table 1: Goodness-of-fit indices for the conducted CFAs. The upper third describes the model fits of the unidimensional models (text-based and IPIP40). The middle third shows the model fit and the latent correlation for each of the two-dimensional models (2-dim.) as well as the results of the model comparisons by means of χ^2-difference tests. The lower third (1-dim.) shows the BF_{comm}-models and their comparison with the respective two-factor BF_{corr}-models.

Model			χ^2	df	$\frac{\chi^2}{df}$	p	CFI	TLI	RMSEA [$CI_{90\%}$]	α_{Cr}	Φ	$p(\Phi)$
Text	E		11.23	9	1.25	.260	.94	0.90	.04 [.00, .12]	†	—	—
	N		27.15	20	1.36	.131	.90	0.86	.05 [.00, .10]	†	—	—
	C		32.08	20	1.60	.043	.86	0.80	.07 [.01, .11]	†	—	—
	A		10.70	14	0.76	.710	1.0	1.18	.00 [.00, .07]	†	—	—
	O		49.57	27	1.84	.005	.44	0.25	.08 [.04, .12]	†	—	—
IPIP40	E		39.33	20	1.97	.006	.99	0.99	.09 [.05, .13]	.90	—	—
	N		149.12	20	7.46	.000	.92	0.89	.23 [.20, .26]	.83	—	—
	C		42.05	20	2.10	.003	.99	0.98	.10 [.05, .14]	.87	—	—
	A		15.12	20	0.76	.770	1.0	1.03	.00 [.00, .05]	.73	—	—
	O		52.02	20	2.60	.000	.98	0.97	.11 [.08, .15]	.82	—	—
2-dim.	E	orth.	99.25	77	1.29	.045	.99	0.99	.05 [.01, .07]	†	—	—
		corr.	95.05	76	1.25	.069	.99	0.99	.04 [.00, .07]	†	.10	.352
		$\Delta\chi^2$	0.84	1	0.84	.359	—	—	—	–	—	—
	N	orth.	270.55	104	2.60	.000	.90	0.89	.11 [.10, .13]	†	—	—
		corr.	252.74	103	2.45	.000	.91	0.90	.11 [.09, .13]	†	.22	.020
		$\Delta\chi^2$	4.24	1	4.24	.040	—	—	—	—	—	—
	C	orth.	167.16	104	1.61	.000	.97	0.97	.07 [.05, .09]	†	—	—
		corr.	130.34	103	1.27	.036	.99	0.99	.05 [.01, .07]	†	.30	.013
		$\Delta\chi^2$	4.18	1	4.18	.041	—	—	—	—	—	—
	A	orth.	78.16	90	0.87	.809	1.0	1.05	.00 [.00, .03]	†	—	—
		corr.	74.80	89	0.84	.859	1.0	1.06	.00 [.00, .03]	†	.14	.296
		$\Delta\chi^2$	0.86	1	0.86	.347	—	—	—	—	—	—
	O	orth.	187.94	119	1.58	.000	.96	0.95	.07 [.05, .09]	†	—	—
		corr.	167.82	118	1.42	.002	.97	0.96	.06 [.04, .08]	†	.32	.026
		$\Delta\chi^2$	4.17	1	4.17	.041	—	—	—	—	—	—
BF_{comm}	E	comm.	124.04	77	1.61	.001	0.98	0.98	.07 [.05, .09]	†	—	—
		$\Delta\chi^2_{corr}$	17.81	1	17.81	.000	—	—	—	—	—	—
	N	comm.	301.15	104	2.90	.000	0.88	0.87	.12 [.11, .14]	†	—	—
		$\Delta\chi^2_{corr}$	25.78	1	25.78	.000	—	—	—	—	—	—
	C	comm.	178.13	104	1.71	.000	0.97	0.96	.08 [.06, .10]	†	—	—
		$\Delta\chi^2_{corr}$	18.79	1	18.79	.000	—	—	—	—	—	—
	A	comm.	103.25	90	1.15	.161	0.95	0.95	.04 [.00, .06]	†	—	—
		$\Delta\chi^2_{corr}$	17.85	1	17.85	.000	—	—	—	—	—	—
	O	comm.	179.28	119	1.51	.000	0.96	0.96	.06 [.04, .08]	†	—	—
		$\Delta\chi^2_{corr}$	6.58	1	6.58	.010	—	—	—	—	—	—

Note: E: Extraversion, N: Neuroticism, C: Conscientiousness, A: Agreeableness, O: Openness. CFI: Comparative Fit Index, TLI: Tucker-Lewis Index, RMSEA [$CI_{90\%}$]: Root-Mean-Square-Error-of-Approximation with 90 % Confidence interval for the RMSEA, Φ: standardized latent covariance (latent correlation) between text- and test-based personality factor, $p(\Phi)$: p-value of the latent correlation, $\Delta\chi^2$: Change in model fit based on the χ^2-difference test between 1) BF_{ortho}- and BF_{corr}-models and 2) BF_{comm}- and BF_{corr} models.

† Reliability coefficients cannot be estimated for the text-based scales because of the different scale levels of the linguistic markers (see Jorgensen et al., 2021, p. 154). An alternative calculation of internal consistencies across the undichotomized text scales at the rational-scale level by Pearson correlation is not meaningful due to the zero frequencies and skewed frequency distribution of some text variables, so we refrain from recording these coefficients for the text-based scales and report them only for IPIP40 scales.

Table 2: Standardized loadings λ, their standard errors ($S.E.$) and p-values of the linguistic markers for each unidimensional CFA of the text-based personality scales.

	Ling. Marker	λ	$S.E.$	p		Ling. Marker	λ	$S.E.$	p
E	1^{st} Person Pl.	.851	0.154	.000	A	Family	.762	0.149	.000
	Family	.503	0.134	.000		Home	.590	0.130	.000
	Soc. Processes	.416	0.429	.000		1^{st} Person Pl.	.441	0.171	.000
	2^{nd} Person	.361	0.139	.010		Anger	.382	0.114	.000
	Anxiousness	.269	0.146	.065		Friends	.174	0.120	.904
	Hesitancy	–.108	0.153	.322		Leisure	.075	0.112	.000
						Pos. Emotions	–.068	0.127	.009
N	Affect	.659	0.234	.000	O	Body	.773	0.196	.000
	Sadness	.524	0.115	.000		Sexuality	.460	0.153	.003
	Work	–.384	0.172	.001		Time	–.299	0.331	.019
	Anxiousness	.333	0.128	.009		Perception	.238	0.097	.051
	Self-references	.270	0.281	.002		Religion	.234	0.160	.143
	2^{nd} Person	–.114	0.095	.231		Pronouns	.085	0.380	.518
	Causality	–.056	0.092	.513		Movement	–.084	0.108	.429
	Discrepancy	–.006	0.104	.945		Sadness	.067	0.111	.579
						Hesitancy	.012	0.178	.926
C	Anger	–.678	0.120	.000					
	Body	–.635	0.127	.000					
	Swear words	–.617	0.163	.044					
	Death	–.569	0.114	.000					
	Neg. Emotions	–.391	0.130	.000					
	Achievement	.288	0.149	.009					
	Work	.233	0.171	.044					
	Negations	.013	0.112	.904					

Note: E: Extraversion, N: Neuroticism, C: Conscientiousness, A: Agreeableness, O: Openness.

lowest loading of the openness scale was found for the item "I tend to have a conservative political attitude" with $\lambda = -.314$.

Examination of validity

The IPIP40 served as a criterion for validating the text-based assessment of the Big-Five. With the exception of the neuroticism scale, which did not attain acceptable fit, the scales of the IPIP40 had acceptable model fits and reliability coefficients (see Table 1); thus, we considered them suitable for the present purpose. Validation was performed by analyzing the standardized latent covariances Φ ("latent correlations") between mutually corresponding text- and test-based personality factors (see Figure 1). Summary model metrics of the estimated CFAs are given in the middle third of Table 1.

Overall, the correlation coefficients (used here in the sense of validity coefficients) showed values between .1 and .32 (Table 1, penultimate column). For the scales Neuroticism, Conscientiousness, and Openness, the latent correlations were Φ significantly different from zero, and χ^2-difference tests between the BF_{corr}- and the BF_{ortho}-models therefore also showed significant model improvement for these scales when taking into account the correlation between text- and test-based assessed construct. The scales Extraversion and Agreeableness were characterized by good model fits of the two-dimensional models, but here neither the latent correlations Φ nor χ^2-difference tests between the corresponding BF_{corr}- and BF_{ortho}-models were significant.

Analysis of the BF_{comm}-models yielded acceptable model fits (see bottom third of Table 1): The agreeableness dimension yielded a non-significant result. Although the χ^2 values of the other models were significant, their fit indices showed good fit. Only for the common-factor model Neuroticism we found poor model fit. However, for each personality dimension, the model with a single common factor fit significantly worse than its corresponding two-factor model. These results indicate that the text- and the test-based assessment of each Big-Five factor are not equivalent in their current form.

Discussion

In the present study, a new language-based method for measuring the Big-Five was investigated with regard to its psychometric qualities. A first validation attempt was made against a self-assessment instrument.

Validation criterion IPIP40

The scales of the IPIP40 used for validation showed overall acceptable model fits in unidimensional CFAs. The scales Neuroticism and Openness showed somewhat worse model fits compared to the other scales, whereby especially the scale Neuroticism seems to be somewhat problematic. A possible explanation for the poorer fit of this scale is a series of similar items for which residual correlations can be assumed: Items 2 ("I feel comfortable the way I am.") and 36 ("I am satisfied with myself.") and both items 30 ("I rarely feel down.") and 40 ("I often feel down.") each depict very similar issues. An inspection of the modification indices of the CFA across the neuroticism scale indicated an improvement in model fit of $\Delta\chi^2_{(2/36)} = 35.43$ when allowing for correlation between the first two items (2 and 36), and an improvement of $\Delta\chi^2_{(30/40)} = 22.08$ when allowing for correlation between the latter pair of items (30 and 40), reinforcing this conjecture.

Also for the openness scale of the IPIP40, dependencies between the items that go beyond the underlying factor of Openness can be assumed: The scale has three items (12, 17, 25), each of which asks about liking for or disinterest in art. Another two items (27 and 39) ask about the tendency toward conservative or alternative political attitudes, respectively. For both triplets/pairs of items, high negative correlations independent of Openness can be expected and the slightly worse model fit of the openness scale may be explained that way ($\Delta\chi^2_{(12/17)} = 1.17$, $\Delta\chi^2_{(12/25)} = 2.82$, $\Delta\chi^2_{(17/25)} = 0.91$, $\Delta\chi^2_{(27/39)} = 3.01$). Given the high loadings for those items that query (dis)interest in art, the openness scale seems to reflect the extent of preference for art. These criticisms aside, the IPIP40 seems suitable as a validation criterion given the satisfactory model fits and reliability coefficients.

Language-based personality scales

With the exception of the openness dimension, the text-based scales of the Big-Five show moderate to good model fits, and the selected linguistic markers seem to reflect the remaining four personality factors of the Big-Five well. However, the loadings of these scales turned out to be only moderately high.

Consistent with theoretical assumptions about the construct extraversion, this personality trait seems to manifest itself textually primarily through mentions of other people and groups. The good model fit of the unidimensional CFA on this factor suggests that the extraversion scale represents a unified construct. Poor loadings for *Anxiousness* (positive loading) and *Hesitancy* (negative but very weak) can be considered an indication of discriminant validity.

On the dimension of neuroticism, expressions of *Affect* and *Sadness* loaded most strongly, *Anxiousness* somewhat less so. Expressions concerning *Work* loaded negatively and

Causality, *2nd Person*, and *Discrepancy* did not load at all. The last three, therefore, do not seem to play a role in capturing neuroticism. *Self-references* also showed only weak (but significant) loading. The associations between Neuroticism and the linguistic markers *Causality*, *2nd Person*, and *Discrepancy* that are found in existing literature are not shown to be predictable by this personality construct in the present study.

The conscientiousness scale is characterized primarily by strong negative loadings (i. e., the absence) of certain textual features: little expression of *Anger*, *Body*-related terms, *Swear words*, and *Death* seem central to this scale, with loadings between $-.569$ and $-.678$. In particular, the infrequent expression of *Swear words* and *Anger* is consistent with the theoretically assumed purposeful and dutiful character of conscientious individuals, characterized by willpower rather than affect (Berth & Goldschmidt, 2006). The lower but statistically significant loading for *Achievement* plausibly complements this picture of linguistically manifested conscientiousness.

For Agreeableness, high loadings are found for *Family* and *Home*, medium loadings are found for the linguistic marker *1st Person Plural*. The rather low but significant loading for *Anger* as well as the zero loading for *Positive Emotions* contradicts the theoretically assumed image of highly agreeable persons who are benevolent towards their environment.

The text-based openness scale, which has the worst model fit compared to the other scales, is characterized primarily by high loadings in the categories of *Body*- and *Sexuality*-related terms. According to Berth and Goldschmidt (2006), openness is defined by interest in (and the extent of engagement with) new experiences and impressions. The concept of openness includes "enjoying having new experiences": Even if the preoccupation with the topics *Body* and *Sexuality* speaks for a high expression of this characteristic, it remains doubtful that these depict the central characteristics of this dimension. Rather, the linguistic openness scale does not seem to represent a uniform construct at the current stage of development, which is also expressed by the moderately convincing model fit of this scale in the unidimensional analysis.

One possible reason for the partially low loadings of the linguistic markers is multidimensionality within individual items: The markers *Hesitancy*, *Anger*, *Body*, *2nd Person*, *1st Person Plural*, *Anxiousness*, and *Family* showed correlations with *several* Big-Five dimensions in the results of Hirsh and Peterson (2009) and Yarkoni (2010) and thus also loaded on two latent factors each in the present study. Given this multidimensionality of the linguistic markers in question, it is obvious that a single personality dimension cannot fully explain the variance in these markers. This would require a decomposition of the relevant linguistic markers into their individual, one-dimensional aspects, which would be possible by adapting the category system in the direction of finer language categories.

Validity

The first validation attempt of the text-based Big-Five scales carried out here using the IPIP40 as a criterion resulted in latent correlation coefficients of .1 to .32 (see Table 1). Thus, it showed only a low level of agreement between the text-based and test-based measurement (significant for Neuroticism, Conscientiousness, and Openness, however). The analysis of equivalence of text- and test-based measure also turned out poorly: For each dimension, a two-factor solution was superior to a one-factor solution, suggesting that the text-based measure can not replace the questionnaire assessment at the present maturity level of the approach.

Still, in view of the novelty of the instrument and its early stage of development, even rather low validity coefficients are of interest, since they can certainly be perceived as indications of the feasibility of the assessment carried out experimentally here. We see potential for development in the present findings; in particular, the scales Neuroticism, Conscientiousness, and Openness seem promising with slightly higher validity coefficients. The well-known difficulty that the validation of OPTs often turns out sobering, as questionnaires and objective tests often measure different aspects of the same construct (Ortner et al., 2015), seems to us to be a plausible explanation that has to be taken into account when considering the present results.

Limitations

The dichotomization of some very infrequently used language categories, necessitated by the specific data situation, has to be seen critically. To face this issue, future approaches might include a modified writing instruction, which explicitly asks for writing about several topics. Also the assessment over the course of several writing sessions, in which topics specifically adapted to the construct of interest are dealt with, might counteract this issue infrequent linguistic categories.

For the definition of the textual scales of personality, the language categories included in the German LIWC dictionary provided an overall framework but were limiting at the same time: Although the LIWC dictionary includes many categories especially relevant for psychological questions, these are relatively coarse. Also, word meanings can vary depending on context, and irony and sarcasm cannot be captured by simple word-counting. In view of the diversity and changeability of language, it must be regarded as insufficiently differentiated to represent all aspects of linguistic utterances with 77 categories. It would contribute to a more differentiated language assessment to develop own category systems (“dictionaries”), which are more finely defined and illuminate more differentiated aspects.

Further refinement of the dictionary used will allow for using the proposed method in a clinical/therapeutical context as well. We could show that the text-based approach provides a foundation for assessing personality traits, but the method is not yet applicable in its current form. In this context, the snowball sampling was chosen to gain insight into the potential of our method and will, consequently, not allow for establishing norm tables.

The basis of the model definition applied here was formed by results from Hirsh and Peterson (2009) and Yarkoni (2010). While the former study analyzed personal texts that averaged more than 15,000 words, the latter examined $N = 674$ online blogs. Here, it is critical to note that speech material from publicly available online blogs might differ systematically from personal texts. Furthermore, due to social desirability or pragmatic goals of bloggers (in the sense of distinguishing the levels of syntax, semantics, and pragmatics), the relationship between the Big-Five and the language categories studied may be biased. At the same time, however, due to the complementary nature of the texts studied, it is also advantageous that the results of analyses of diverse language material were used to define the model. It should be noted that even these writing opportunities, which have not yet been optimized in terms of psychological-diagnostic questions, yielded promising indications.

Conclusion and outlook

After initial piloting of the personality instrument investigated in this study, the presented procedure appears to be promising, although it is not yet able to capture the Big-Five with sufficient accuracy and validity at the current stage of development. Especially for the scales Extraversion, Conscientiousness, and Agreeableness, however, good model fits and loadings of linguistic markers were found, which seem promising as a basis for future development of the innovative instrument. The linguistic scales Neuroticism, Conscientiousness, and Openness also showed correlations with the corresponding IPIP40 scales, which can be taken as an indication of validity.

For the future advancement of the instrument, the use of more differentiated language categories would be desirable as a first step. This can be implemented by developing a specifically tailored LIWC dictionary that captures more diverse and more fine-grained aspects of language. The LIWC software supports the use of specially defined dictionaries. We expect better results in terms of model fit and equivalence to self-report measures when using more differentiated linguistic markers. In addition, to capture language content in texts in a more sophisticated way, the use of text parsers for syntactic analysis would be useful (e. g., elaborated in Jaf & Calder, 2019 and Collins, 2003): By taking grammatical structures into account, valid operationalization of content meaning of linguistic utterances can be supported. The analysis of more comprehensive textual

material, ideally collected in the course of multiple writing sessions, would also be desirable in future studies: Multiple writing sessions could provide more representative and less situationally weighted insight into language behavior.

The proposed personality instrument can also be adapted with regard to personality models other than the Big-Five and validated on the basis of corresponding inventories. It would be possible to adapt the instrument to the HEXACO personality model (Ashton & Lee, 2008), which includes a sixth dimension, honesty-humility, in addition to slightly modified versions of the "big five" factors. Guidance for defining a linguistic honesty-humility factor is provided by, e. g., Hancock, Curry, Goorha, and Woodworth (2007) and Newman, Pennebaker, Berry, and Richards (2003) with findings on linguistic patterns of lying. Validation against the HEXACO-60 (Ashton & Lee, 2009) would be conceivable and could provide important information about linguistic patterns of personality due to the slightly different factor definition in the HEXACO model compared to the Big-Five model.

Transferring the text analytic approach to capture the Big-Five presented in this study to psychological constructs other than personality seems equally worth pursuing. For example, findings on linguistic correlates of narcissistic personality disorder (Carey et al., 2015; Holtzman et al., 2019) may serve as a basis for developing a text-analytic instrument to assess narcissism. Applications in the field of psychotherapy research, such as research on mentalization-based psychotherapy, are also conceivable: Many procedures for assessing mentalizing ability are based on ratings of transcripts whose focus is on content rather than therapeutic processes (Shaw, Lo, Lanceley, Hales, & Rodin, 2020). The use of an objective, language analytic approach, such as the one presented here, could allow for comparisons of a person's level of mentalizing ability between therapy sessions and thus objectively map changes in mentalizing ability over time.

In summary, the presented quantitative language analytic instrument for personality diagnostics gives promising indications for future development. Despite the use of a non-optimized language model, the scales Extraversion and Agreeableness showed good model fits and the scales Neuroticism, Conscientiousness, and Openness showed promising validity coefficients. After thorough refinement of the underlying category system, this method could find use as an instrument for objective assessment of personality dimensions.

References

Alizadeh, Z., Feizi, A., Rejali, M., Afshar, H., Keshteli, A. H., & Adibi, P. (2018). The Predictive Value of Personality Traits for Psychological Problems (Stress, Anxiety and Depression): Results from a Large Population Based Study. *Journal of Epidemiology and Global Health*, *8*(3-4), 124–133. doi: 10.2991/j.jegh.2017.11.003

Allport, G. W., & Odbert, H. (1936). Trait-Names: A Psycho-lexical Study. No.211. *Psychological Review Monographs*, *47*(1). doi: 10.1037/h0093360

Alvarez-Moya, E. M., Jiménez-Murcia, S., Granero, R., Vallejo, J., Krug, I., Bulik, C. M., & Fernández-Aranda, F. (2007). Comparison of Personality Risk Factors in Bulimia Nervosa and Pathological Gambling. *Comprehensive psychiatry*, *48*(5), 452–457. doi: 10.1016/j.comppsych.2007.03.008

Ashton, M. C., & Lee, K. (2008). The HEXACO Model of Personality Structure and the Importance of the H Factor. *Social and Personality Psychology Compass*, *2*(5), 1952–1962. doi: 10.1111/j.1751-9004.2008.00134.x

Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A Short Measure of the Major Dimensions of Personality. *Journal of Personality Assessment*, *91*(4), 340–345. doi: 10.1080/00223890902935878

Bäckström, M., & Björklund, F. (2014). Social Desirability in Personality Inventories. *Journal of Individual Differences*, *35*(3), 144–157. doi: 10.1027/1614-0001/a000138

Bäckström, M., Björklund, F., & Larsson, M. R. (2009). Five-Factor Inventories have a Major General Factor Related to Social Desirability Which Can be Reduced by Framing Items Neutrally. *Journal of Research in Personality*, *43*(3), 335–344. doi: 10.1016/j.jrp.2008.12.013

Bedi, G., Carrillo, F., Cecchi, G. A., Fernandez Slezak, D., Sigman, M., Mota, N., ... Corcoran, C. (2015). Automated Analysis of Free Speech Predicts Psychosis Onset in High-Risk Youths. *npj Schizophrenia*, *1*. doi: 10.1038/npjschz.2015.30

Berth, H., & Goldschmidt, S. (2006). Testinformation: NEO-PI-R. NEO-Persönlichkeitsinventar nach Costa und McCrae. Revidierte Fassung [Testinfor-

mation: NEO-PI-R. NEO-Personality Inventory by Costa and McCrae. Revised Version]. *Diagnostica*, *52*(2), 95–103. doi: 10.1026/0012-1924.52.2.95

Borkenau, P., & Ostendorf, F. (1993). *NEO-Fünf-Faktoren Inventar (NEO-FFI) nach Costa und McCrae [NEO-Five Factor Inventory (NEO-FFI) adapted from Costa and McCrae]*. Göttingen: Hogrefe.

Boyd, R. L., & Pennebaker, J. W. (2015). A Way with Words: Using Language for Psychological Science in the Modern Era. In C. V. Dimofte, C. P. Haugtvedt, & R. F. Yalch (Eds.), *Consumer Psychology in a Social Media World* (pp. 222–236). Routledge: Taylor & Francis Group.

Cardenal, V., Cerezo, M. V., Martínez, J., Ortiz-Tallo, M., & José Blanca, M. (2012). Personality, Emotions and Coping Styles: Predictive Value for the Evolution of Cancer Patients. *The Spanish Journal of Psychology*, *15*(2), 756–767. doi: 10.5209/rev_SJOP.2012.v15.n2.38887

Carey, A., Brucks, M., Küfner, A., Holtzman, N., Deters, F., Back, M., ... Mehl, M. (2015). Narcissism and the Use of Personal Pronouns Revisited. *Journal of Personality and Social Psychology*, *109*, e1–e15. doi: 10.1037/pspp0000029

Cloninger, S. (2009). Conceptual Issues in Personality Theory. In P. J. Corr & G. Matthews (Eds.), *The Cambridge Handbook of Personality Psychology* (pp. 3–26). Cambridge and New York: Cambridge University Press.

Collins, M. (2003). Head-Driven Statistical Models for Natural Language Parsing. *Computational Linguistics*, *29*(4), 589–637. doi: 10.1162/089120103322753356

Costa, P., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) Professional Manual*. Odessa, FL: Psychological Assessment Resources.

Crowne, D. P., & Marlowe, D. (1960). A New Scale of Social Desirability Independent of Psychopathology. *Journal of Consulting Psychology*, *24*(4), 349–354. doi: 10.1037/h0047358

Elvevåg, B., Foltz, P., Rosenstein, M., & Delisi, L. (2010). An Automated Method to Analyze Language Use in Patients with Schizophrenia and their First-Degree Relatives. *Journal of Neurolinguistics*, *23*, 270–284. doi: 10.1016/j.jneuroling .2009.05.002

Fiske, D. W. (1949). Consistency of the Factorial Structures of Personality Ratings from Different Sources. *Journal of Abnormal Psychology*, *44*(3), 329–344. doi: 10.1037/h0057198

Galton, F. (1884). Measurement of Character. *Fortnightly Review*, *36*, 179–185.

Goldberg, L. R. (1981). Language and Individual Differences: The Search for Universals in Personality Lexicons. In L. Wheeler (Ed.), *Review of Personality and Social Psychology* (Vol. 2, pp. 141–165). Beverly Hills, CA: Sage.

Goldberg, L. R. (1993). The Structure of Phenotypic Personality Traits. *The American Psychologist*, *48*(1), 26–34. doi: 10.1037/0003-066X.48.1.26

Goldberg, L. R. (1999). A Broad-Bandwidth, Public Domain, Personality Inventory Measuring the Lower-Level Facets of Several Five-Factor Models. In I. Mervielde, I. Deary, P. de Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe* (pp. 7–28). Tilburg, Niederlande: Tilburg University Press.

Grant, S., Langan-Fox, J., & Anglim, J. (2009). The Big Five Traits as Predictors of Subjective and Psychological Well-Being. *Psychological Reports*, *105*, 205–231. doi: 10.2466/PR0.105.1.205-231

Günther, E., & Quandt, T. (2016). Word Counts and Topic Models. *Digital Journalism*, *4*(1), 75–88. doi: 10.1080/21670811.2015.1093270

Hancock, J. T., Curry, L. E., Goorha, S., & Woodworth, M. (2007). On Lying and Being Lied To: A Linguistic Analysis of Deception in Computer-Mediated Communication. *Discourse Processes*, *45*(1), 1–23. doi: 10.1080/01638530701739181

Hartig, J., Jude, N., & Rauch, W. (2003). *Entwicklung und Erprobung eines deutschen Big-Five-Fragebogens auf Basis des International Personality Item Pools (IPIP40)* [Development and Testing of a German Big-Five Questionnaire Based on the International Personality Item Pool (IPIP40)] (No. 1). Frankfurt.

Hirsh, J. B., & Peterson, J. B. (2009). Personality and Language Use in Self-Narratives. *Journal of Research in Personality*, *43*(3), 524–527. doi: 10.1016/j.jrp.2009.01.006

Holtzman, N., Tackman, A., Carey, A., Brucks, M., Küfner, A., Deters, F., … Mehl, M. (2019). Linguistic Markers of Grandiose Narcissism: A LIWC Analysis of

15 Samples. *Journal of Language and Social Psychology*, *38*, 773–786. doi: 10.1177/0261927X19871084

IBM. (2020). *IBM Watson: Natural Language Understanding.* Armonk, NY, US. Retrieved 18.02.2021, from `https://www.ibm.com/cloud/watson-natural-language-understanding`

Ireland, M., & Mehl, M. (2014). Natural Language Use as a Marker of Personality. In T. M. Holtgraves (Ed.), *Oxford Handbook of Language and Social Psychology* (pp. 201–218). New York: Oxford University Press. doi: 10.1093/oxfordhb/9780199838639.013.034

Jaf, S., & Calder, C. (2019). Deep Learning for Natural Language Parsing. *IEEE Access*, *7*, 131363–131373. doi: 10.1109/ACCESS.2019.2939687

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2021). *semTools: Useful Tools for Structural Equation Modeling (Version 0.5-4).* Retrieved from `https://CRAN.R-project.org/package=semTools`

Kacewicz, E., Slatcher, R. B., & Pennebaker, J. W. (2007). Expressive Writing: An Alternative to Traditional Methods. In L. L'Abate (Ed.), *Low-Cost Approaches to Promote Physical and Mental Health* (pp. 271–284). New York, NY: Springer New York. doi: 10.1007/0-387-36899-X_13

Körner, A., Geyer, M., & Brähler, E. (2002). Das NEO-Fünf-Faktoren Inventar (NEO-FFI) [German validation of the NEO Five-Factor Inventory]. *Diagnostica*, *48*(1), 19–27. doi: 10.1026//0012-1924.48.1.19

Kubinger, K. D. (2006). Ein Update der Definition von Objektiven Persönlichkeitstests: Experimentalpsychologische Verhaltensdiagnostik [An Update of the Defintion of Objective Personality Tests: Experimental Psychological Behavioral Diagnostics]. In T. M. Ortner, R. T. Proyer, & K. D. Kubinger (Eds.), *Theorie und Praxis Objektiver Persönlichkeitstests* [Theory and Practice of Objective Personality Tests] (pp. 38–52). Bern: Huber.

Laserna, C., Seih, Y.-T., & Pennebaker, J. W. (2014). Um . . . Who Like Says You Know: Filler Word Use as a Function of Age, Gender, and Personality. *Journal of Language and Social Psychology*, *33*, 328–338. doi: 10.1177/0261927X14526993

Limesurvey GmbH. (2013). *LimeSurvey: An Open Source Survey Tool.* Hamburg,

Germany: Author. Retrieved from http://www.limesurvey.org

McCrae, R. R., & John, O. P. (1992). An Introduction to the Five-Factor Model and Its Applications. *Journal of Personality*, *60*(2), 175–215. doi: 10.1111/j.1467-6494.1992.tb00970.x

Mehl, M. (2006). Quantitative Text Analysis. In M. Eid & E. Diener (Eds.), *Handbook of Multimethod Measurement in Psychology* (pp. 141–156). Washington, DC: American Psychological Association. doi: 10.1037/11383-011

Meier, T., Boyd, R. L., Pennebaker, J. W., Mehl, M. R., Martin, M., Wolf, M., & Horn, A. B. (2018). *"LIWC auf Deutsch": The Development, Psychometrics, and Introduction of DE-LIWC2015*. Retrieved from https://osf.io/tfqzc/ doi: 10.31234/osf.io/uq8zt

Moosbrugger, H., & Kelava, A. (2020). Qualitätsanforderungen an Tests und Fragebogen ("Gütekriterien") [Quality requirements for Tests and Questionnaires ("Quality Criteria")]. In H. Moosbrugger & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* [Test Theory and Questionnaire Design] (pp. 13–38). Berlin, Heidelberg: Springer.

Mota, N. B., Vasconcelos, N. A. P., Lemos, N., Pieretti, A. C., Kinouchi, O., Cecchi, G. A., … Ribeiro, S. (2012). Speech Graphs Provide a Quantitative Measure of Thought Disorder in Psychosis. *PloS one*, *7*, 1–9. doi: 10.1371/journal.pone.0034928

Musch, J., Brockhaus, R., & Bröder, A. (2002). Ein Inventar zur Erfassung von zwei Faktoren sozialer Erwünschtheit [An Inventory to Assess Two Factors of Social Desirabilty]. *Diagnostica*, *48*(3), 121–129. doi: 10.1026//0012-1924.48.3.121

Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural Language Processing: An Introduction. *Journal of the American Medical Informatics Association : JAMIA*, *18*(5), 544–551. doi: 10.1136/amiajnl-2011-000464

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying Words: Predicting Deception from Linguistic Styles. *Personality & Social Psychology Bulletin*, *29*(5), 665–675. doi: 10.1177/0146167203029005010

Ones, D. S., Viswesvaran, C., Dilchert, S., & Deller, J. (2006). Considering response distortion in personality measurement for industrial, work and organizational

psychology research and practice. *Psychology Science*, *48*(3).

Ortner, T. M., Proyer, R., & Proyer, T. (2015). Objective Personality Tests. In T. M. Ortner & F. J. R. van de Vijver (Eds.), *Behavior-based Assessment in Psychology* (pp. 133–149). Göttingen, Germany: Hogrefe.

Ortner, T. M., Proyer, R. T., & Kubinger, K. D. (Eds.). (2006). *Theorie und Praxis Objektiver Persönlichkeitstests* [Theory and Practice of Objective Personality Tests] (1. ed.). Bern: Huber.

Pawlik, K. (2006). Objektive Tests in der Persönlichkeitsforschung [Objective Tests in Personality Research]. In T. M. Ortner, R. T. Proyer, & K. D. Kubinger (Eds.), *Theorie und Praxis Objektiver Persönlichkeitstests* [Theory and Practice of Objective Personality Tests] (pp. 16–23). Bern: Huber.

Pennebaker, J. W., Booth, R., Boyd, R. L., & Francis, M. E. (2015). *Linguistic Inquiry and Word Count: LIWC2015: Operator's Manual.* Austin, Texas. Retrieved from `www.LIWC.net`

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The Development and Psychometric Properties of LIWC2015.* Austin, Texas.

Pennebaker, J. W., Mehl, M., & Niederhoffer, K. (2003). Psychological Aspects of Natural Language Use: Our Words, Our Selves. *Annual Review of Psychology*, *54*, 547–577. doi: 10.1146/annurev.psych.54.101601.145041

PRECIRE Technologies GmbH. (2021). *PRECIRE.* Aachen, Germany. Retrieved 18.02.2021, from `https://precire.com/`

R Core Team. (2017). *R: A Language and Environment for Statistical Computing (Version 3.4.3).* Vienna, Austria. Retrieved from `https://www.R-project.org/`

Rammstedt, B., & Danner, D. (2017). Die Facettenstruktur des Big Five Inventory (BFI): Validierung für die deutsche Adaptation des BFI [The Facet Structure of the Big Five Inventory (BFI): Validation for the German Adaptation of the BFI]. *Diagnostica*, *63*(1), 70–84. doi: 10.1026/0012-1924/a000161

Rosseel, Y. (2012). *lavaan: An R Package for Structural Equation Modeling (Version 0.6-9).* Retrieved from `http://www.jstatsoft.org/v48/i02/`

Rothstein, M. G., & Goffin, R. D. (2006). The Use of Personality Measures in Personnel Selection: What Does Current Research Support? *Human Resource Management Review*, *16*(2), 155–180. doi: 10.1016/j.hrmr.2006.03.004

Schmidt-Atzert, L., Krumm, S., & Amelang, M. (Eds.). (2021). *Psychologische Diagnostik* [Psychological Diagnostics] (6., completely revised ed.). Berlin, Heidelberg: Springer. doi: 10.1007/978-3-662-61643-7

Shaw, C., Lo, C., Lanceley, A., Hales, S., & Rodin, G. (2020). The Assessment of Mentalization: Measures for the Patient, the Therapist and the Interaction. *Journal of Contemporary Psychotherapy*, *50*(1), 57–65. doi: 10.1007/s10879-019-09420-z

Stöber, J. (1999). Die Soziale-Erwünschtheits-Skala-17 (SES-17): Entwicklung und erste Befunde zu Reliabilität und Validität [The Social Desirability Scale-17 (SDS-17): Development and first results on reliability and validity]. *Diagnostica*, *45*(4), 173–177. doi: 10.1026//0012-1924.45.4.173

Tackman, A., Sbarra, D., Carey, A., Donnellan, M., Horn, A., Holtzman, N., … Mehl, M. (2019). Depression, Negative Emotionality, and Self-Referential Language: A Multi-Lab, Multi-Measure, and Multi-Language-Task Research Synthesis. *Journal of Personality and Social Psychology*, *116*, 817–834. doi: 10.1037/pspp0000187

Tausczik, Y., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, *29*, 24–54. doi: 10.1177/0261927X09351676

Ullman, J. B. (2013). Structural Equation Modelling. In B. G. Tabachnick & L. S. Fidell (Eds.), *Using Multivariate Statistics* (pp. 681–785). Boston, Mass. and London: Pearson.

Yarkoni, T. (2010). Personality in 100,000 Words: A Large-Scale Analysis of Personality and Word Use among Bloggers. *Journal of Research in Personality*, *44*, 363–373. doi: 10.1016/j.jrp.2010.04.001

Ziegler, M. (2015). "F*** You, I Won't Do What You Told Me!" – Response Biases as Threats to Psychological Assessment. *European Journal of Psychological Assessment*, *31*(3), 153–158. doi: 10.1027/1015-5759/a000292