Journal of Research in Personality 44 (2010) 427–435

Contents lists available at ScienceDirect

Journal of Research in Personality

journal homepage: www.elsevier.com/locate/jrp

Tell me a story and I will tell you who you are! Lens model analyses of personality and creative writing

Albrecht C.P. Küfner *, Mitja D. Back, Steffen Nestler, Boris Egloff

Department of Psychology, Johannes Gutenberg University Mainz, 55099 Mainz, Germany

ARTICLE INFO

Article history:
Available online 7 May 2010

Keywords:
Personality
Big Five
Accuracy
Person perception
Lens model analyses
Personality judgment
LIWC
Text analysis
Writing

ABSTRACT

We elucidated the accuracy of personality judgments based on creative writing by means of lens model analyses. Targets ($N = 79$) wrote short stories with five predefined words. Observers rated the Big Five dimensions and general knowledge of the targets who wrote these stories. Three main findings were revealed: (a) the Big Five and general knowledge were consensually judged by observers; (b) judgments of openness to experience, agreeableness, and general knowledge were accurate; and (c) accuracies were achieved due to the correct usage of valid cues. Additionally, we replicated all results in a second sample of targets ($N = 126$). Results are discussed in comparison to other areas of personality judgment research.

1. Introduction

The accuracy of personality judgments at zero-acquaintance is a major topic of personality research that is studied across a wide range of social contexts, including real-life interactions (Levesque & Kenny, 1993; Watson, 1989), physical appearances (Borkenau, Brecke, Mottig, & Paelecke, 2009; Borkenau & Liebler, 1992; Naumann, Vazire, Rentfrow, & Gosling, 2009), and personal living spaces (Gosling, Ko, Mannarelli, & Morris, 2002). An important but seldom studied context in accuracy research is written material—the content of what people express linguistically. Linguistic content is the major vehicle for expressing and understanding social goals, preferences and messages. It is, moreover, of growing importance for our everyday lives, since more and more communication is made via word-based computer-mediated communication (e.g., internet chats, email, online social networks; Back, Schmukle, & Egloff, 2008; Back, Stopfer et al., 2010). Linguistic content might be closely related to personality and personality judgments (e.g., Fast & Funder, 2008; Pennebaker, Mehl, & Niederhoffer, 2003) for two reasons. First, people typically talk about who they are or what they do, think, and feel, automatically revealing something about their personality via the self-related content of their words. Second, people differ in *how* they talk about a given topic—what kinds of words they choose and how they combine them—automatically revealing something about their personality via the style of their linguistic expression. Here, we were interested in linguistic style, and investigated whether and how observers are able to give accurate personality judgments based on fictional short stories.

2. The lens model approach to the accuracy of personality judgments

In zero-acquaintance research (Albright, Kenny, & Malloy, 1988; Kenny & West, 2008) as well as other areas of interpersonal perception, lens model analysis (Brunswik, 1956) is a very suitable approach for investigating the different patterns of accuracies of personality judgments. Following this model, the personality of a *target* is expressed, and through this expression, perceived by *observers* via *cues*. Thus, accurate personality judgments can occur when (a) personality (e.g., extraversion) is expressed in (b) observable cues (e.g., smiling), which are (c) correctly used by observers (e.g., observers correctly judge people who smile more as more extraverted). The correlation between actual personality and a given cue is called *cue validity*; the correlation between each cue and the personality judgment is called *cue utilization*.

The lens model approach enables researchers to analyze the processes by which accuracy is achieved by looking at the actual behaviors and behavioral residues that mediate the relation between personality and personality-judgments. This is a crucial point because it also helps us to understand why certain personality traits are more easily judged (i.e., judgeable) than others (e.g., extraversion being a judgeable trait; Funder, 1999) and why this

* Corresponding author. Fax: +49 6131 3939154.
E-mail address: kuefner@uni-mainz.de (A.C.P. Küfner).

0092-6566/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jrp.2010.05.003

pattern may vary depending on the social context (e.g., openness being particularly judgeable based on personal environments or online profiles; Back, Stopfer et al., 2010; Gosling et al., 2002; Vazire & Gosling, 2004). Whereas the lens model approach has been repeatedly applied in the context of directly observable behaviors and physical appearances (Back, Schmukle et al., 2010; Borkenau & Liebler, 1992; Funder & Sneed, 1993; Gifford, 1994; Naumann et al., 2009), researchers have just started to use it in other domains, such as the domain of written material (Rodriguez, Holleran, & Mehl, 2010).

3. Prior research on word use

Writing and personality go hand in hand. Independent of whether the author is writing a novel, a diary, a pager message or a blog, word use seems to be linked to the author's cognitive, emotional, and social processes (e.g., Back, Küfner, & Egloff, in press; Cohn, Mehl, & Pennebaker, 2004; Fast & Funder, 2008; Hirsh & Peterson, 2009; Pennebaker et al., 2003; Pressman & Cohen, 2007; Yarkoni, 2010). Consequently, writing has been successfully used in psychotherapy (Frisina, Borod, & Lepore, 2004; Lumley, 2004; Pennebaker, 1997), for example. Writing patterns have moreover been shown to be consistent and stable over time (Pennebaker & King, 1999; Pennebaker et al., 2003). By using words, one can express one's personality most directly (Fast & Funder, 2008). First, writing shows people's characteristic feelings, motivations, and thoughts about themselves, their current situation, or their lives in general. Second, writing reveals differences in the kinds of words people use and how they combine them, irrespective of the topic they write about. Thus, differences in what people write about themselves, as well as differences in the writing style itself might be connected to the expression of personality traits.

A number of studies have investigated the linkage of written material (or transcribed spoken word material) and the Big Five personality dimensions (cue validity, in terms of the lens model). Most of them have used linguistic inquiry and word count (LIWC; Pennebaker, Francis, & Booth, 2001), which is currently the most used and best validated software for automatic text analysis in psychological research (Mehl, 2006). Pennebaker and King (1999) found correlations ranging from .10 to .16 for the Big Five and the usage of specific words (e.g., usage of articles and agreeableness correlated −.15) in stream-of-consciousness essays. Language and personality within transcribed life-history interviews were connected to several personality measures (self-reports, acquaintance ratings, and behavioral measures as criteria) up to $r = -.38$ for sexual words and the expression of sympathy toward a partner in a study conducted by Fast and Funder (2008). Similar results have been found by Hirsh and Peterson (2009) in self-narratives (on average, $r = .23$ with the Big Five dimensions).

To our knowledge, there are two studies that have investigated the accuracy of personality judgments based on written material. In the first study, Holleran and Mehl (2008) investigated the accuracy of personality judgments based on 20-min stream-of-consciousness essays. Their results showed significant accuracies for all Big Five traits, with correlations ranging from .29 (openness to experience) to .50 (conscientiousness). In the second study, Rodriguez et al. (2010) investigated how accurately lay observers could judge a target's subclinical depression from two different forms of written self-descriptions (personal diary vs. online blog). Average accuracy of these judgments was $r = .53$. As the authors noted, depression is highly evaluative, and thus, in most social situations, it is one of the traits that is less observable and therefore less judgeable (see also John & Robins, 1993; Vazire, 2010). However, when writing, people seem to give away more valid information about their thoughts and feelings. To explain these findings, Rodriguez and colleagues additionally applied lens model analyses using LIWC variables as cues. Results revealed that positive emotion words for both forms of self-descriptions were valid (i.e., were associated with targets' depression) and were correctly used by observers to judge targets' depression. For personal diaries, this applied also to sadness words (e.g., crying, agony), and cognitive mechanism words (e.g., cause, know). For the online blog condition, observers correctly used valid cues consisting of swear words, usage of past and present tense, and sleep words.

In sum, there are now a number of interesting results indicating that (a) personality expresses itself in self-related linguistic content, and (b) lay observers are indeed able to infer a person's personality when reading stream-of-consciousness essays or self-descriptions of that person. As described above, two broad aspects of writing might have contributed to the (quite remarkable) accuracy found in these studies. Valid information might have been made available for observers through (a) the content of self-related thoughts and experiences of their own lives and (b) the use and combination of words (writing style) per se, irrespective of the content described. Obviously, stream-of-consciousness essays as well as self-descriptions necessarily contain mostly self-related information. Thus, expressed content related to the targets' own lives and mind-sets certainly has contributed to the accurate perceptions of targets' personalities. But what about writing style? Does the style of writing alone tell us something about the personality of the writer, and are observers able to detect this information, thereby being able to accurately judge the personality of the writer? Thus, can fictional written material, without a contextual connection to an author's self, reflect his or her personality, and to what degree is this perceivable by lay observers?

4. Overview of the present studies

In the present studies, we investigated the accuracy of personality judgments based on creative writing. To this end, we let targets write fictional short stories, which were then given to lay observers to read and rate the targets on their personality characteristics. To understand accuracy patterns, different cues of writing style and their connection to the targets' personality scores and the personality judgments were analyzed by applying lens model analyses (Brunswik, 1956).

Four methodological features of the present research should be highlighted that were thought to strengthen the validity of possible empirical results. First, we let targets write fictional short stories constructed "around" the same predefined and not self-related words. This enabled us to analyze the usage and combinations of the actual words, controlling for the contextual information. This changed the focus from what people write about themselves to how they write per second.

Second, to gain a more comprehensive insight into the linguistic cues that might mediate the accuracy of personality judgments, we wanted to complement the word-count strategy of LIWC (Pennebaker et al., 2001). Particularly, to measure more complex linguistic structures like sarcasm or metaphoric speech within a given text, one needs other analytic strategies (Mehl, 2006; Wolf et al., 2008). Here, we additionally incorporated more holistic and tone-related cues. To this end, we transferred rating methods from established systems of behavioral observation (Back, Schmukle, & Egloff, 2009; Funder, Furr, & Colvin, 2000) onto written material (cf. Back et al., 2008).

Third, we investigated two independent target samples to replicate consensus and accuracy findings. In doing so, we also aimed to tackle the replicability of the lens model findings that are thought to explain the accuracy of personality judgments via the actually observable linguistic cues of short stories (cue validities and cue utilizations). Lens model analyses give us important

insight into the processes that might have explained why personality traits of a given set of targets could or could not be detected. Each specific set of targets might however also include unique patterns of cues, and some of the cue relations might therefore not be generalizable to the other set of targets. Presumably because the measurement and analysis of actually observable cues is a painstaking effort (cf. Back & Egloff, 2009), a replication of lens model results is almost never undertaken (see Borkenau & Liebler, 1993; Kenny, Horner, Kashy, & Chu, 1992; for exceptions in the context of nonverbal behavior and physical appearance). Replicated personality-cue and cue-judgment relations across target samples would, however, be a most valuable extension, as this would substantially strengthen the processual explanations outlined by lens model analyses.

Fourth, to allow for a broader look at the accuracy of personality judgments based on written material, we obtained not only self-report measures of the Big Five, but also included peer-reports and objective measures of general knowledge.

With regard to the specific personality traits for which accurate judgments are achieved based on linguistic style, a differentiated pattern of results might evolve. In particular, we expected openness and general knowledge of targets to be detectable in written short stories. A writing task is an intellectually challenging activity (e.g., What do I write? What words do I use?) that requires the production of something new. Consequently, traits related to originality and intellectual ability such as openness and general knowledge should predict observable aspects of writing style (Beier & Ackermann, 2001; McCrae, 1987; Wolfradt & Pretz, 2001), leading to accurate judgments of targets' openness and knowledge by lay observers. Consequently, openness and general knowledge judgments should be most accurate.

In studies regarding nonverbal behavior or physical appearance (Borkenau & Liebler, 1992; Funder & Sneed, 1993; Kenny, 1994; Levesque & Kenny, 1993; Naumann et al., 2009), extraversion and conscientiousness are the most judgeable traits. By contrast, written short stories might lack the kinds of valid cues (e.g., expressive behavior, stylish or orderly clothing) that are needed for accurate judgments of these traits. Neuroticism and agreeableness are very difficult to detect in zero-acquaintance studies relying on nonverbal behavior or physical appearance (Connolly, Kavanagh, & Viswesvaran, 2007; John & Robins, 1993; Kenny & West, 2008). Interestingly however, there seems to be valid information regarding emotions and social processes within people's self-descriptions and streams-of-thought leading to accurate judgments for neuroticism and agreeableness when self-related content is observed (Holleran & Mehl, 2008). Whether this applies also to the style of writing awaits an empirical investigation.

5. Study 1

5.1. Method

5.1.1. Procedure and target personality

The target sample consisted of 79 (56 female) university students. The average age was 23.34 years (SD = 5.00), ranging from 18 to 45. All targets were rewarded with research participation credit. First, targets completed a short form of the German version of the Big Five Inventory (BFI-S; Schupp & Jean-Yves, 2008)[1] and the multiple-choice vocabulary test as a measure of general linguistic knowledge (Lehrl, 2005). Next, targets had to write a short story that had to include the words "plane crash," "parlormaid," "fireworks," "Middle Ages," and "supermarket" (German originals: Flugzeugabsturz, Zimmermädchen, Feuerwerk, Mittelalter, Supermarkt). The choice of these particular words was based on a group discussion and subsequent ratings of a number of candidate words. To evoke the writing of creative and fictional short stories, those five words were chosen that (a) had a minimal relation to one another and (b) are of little direct thematic self-relation. Targets were given 7 min to write the short story. The word count ranged from 32 to 177 (M = 108.54, SD = 30.60). Short stories were then given to an independent group of observers who individually read the transcribed stories and rated the author of every story regarding his or her personality.

5.1.2. Cue measures

The selection of cues was designed to ensure maximal coverage of the meaningful and observable information at hand, and was inspired by other established systems of behavioral observation (Back et al., 2009; Funder et al., 2000; Rodriguez et al., 2010). Aspects of writing style that were objectively countable were measured by using LIWC (Pennebaker et al., 2001), which applies a word-count strategy searching for the presence of over 6400 words and word stems in any given text file (German dictionary; Wolf et al., 2008). The categories of the dictionary include standard grammatical features (e.g., pronouns, articles), basic psychological processes (e.g., positive and negative emotion words, cognitive mechanism words), and various personal concerns (e.g., achievement, leisure). The following LIWC categories were used in the present research: optimism words, sadness words, anger words, negations, communication words, references to others, and friend-related words.

To obtain more holistic aspects of writing style, we additionally obtained a number of rating cues. Three trained observers rated the stories on how happy, aggressive, warm, creative, meaningfully finished, fluent, textually demanding, linguistically demanding, and erroneous they were. Moreover, they rated the number of foreign words, cuss words, stylistic devices, and the usage of details. Inter-rater reliabilities were satisfactory with a mean alpha of .66 (using fisher's r-to-z formula), ranging from .42 for fluency to .82 for cuss words. Additionally, one rater categorized all stories with respect to the relation of characters within each story (i.e., if they were in any way related or not). Z-standardized cue value scores were used throughout subsequent analyses. To analyze the effects of personality on writing style as well as of writing style on personality judgments on a more meaningful level, we computed the following four cue aggregates across single cues: *sophisticated writing* (cues: finished, erroneous (inverse), linguistically demanding, usage of stylistic devices, fluent, usage of foreign words), *creative expression* (creative, textually demanding, details), *positive emotions in words* (warm, sadness words (inverse), optimism words, happy), and *social orientation in words* (negations (inverse), anger words (inverse), aggressive (inverse), swear words (inverse), no relation (inverse), communication words, reference to others, friend-related words).[2]

[1] Internal consistencies of the BFI-S were as follows: neuroticism = .75; extraversion = .89; openness = .65; agreeableness = .56 and conscientiousness = .75.

[2] Internal consistencies of cue aggregates were .84 for sophisticated writing, .84 for creative expression, .45 for positive emotions in words, and .35 for social orientation in words. Similar results were obtained in Study 2 with internal consistencies of .74, .88, .54, and .43, respectively. In light of the limited number of multiply determined, single behavioral indicators, we regarded these numbers as satisfactory. Additionally, we would like to point out that when analyzing behavioral measures of personality, it is necessary to consider the individual specificity of behavioral responses (see Asendorpf (1988), for a detailed discussion): many behavioral indicators of a trait (e.g., anxiousness) that can be identified at the group level (e.g., talking nervous, nervous hand movements) are not parallel indicators at the individual level (e.g., some anxious individuals may talk nervous, and may not show nervous hand movements whereas other anxious individuals show nervous hand movements, only), resulting in low internal consistencies although the aggregate (e.g., mean of nervous talking and nervous hand movements) is a valid measure of anxious behavior.

5.1.3. Observer ratings

Observers were five female and five male students of different majors who received monetary compensation for their participation in this study. Their ages ranged from 23 to 30 years, with a mean age of 26.10 years (*SD* = 2.18). The observers had to rate each story's author on the Big Five using a third-person version of the BFI-S (Schupp & Jean-Yves, 2008). Additionally, observers rated the targets regarding their general linguistic knowledge ("This person is verbally intelligent") using a 6-point scale from 1 (*not at all*) to 6 (*very much*).

6. Results and discussion

Single and average inter-rater consensus and accuracies of Big Five and knowledge judgments are depicted in Table 1. One can see that observers agreed in their judgments for all Big Five dimensions as well as for general knowledge. Furthermore, observers' judgments of targets' openness, agreeableness, and general knowledge converged partly with the criterion measures. Thus, as expected, the style of writing fictional short stories revealed something about the targets' openness and general knowledge. Accuracies were remarkably strong, particularly for general knowledge, given the short, purely written, and depersonalized information. Interestingly, agreeableness, a trait that is barely judgeable in face-to-face zero-acquaintance contexts (Connolly et al., 2007), could also be judged accurately just on the basis of fictional short stories.

In order to address how accuracy for these judgments was obtained, we applied lens model analyses. As can be seen on the left side of Table 2, openness was positively related to creative expression and positive emotions in words, agreeableness predicted positive emotions as well as social orientation in words, and general knowledge was positively associated with sophisticated writing and creative expression. On the right side of Table 2, one can see that for judging openness, observers made extensive use of the cue aggregates sophisticated writing and creative expression and to a lesser extent positive emotions in words. For judging agreeableness, they used positive emotions in words and social orientation in words. General knowledge was judged using mainly sophisticated writing, creative expression, and positive emotions in words. Altogether, cue validities and cue utilizations suggest why observers were somewhat correct in judging targets' openness agreeableness and general knowledge of targets: For these traits, they drew on cues that provided valid information about the targets' personalities.

7. Study 2

In the second sample, we aimed to replicate the results found in Study 1. That is, we used a second set of targets to find out whether personality judgments based on fictional short stories are (a) consensually shared for all personality dimensions, and (b) accurate for openness, agreeableness, and general knowledge. These are findings that should be explainable by the observable cues of writing style. To strengthen the generalizability of results, we broadened our criterion measures by adding peer ratings of targets' Big Five personality as well as a broader measure of general knowledge.

7.1. Method

The target sample consisted of 126 (74 female) university students. The average age was 22.37 years (*SD* = 3.81), ranging from 19 to 41 years. All targets were rewarded with research participation credit.

Targets completed a short form of the German version of the Big Five Inventory (BFI-10; Rammstedt & John, 2007). Additionally they were asked to complete a 10-min knowledge test, which consisted of 19 general knowledge questions in a multiple-choice format and seven open-ended questions pertaining to more current topics (Back et al., 2009). Additionally, targets were asked to nominate two close peers who know them best to fill out the peer-rating form of the NEO Five Factor Inventory (Borkenau & Ostendorf, 1993). Peer ratings were returned via mail in prepaid envelopes. Aggregates of self- and peer ratings were used as criterion measures for the Big Five.[3] The writing task was as described in Study 1. The word count ranged from 38 to 215 (*M* = 116.92, *SD* = 35.20). The measurement of cues and the procedure for obtaining the personality judgments were as described in Study 1. The observers had to rate each story's author on the Big Five using a third-person version of the BFI-10 (BFI-10; Rammstedt & John, 2007). Additionally, observers rated the targets regarding their general knowledge ("This person has a high general knowledge") using a 6-point scale from 1 (*not at all*) to 6 (*very much*).

8. Results and discussion

Inter-rater consensus and accuracy of observer judgments for targets' self-ratings, peer ratings, and the aggregated personality measure are depicted in Table 3. Again, observers made consensual judgments for all Big Five dimensions as well as for general knowledge. Furthermore, observers were again somewhat able to accurately judge targets' agreeableness, openness, and general knowledge.[4]

Table 1
Consensus and accuracy of personality judgments based on short stories (Study 1).

	Consensus		Accuracy
	Single	Average	
Neuroticism	.26**	.78**	−.07
Extraversion	.25**	.77**	.09
Openness	.28**	.80**	.19*
Agreeableness	.34**	.84**	.31**
Conscientiousness	.23**	.75**	.11
General knowledge	.26**	.74**	.50**

Note. Consensus across 10 observers is shown as mean single-rater intraclass correlation, ICC (2, 1), and average-rater intraclass correlation, ICC (2, *k*). Accuracy is the correlation between observer ratings and targets' self-reports. *N* = 79.
*$p < .05$. **$p < .01$ (one-tailed).

[3] Internal consistencies of the BFI-10 were as follows: neuroticism = .57; extraversion = .66; openness = .57; agreeableness = .02 and conscientiousness = .52. These numbers should be interpreted with caution as the two items of the BFI-10 were explicitly chosen to represent different aspects of each broad Big Five dimension (Rammstedt & John, 2007). As an effect, although test-retest reliabilities of the BFI-10 are satisfactory, internal consistencies are often not (see Gosling, Rentfrow, & Swann (2003), for a similar argument). Clearly, the absence of any overlap for the agreeableness items in the present study is unsatisfactory. Internal consistencies of the peer-reports were .84 (neuroticism), .75 (extraversion), .75 (openness), .81 (agreeableness), and .90 (conscientiousness) for peer 1, and .88 (neuroticism), .83 (extraversion), .77 (openness), .85 (agreeableness), and .91 (conscientiousness) for peer 2. When aggregating the two scores for each dimension across peers, reliabilities were as follows: neuroticism = .57; extraversion = .63; openness = .62; agreeableness = .72 and conscientiousness = .72. Self-ratings and peer ratings of targets' personality were correlated as follows: neuroticism = .31**; extraversion = .59**; openness = .36**; agreeableness = .38**; conscientiousness = .63**.

[4] Given the internal inconsistency of the self-reported agreeableness measure used in Study 2 we additionally calculated accuracies of perceiver ratings separately for the two single agreeableness items of the BFI-10. Results show that when using self-reported agreeableness as a criterion measure, accuracy was mainly driven by the reversed agreeableness item of "I see myself as someone who tends to find fault with others" (item 7 reversed; $r = .29$, $p < .01$) and not by the other agreeableness item (item 2; "I see myself as someone who is generally trusting", $r = -.02$ *ns*).

Table 2
A lens model analysis of personality judgments based on short stories (Study 1).

Cue validity			Cue aggregates (lens)	Cue utilization		
O	A	Knowledge		O	A	Knowledge
.09	−.04	.56**	Sophisticated writing	.71**	−.04	.83**
.25*	−.03	.34**	Creative expression	.82**	−.05	.73**
.26*	.23*	.07	Positive emotions in words	.39**	.28*	.43**
.05	.30**	−.06	Social orientation in words	.02	.47**	.02

Note. O = openness, A = agreeableness. $N = 79$.
*$p < .05$. **$p < .01$ (two-tailed).

Table 3
Consensus and accuracy of personality judgments based on short stories (Study 2).

	Consensus		Accuracy		
	Single	Average	Self	Peer	Aggregate
Neuroticism	.11**	.55**	−.02	−.03	−.02
Extraversion	.20**	.71**	.05	.07	.06
Openness	.20**	.72**	.20*	.17*	.22**
Agreeableness	.32**	.82**	.18*	.15+	.19*
Conscientiousness	.16**	.66**	.08	.03	.08
General knowledge	.14**	.62**	.26**		

Note. Consensus across 10 observers is shown as mean single-rater intraclass correlation, ICC (2, 1), and average-rater intraclass correlation, ICC (2, *k*). Accuracy is the correlation between observer ratings and targets' self-reports. $N = 126$.
+$p < .10$. *$p < .05$. **$p < .01$ (one-tailed).

As in Study 1, we applied lens model analyses to address how accuracy of judgments was obtained by observers. In Table 4, validities and utilizations are depicted for the four cue aggregates. Openness and general knowledge were positively connected to sophisticated writing, creative expression, and positive emotions in words, and agreeableness was positively associated with social orientation in words (see left side of Table 4). On the right side of Table 4, one can again see the utilization writing style cues. Sophisticated writing, creative expression, and positive emotions in words led to judgments of openness and general knowledge; agreeableness judgments were based on the positivity of emotions and the social orientation in words. Again, the pattern of cue validities and cue utilizations nicely shows why observers were somewhat accurate in judging openness, general knowledge, and agreeableness: They correctly used valid cues for their judgments.

9. Combined analyses

The results of the two studies match well. First, observers generally agreed in their personality judgments. Second, judgments were generally correct for openness, agreeableness, and general knowledge, but not for neuroticism, extraversion, and conscientiousness. Third, the same kinds of writing style cues were able to explain these accuracies in both studies: creative expression and positive emotion in words were predicted by targets' openness and in turn these cues predicted openness judgments. Social orientation in words was predicted by targets' agreeableness and in turn it predicted agreeableness judgments. Finally, sophisticated writing as well as creative expression was predicted by the targets' general knowledge and they predicted general knowledge judgments.

Given this remarkably similar pattern of results, we *z*-standardized all measures within each study, combined both samples ($N = 205$), and reran all lens model analyses. We also checked whether cue aggregates mediated the accuracy of personality judgments by performing a multiple mediator analysis (see Kenny, Kashy, & Bolger, 1998). Specifically, we used Mplus (Muthén & Muthén, 2006) to fit a path model to the data in which separate paths were specified from the criterion to the cue aggregates and from the cue aggregates to the personality judgments, respectively, and in which a direct path was specified from the criterion to the personality judgments. Also, we allowed cue aggregates to covary with each other. An analysis of this model, thus, tests simultaneous mediation of all cue aggregates, and it enables to compute the magnitude of the direct effect, the indirect effect of all cue aggregates and the specific indirect effects of each cue aggregate."

Fig. 1 depicts the lens model analysis for openness and its judgments based only on short stories. The overall accuracy for judging openness was $r = .21$. The path analysis revealed that the total effect of the targets' openness on the observer judgments of openness was indeed completely mediated by the cue aggregates (indirect effect = .21, $p < .01$; direct effect = .01, *ns*). The indirect effects for the single-cue aggregates were .02 (*ns*) for sophisticated writing, .16 ($p < .01$) for creative expression, .02 ($p < .10$) for positive emotions in words, and .00 (*ns*) for social orientation in words. This indicates that the cue aggregate of creative expression was most responsible for the observers' ability to judge targets' openness.

Fig. 2 shows the lens model analysis for agreeableness judgments. The overall accuracy for judging agreeableness was $r = .24$. The path analysis showed that cue aggregates partly mediated the influence of targets' agreeableness on the observer judgments of agreeableness (indirect effect = .14, $p < .01$; direct effect = .10, $p < .10$). The indirect effects for the single-cue aggregates were .00 (*ns*) for sophisticated writing, .00 (*ns*) for creative expression, .05 ($p < .10$) for positive emotions in words, and .09 ($p < .01$) for social orientation in words. Thus, it was mainly the higher social orientation in words of agreeable targets that allowed observers to correctly judge them as agreeable.

Table 4
A lens model analysis of personality judgments based on short stories (Study 2).

Cue validity							Cue aggregates (lens)	Cue utilization		
O agg.	O self	O peer	A agg.	A self	A peer	Knowledge		O	A	Knowledge
.29**	.20*	.30**	.02	−.01	.02	.39**	Sophisticated writing	.62**	−.03	.66**
.24**	.18*	.22*	.02	−.02	.05	.31**	Creative expression	.77**	−.04	.69**
.22*	.17+	.15	.07	.15+	−.05	.23*	Positive emotions in words	.47**	.38**	.31**
−.11	−.08	−.11	.21*	.15+	.19*	−.03	Social orientation in words	−.01	.46**	−.07

Note. O = openness, A = agreeableness, agg. = aggregated, self = self-reports, peer = peer reports. $N = 126$.
+$p < .10$. *$p < .05$. **$p < .01$ (two-tailed).

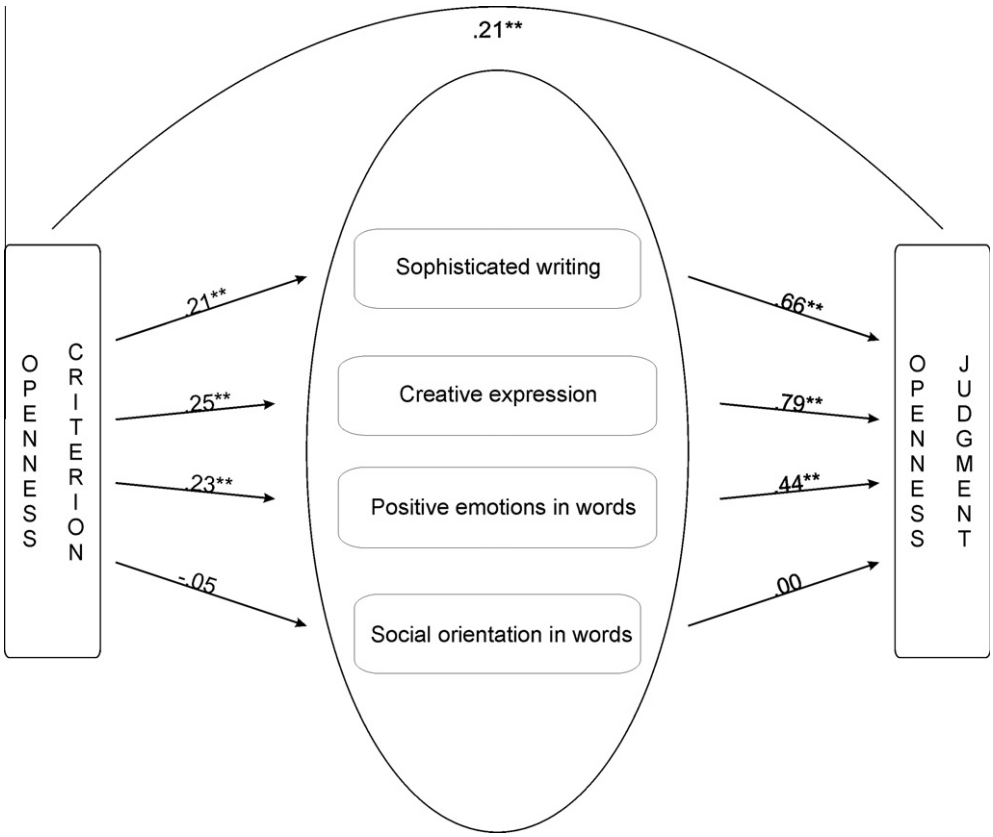

Fig. 1. Lens model analysis of openness judgments based on short stories. The value on the curve reflects the correlations of targets' openness scores and observers' openness judgments. Personality cue relations are depicted by the arrows on the left side, and refer to associations between openness and cue aggregates of short stories. The arrows on the right side show the correlations of these cue aggregates and personality judgments of openness made by observers. $N = 205$. $^{**}p < .01$.

Fig. 3 depicts the lens model analysis for general knowledge and its judgments. The overall accuracy for judging general knowledge was $r = .35$. The path analysis revealed that the total effect of the targets' general knowledge on the observers' judgments of general knowledge was completely mediated by the cue aggregates (indirect effect = .31, $p < .01$; direct effect = .04, *ns*). The indirect effects for the single-cue aggregates were .20 ($p < .01$) for sophisticated writing, .11 ($p < .01$) for creative expression, .01 (*ns*) for positive emotions in words, and .00 (*ns*) for social orientation in words. This indicates that more knowledgeable targets were indeed judged as more knowledgeable because of their sophisticated writing, and to a lesser extent, their creative writing expression. Altogether, these results show nicely that observers were able to accurately judge openness, agreeableness, and general knowledge of targets because they drew on specific and valid cues expressed in the writing styles of short stories.

10. General discussion

The goal of the present study was to show how personality and personality judgments are linked to writing style, and thereby to explain the accuracy of personality judgments based on short stories. To analyze the role of writing style—irrespective of the content of writing—we let targets write fictional short stories based on predefined words. When the authors of the present study first looked over the pages of the written material, three things quickly became obvious: The short stories we had accumulated had a very heterogeneous writing style; when reading them, we were inevitably left with an impression of the personality of the author, and we mostly agreed with regard to our spontaneous impressions.

10.1. Accuracy of personality judgments based on fictional short stories

In accordance with our own spontaneous impressions, we found high inter-rater agreement for all personality dimensions and general knowledge when analyzing lay observer judgments based on the reading of the short stories. More important, however, is the question of accuracy: were personality judgments correct? Prior research on the linguistic expression of self-related content shows quite impressive accuracies of stranger judgments (Holleran & Mehl, 2008; Rodriguez et al., 2010). Here, we concentrated on the style of linguistic expression alone. Our results show that writing style per second is indeed connected to an author's personality, even when the written material is fictional and not directly self-related. There is thus an additional and deeper-rooted relation between a person's writing and personality than just writing about oneself or one's own thoughts.

What are the specific traits that can be inferred from writing style? Research on the accuracy of personality judgments that analyze written material (e.g., self-descriptions, streams-of-thought) typically find some degree of accuracy for all Big Five dimensions (Holleran & Mehl, 2008). Based on linguistic style alone, the present observers were able to accurately judge openness, agreeableness, and general knowledge to a certain degree. In contrast, neuroticism, extraversion, or conscientiousness could not be detected in the targets' writing styles.

10.2. The valid cues of writing style

By using a wide range of word-count based and more holistic cues of writing style, we were able to reveal why judgments of

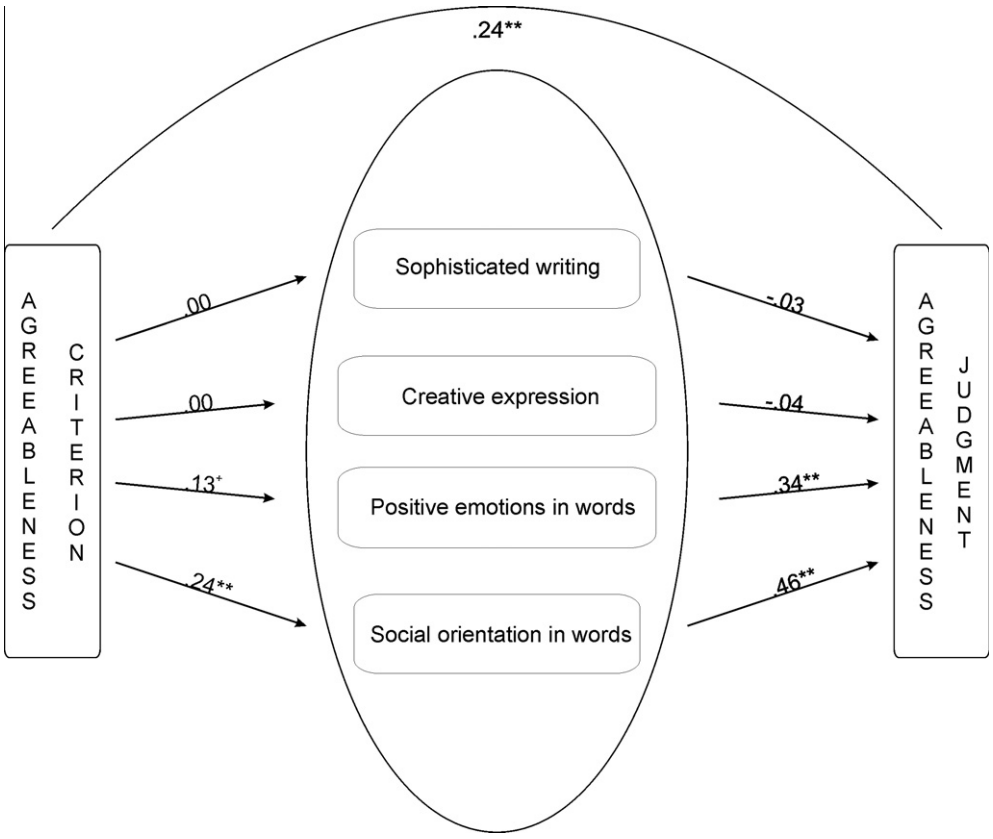

Fig. 2. Lens model analysis of agreeableness judgments based on short stories. The value on the curve reflects the correlations of targets' agreeableness scores and observers' agreeableness judgments. Personality cue relations are depicted by the arrows on the left side, and refer to associations between agreeableness and cue aggregates of short stories. The arrows on the right side show the correlations of those cue aggregates and personality judgments of agreeableness made by observers. $N = 205$. $^{+}p < .10$. $^{**}p < .01$.

openness, agreeableness, and general knowledge were accurate. Lens model analyses showed that (a) there is some valid and observable information for openness, agreeableness, and general knowledge contained in the targets' writing styles, and (b) readers are sensitive to that information such that they detect and correctly use the valid cues that a story transports.

Targets who are high on openness tend to write creative stories probably due to their rather unconventional thinking and rather high powers of imagination. As observers judged the author of a story with a more creative style as more open, accurate openness judgments were achieved. Due to their friendliness, more agreeable targets are prone to write stories in a pro-social style. Observers are also sensitive to a writing style containing a social orientation, and by that, they judged targets' agreeableness correctly. As for general knowledge, its direct influence on language use seems pretty obvious. The variety in a person's vocabulary and knowledge of facts should increase as his or her general knowledge improves. Observers were very sensitive to the validity of complex and sophisticated language use in indicating general knowledge. As a consequence, general knowledge was judged most accurately. By replicating all results across different target samples and different criterion measures (including peer ratings and objective measures of knowledge), we were able to find strong evidence for the generalizability of these findings.

10.3. Implications for future research on personality judgments based on writing

A number of methodological supplements might help to achieve a deeper insight into the processes and moderators of personality perception based on written material. It would, for example, be interesting to compare the accuracies of personality judgments based on a variety of different writing tasks. In doing so, one could try to systematically vary the situational context (e.g., different instructions and cover stories), contextual boundaries (e.g., predefined words or not), and self-focus of creative writing (e.g., writing about a self-related topic or not).

An investigation into the influence of anticipated readers (e.g., confederates vs. persons in charge) is another interesting question. Targets might think of a certain reader in mind when writing, so it would be favorable to additionally measure impression management or self-presentation concerns. These are possible moderator variables of accuracy of personality judgments. Finally, it would be desirable to additionally include actual behavior measures (Back & Egloff, 2009) as criterion measures, a number of additional and more contextualized personality measures, as well as additional observable linguistic patterns as cues.

10.4. The whens and whys of accurate personality judgments at zero-acquaintance

Together with prior zero-acquaintance accuracy research, the present findings provide an interesting and differentiated pattern of results concerning the following question: What kinds of personality traits can be accurately judged given what kinds of information? In terms of moderators of accuracy outlined in Funder's Realistic Accuracy Model (1995, 1999), this question pertains to a trait × information interaction called *diagnosticity*. The available empirical data suggest that diagnosticity is indeed an important moderator of the accuracy of personality judgments at zero-acquaintance: some traits can be judged accurately in some situations, whereas others can be accurately judged in others. Openness seems to manifest in what people tell about themselves, a person's self-related content (e.g., Back, Schmukle et al., 2010; Back, Stopfer

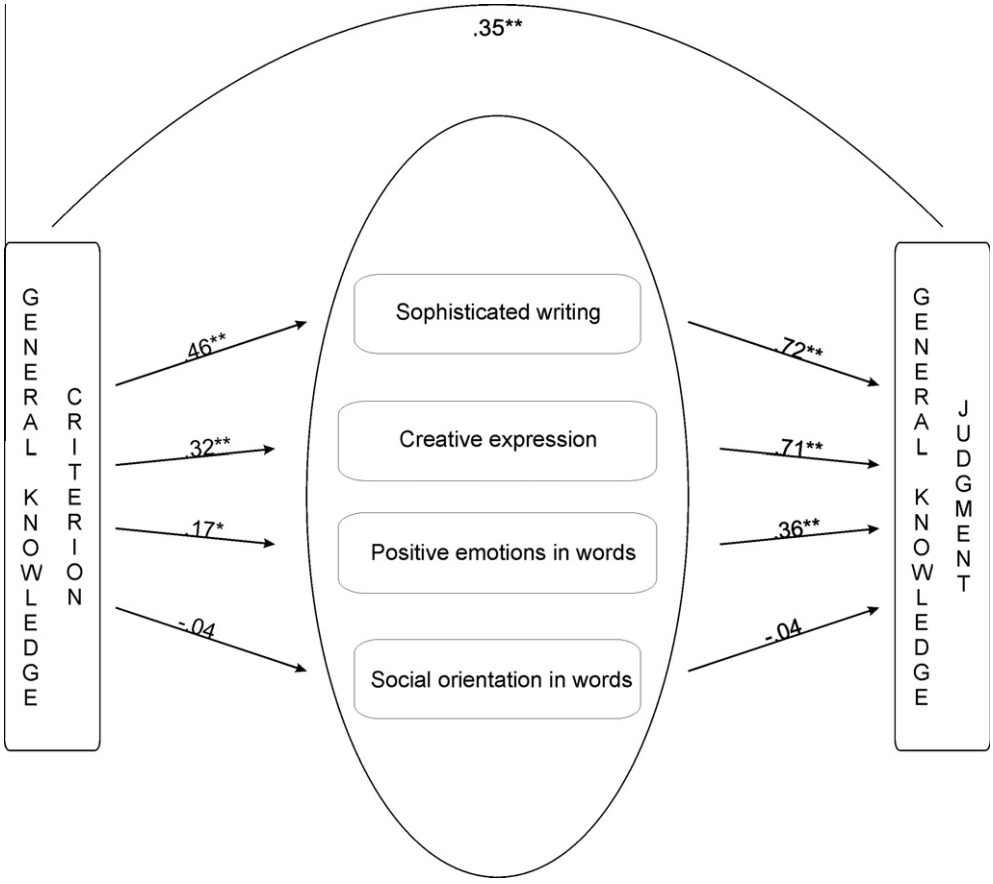

Fig. 3. Lens model analysis of general knowledge judgments based on short stories. The value on the curve reflects the correlations of targets' general knowledge and observers' knowledge judgments. Personality cue relations are depicted by the arrows on the left side, and refer to associations between general knowledge and cue aggregates of short stories. The arrows on the right side show the correlations of these cue aggregates and personality judgments of general knowledge made by observers. $N = 205$. $^{*}p < .05$. $^{**}p < .01$.

et al., 2010; Holleran & Mehl, 2008), as well as in how they tell it (i.e., a person's style of linguistic expression as measured in the present study). The last item is probably due to the intellectual and creativity dependent character of linguistic expression (and other analogous types of creative expression) itself.[5] Accuracy of agreeableness judgments is rather seldom found. In "real-life" zero-acquaintance studies, agreeableness could not be judged accurately (Connolly et al., 2007), whereas in studies including self-connected content, it could (Back, Stopfer et al., 2010; Holleran & Mehl, 2008). What we were able to show here is that agreeableness can be accurately judged even if the information one gets is based on only the style of linguistic expression.

Neuroticism and related traits seem to appear and be shown only in self-related material (Holleran & Mehl, 2008; Rodriguez et al., 2010). Neither in face-to-face contexts (Connolly et al., 2007) nor based on style of linguistic expression (present study) have observers been able to judge neuroticism accurately. Thus, when no self-related content is provided, neuroticism remains somewhat of an unjudgeable personality trait.

Extraversion and conscientiousness are visible, and by this, they are judgeable, either when self-content is provided (Back, Stopfer et al., 2010; Holleran & Mehl, 2008) or in directly observable real-life situations (e.g., Borkenau & Liebler, 1992; Funder & Sneed, 1993; Levesque & Kenny, 1993; Watson, 1989), but not in the style of linguistic expression per se (present study). Or put differently: self-related content (somewhat by definition) contains valid information about each Big Five dimension; appearance (physical appearance and nonverbal behavior) reveals something about extraversion and conscientiousness; and style of expression informs us about openness and agreeableness. It might be worth adopting such a perspective of diagnosticity when summarizing existing research on the accuracy of personality judgments at zero-acquaintance (more fully and systematically than we were able to do here). This should help us to get a better overview concerning what is known in accuracy research and what still needs to be investigated. Additionally, such a perspective might also be of practical relevance because ultimately it allows for predicting the predictability of personality traits in a given situation. Therefore, one would need to have an empirically based overview outlining (a) how much valid information each cue domain (e.g., self-related content, linguistic style of content expression, nonverbal behavior, physical appearance) contains concerning each trait, and (b) what kinds of cue domains are present in a defined social situation. Given this, it should be directly inferable what kind of personality traits can be judged in a given situation.

[5] Openness can also be judged accurately when self-related content is not expressed linguistically, but via the distinctiveness and diversity of leisure preferences such as books, CDs, and so forth (Gosling et al., 2002; Rentfrow & Gosling, 2006). Moreover, openness is revealed when creativity of style does not pertain to linguistic expression, but expression in other kinds of creativity related tasks (e.g., Back et al., 2009; Borkenau, Mauer, Riemann, Spinath, & Angleitner, 2004).

11. Conclusion

The present study examined the relation of personality, fictional short stories, and personality judgments based on the stories. In

sum, we were able to show that writing style per se contains valuable information regarding the targets' personality, particularly the targets' openness, agreeableness, and general knowledge—information that is moreover correctly used by strangers to infer the targets' personality. The differentiated pattern of results is another piece of knowledge to help fill the blank spots on the scientific map of what kinds of personality traits can be accurately judged in what kinds of situations.

Acknowledgment

This research was supported by Grant BA 3731/1-1 from the German Research Foundation (DFG) to Mitja Back, and Grant NE 1485/2-1 from the German Research Foundation (DFG) to Steffen Nestler.

References

Albright, L., Kenny, D. A., & Malloy, T. E. (1988). Consensus in personality judgments at zero acquaintance. *Journal of Personality and Social Psychology, 55*, 387–395.

Asendorpf, J. B. (1988). Individual response profiles in the behavioral assessment of personality. *European Journal of Personality, 2*, 155–167.

Back, M. D., & Egloff, B. (2009). Yes we can! A plea for direct behavioural observation in personality research. *European Journal of Personality, 23*, 403–405.

Back, M. D., Küfner, A. C. P., & Egloff, B. (in press). The emotional timeline of September 11, 2001. *Psychological Science*.

Back, M. D., Schmukle, S. C., & Egloff, B. (2008). How extraverted is honey.bunny77@hotmail.de? Inferring personality from e-mail addresses. *Journal of Research in Personality, 42*, 1116–1122.

Back, M. D., Schmukle, S. C., & Egloff, B. (2009). Predicting actual behavior from the explicit and implicit self-concept of personality. *Journal of Personality and Social Psychology, 97*, 533–548.

Back, M. D., Schmukle, S. C., & Egloff, B. (2010). Why are narcissists so charming at first sight? Decoding the narcissism-popularity link at zero acquaintance. *Journal of Personality and Social Psychology, 98*, 132–145.

Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., et al. (2010). Facebook profiles reflect actual personality not self-idealization. *Psychological Science, 21*, 372–374.

Beier, M. E., & Ackermann, P. L. (2001). Current-events knowledge in adults: An investigation of age, intelligence, and nonability determinants. *Psychology and Aging, 16*, 615–628.

Borkenau, P., Brecke, S., Mottig, C., & Paelecke, M. (2009). Extraversion is accurately perceived after a 50-ms exposure to a face. *Journal of Research in Personality, 43*, 703–706.

Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology, 62*, 645–657.

Borkenau, P., & Liebler, A. (1993). Consensus and self-other agreement for trait inferences from minimal information. *Journal of Personality, 61*, 477–496.

Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology, 86*, 599–614.

Borkenau, P., & Ostendorf, F. (1993). *NEO-Fünf-Faktoren Inventar (NEO-FFI)*. Göttingen: Hogrefe.

Brunswik, E. (1956). *Perception and the representative design of experiments*. Berkeley, CA: University of California Press.

Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science, 15*, 687–693.

Connolly, J. J., Kavanagh, E. J., & Viswesvaran, C. (2007). The convergent validity between self and observer ratings of personality: A meta-analytic review. *International Journal of Selection and Assessment, 15*, 110–117.

Fast, L. A., & Funder, D. C. (2008). Personality as manifest in word use: Correlations with self-report, acquaintance report, and behavior. *Journal of Personality and Social Psychology, 94*, 334–346.

Frisina, P. G., Borod, J. C., & Lepore, S. J. (2004). A meta-analysis of the effects of written emotional disclosure on the health outcomes of clinical populations. *Journal of Nervous and Mental Disease, 192*, 629–634.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review, 102*, 652–670.

Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception*. San Diego, CA: Academic Press.

Funder, D. C., Furr, R. M., & Colvin, C. R. (2000). The riverside behavioral Q-sort: A tool for the description of social behavior. *Journal of Personality, 68*, 451–489.

Funder, D. C., & Sneed, C. D. (1993). Behavioral manifestations of personality: An ecological approach to judgmental accuracy. *Journal of Personality and Social Psychology, 64*, 479–490.

Gifford, R. (1994). A lens-mapping framework for understanding the encoding and decoding of interpersonal dispositions in nonverbal behaviors. *Journal of Personality and Social Psychology, 66*, 398–412.

Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: Personality judgments based on offices and bedrooms. *Journal of Personality and Social Psychology, 82*, 379–398.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. Jr., (2003). A very brief measure of the Big Five personality domains. *Journal of Research in Personality, 37*, 504–528.

Hirsh, J. B., & Peterson, J. B. (2009). Personality and language use in self-narratives. *Journal of Research in Personality, 43*, 524–527.

Holleran, S. E., & Mehl, M. R. (2008). Let me read your mind: Personality judgments based on a person's natural stream of thought. *Journal of Research in Personality, 42*, 747–754.

John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality-traits: The Big 5 domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality, 61*, 521–551.

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford Press.

Kenny, D. A., Horner, C., Kashy, D. A., & Chu, L. (1992). Consensus at zero-acquaintance. Replication, behavioral cues, and stability. *Journal of Personality and Social Psychology, 62*, 88–97.

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp. 233–265). Boston, MA: McGraw-Hill.

Kenny, D. A., & West, T. V. (2008). Zero acquaintance: Definitions, statistical model, findings, and process. In N. Ambady & J. J. Skowronski (Eds.), *First impressions* (pp. 129–146). New York: Guilford.

Lehrl, S. (2005). *Mehrfachwahl-Wortschatz-Intelligenztest MWT-B* (5 ed.). Balingen: Spitta Verlag.

Levesque, M. J., & Kenny, D. A. (1993). Accuracy of behavioral predictions at zero acquaintance: A social relations analysis. *Journal of Personality and Social Psychology, 65*, 1178–1187.

Lumley, M. A. (2004). Alexithymia, emotional disclosure, and health: A program of research. *Journal of Personality, 72*, 1271–1300.

McCrae, R. R. (1987). Creativity, divergent thinking, and openness to experience. *Journal of Personality and Social Psychology, 52*, 1258–1265.

Mehl, M. R. (2006). Quantitative text analysis. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 141–156). Washington, DC: American Psychological Association.

Muthén, L.K., Muthén, B.O., (2006). *Mplus. User's Guidem* (4th ed.) [Software manual]. Los Angeles, CA: Muthén & Muthén.

Naumann, L. P., Vazire, S., Rentfrow, P. J., & Gosling, S. D. (2009). Personality judgments based on physical appearance. *Personality and Social Psychology Bulletin, 35*, 1661–1671.

Pennebaker, J. W. (1997). *Opening up: The healing power of expressing emotions*. New York: Guilford Press.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic Inquiry and Word Count (LIWC): LIWC 2001*. Mahwah, NJ: Erlbaum.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology, 77*, 1296–1312.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54*, 547–577.

Pressman, S. D., & Cohen, S. (2007). Use of social words in autobiographies and longevity. *Psychosomatic Medicine, 69*, 262–269.

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality, 41*, 203–212.

Rentfrow, P. J., & Gosling, S. D. (2006). Message in a ballad: The role of music preferences in interpersonal perception. *Psychological Science, 17*, 236–242.

Schupp, J., & Jean-Yves, G. (Eds.). (2008). BFI-S: Big Five Inventory-SOEP (Vol. 12.00). Bonn: GESIS.

Rodriguez, A. J., Holleran, S. E., & Mehl, M. R. (2010). Reading between the lines: The lay assessment of subclinical depression from written self-descriptions. *Journal of Personality, 78*, 575–598.

Vazire, S. (2010). Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology, 98*, 281–300.

Vazire, S., & Gosling, S. D. (2004). E-perceptions: Personality impressions based on personal websites. *Journal of Personality and Social Psychology, 87*, 123–132.

Watson, D. (1989). Strangers' ratings of the five robust personality factors: Evidence of surprising convergence with self-report. *Journal of Personality and Social Psychology, 57*, 120–128.

Wolf, M., Horn, A. B., Mehl, M. R., Haug, S., Pennebaker, J. W., & Kordy, H. (2008). Computer-aided quantitative text analysis: Equivalence and reliability of the German adaptation of the Linguistic Inquiry and Word Count. *Diagnostica, 54*, 85–98.

Wolfradt, U., & Pretz, J. E. (2001). Individual differences in creativity: Personality, story writing, and hobbies. *European Journal of Personality, 15*, 297–310.

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.