



Controlling for Response Bias in Self-Ratings of Personality: A Comparison of Impression Management Scales and the Overclaiming Technique

Sascha Müller & Morten Moshagen

To cite this article: Sascha Müller & Morten Moshagen (2019) Controlling for Response Bias in Self-Ratings of Personality: A Comparison of Impression Management Scales and the Overclaiming Technique, *Journal of Personality Assessment*, 101:3, 229-236, DOI: [10.1080/00223891.2018.1451870](https://doi.org/10.1080/00223891.2018.1451870)

To link to this article: <https://doi.org/10.1080/00223891.2018.1451870>



© 2018 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 12 Apr 2018.



[Submit your article to this journal](#)



Article views: 4336



[View related articles](#)



[View Crossmark data](#)



Citing articles: 3 [View citing articles](#)

Controlling for Response Bias in Self-Ratings of Personality: A Comparison of Impression Management Scales and the Overclaiming Technique

Sascha Müller and Morten Moshagen

Institute of Psychology and Education, Ulm University, Ulm, Germany

ABSTRACT

Self-serving response distortions pose a threat to the validity of personality scales. A common approach to deal with this issue is to rely on impression management (IM) scales. More recently, the overclaiming technique (OCT) has been proposed as an alternative and arguably superior measure of such biases. In this study ($N = 162$), we tested these approaches in the context of self- and other-ratings using the HEXACO personality inventory. To the extent that the OCT and IM scales can be considered valid measures of response distortions, they are expected to account for inflated self-ratings in particular for those personality dimensions that are prone to socially desirable responding. However, the results show that neither the OCT nor IM account for overly favorable self-ratings. The validity of IM as a measure of response biases was further scrutinized by a substantial correlation with other-rated honesty-humility. As such, this study questions the use of both the OCT and IM to assess self-serving response distortions.

ARTICLE HISTORY



Received 17 June 2017
Revised 20 January 2018

Researchers using personality inventories often face the problem that respondents might distort self-reports to present themselves in a more desirable manner (e.g., Goffin & Boyd, 2009; Hough, 1998). Such self-favoring response distortions pose a threat to the validity of a personality assessment relying on self-reports. A common approach to deal with social desirability bias is the use of impression management (IM) scales (e.g., the respective subscale of the Balanced Inventory of Desirable Responding [BIDR]; Paulhus, 1991). IM scales rest on the idea that individuals either claiming highly desirable (but rare) or discarding highly undesirable (but common) attributes are likely to present themselves overly favorably, so that the extent of response bias can be determined and, in turn, statistically controlled.

However, IM scales have been shown to be substantially related to self-reported personality traits, in particular to emotional stability, conscientiousness, and agreeableness (for a meta-analysis, see Ones, Viswesvaran, & Reiss, 1996). These results might suggest that IM scores seem to be confounded with features of personality; that is, that they do not only represent response distortions (style), but also carry substantive trait-like information (substance). Thus, IM scores by themselves cannot unambiguously indicate whether individuals provide distorted self-descriptions or whether they truly possess virtues. In fact, some have argued that IM scales primarily reflect the latter, such as interpersonally oriented self-control (Uziel, 2010) or honesty (de Vries, Zettler, & Hilbig,

2014; Zettler, Hilbig, Moshagen, & de Vries, 2015). Indeed, Paulhus (2011) more recently recommended against using IM scales to control self-presentation on self-reports of personality and introduced the overclaiming technique (OCT; Paulhus, 2011; Paulhus, Harms, Bruce, & Lysy, 2003) as a superior approach. Nonetheless, IM remains a popular and widely used construct.

The OCT attempts to overcome the inherent problem of IM scales by assessing deviations of self-reports from a known criterion value. As such, this technique is intended as a behavioral method rather than a questionnaire (Paulhus & Holden, 2010). Overclaiming represents the tendency to overstate one's knowledge by claiming to be familiar with factually nonexistent terms (Phillips & Clancy, 1972). Unlike IM scales, the OCT exhibits a less clear pattern of associations with broad dimensions of personality, with the most consistent finding being an association with openness (Tonković, Galić, & Jerneić, 2011), due to which Dunlop et al. (2017) interpreted overclaiming as “a result of knowledge accumulated through a general proclivity for cognitive and aesthetic exploration” (p. 810). Nonetheless, the OCT has been reported to be positively associated with measures of self-presentation; that is, narcissism and IM (Paulhus et al., 2003; Randall & Fernandes, 1991; Tracy, Cheng, Robins, & Trzesniewski, 2009; but see Mesmer-Magnus, Viswesvaran, Deshpande, & Joseph, 2006). In addition, the OCT increased the validity of achievement striving as a facet of Big Five conscientiousness for predicting the grade-point average as a

CONTACT Sascha Müller  sascha.mueller@uni-ulm.de  Institute of Psychology and Education, Ulm University, Albert-Einstein-Allee 47, 89081, ULM Germany

© 2019 the Author(s). Published with license by Taylor & Francis Group, LLC.

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License (<http://www.creativecommons.org/licenses/by-nc-nd/4.0/>) which permits non-commercial use, reproduction and distribution of the work as published without adaptation or alteration, without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

measure of academic performance (Bing, Kluemper, Davison, Taylor, & Novicevic, 2011) and distinguished between faked and genuine interviews about attitudes, behavior, and personality (Kemper & Menold, 2014). However, the OCT does not seem to be capable of predicting social desirability ratings of personality items (Kam, Risavy, & Perunovic, 2015). Overclaiming has also been found to be associated with indicators of careless responding (Barber, Barnes, & Carlson, 2013; Ludeke & Makransky, 2016). Likewise, it was recently shown that overclaiming does not share a common core or nomological network with other measures of socially desirable responding (Bensch, Paulhus, Stankov, & Ziegler, *in press*). In light of these discrepant findings, it should be further examined whether the OCT can be seen as a measure of self-serving response biases.

Whereas the OCT relies on an objective criterion, it is more difficult to obtain such a criterion for personality ratings. One approach is to determine the deviation of self-ratings from other-ratings of personality. Self-other knowledge of personality exhibits obvious asymmetries (Vazire, 2010) with correlations between self- and other-ratings in contemporary models of personality (e.g., the Big Five or the HEXACO model) typically averaging around .50 (Ashton & Lee, 2009; Connelly & Ones, 2010; Connolly, Kavanagh, & Viswesvaran, 2007). These discrepancies between self- and other-reports of personality are due to a number of factors, the most important one arguably being differing informational foundations (Kandler, 2012; Vazire, 2010). Another possible reason for the observed differences in self- and other perception are response distortions in self-ratings due to socially desirable responding (Connolly et al., 2007; Connelly & Ones, 2010; Funder, 1995). Given that observers usually lack motives for deliberate distortion, other-reports should be less prone to response biases. Indeed, evidence suggests that well-acquainted observers are able to provide accurate ratings of personality (e.g., Kolar, Funder, & Colvin, 1996; McCrae & Weiss, 2007) and that other-reports increase the validity in predicting job performance (Mount, Barrick, & Strauss, 1994; Oh, Wang, & Mount, 2011) or features of personality disorders (Miller, Pilkonis, & Clifton, 2005). Thus, other-reports have the potential to expose motivated distortions of self-rated personality (Vazire & Carlson, 2011) and can therefore be used as a testbed to evaluate measures attempting to assess response distortions. If a measure is indeed suitable for controlling self-favoring response biases, it should at least partially account for differences in self- and other-reports about personality.

Existing studies concerning the ability of IM scales to account for the discrepancy between self- and observer reports provided only scattered evidence in favor of IM scales. For example, Piedmont, McCrae, Riemann, and Angleitner (2000) did not find IM scales to account for shared variance between self- and other-reports of personality. Similarly, Borkenau and Ostendorf (1992) neither found moderator nor suppression effects of various IM scales. Borkenau and Zaltauskas (2009) reported that IM did not predict the similarity of self- and other-rated personality profiles. Konstabel, Aavik, and Allik (2006) found

controlling for IM to reduce self-other discrepancies of the Big Five only for neuroticism. In contrast, in the only study so far investigating overclaiming in the context of self- and other-ratings, overclaiming predicted inflated self-reports of a factor primarily made up of extraversion and openness (Paulhus et al., 2003).

As of now, no study has directly compared IM scales and the OCT as a means to account for discrepancies in self- and other-reports. Furthermore, all of the aforementioned studies were based on the Big Five model of personality. However, the honesty-humility factor of the HEXACO model of personality (Ashton & Lee, 2009; Lee & Ashton, 2004) appears to be especially interesting in the context of socially desirable responding, given that it bears close resemblance to IM scales. In fact, IM and honesty-humility frequently exhibit substantially positive associations around $r = .50$ (Dunlop et al., 2017; Zettler et al., 2015). Similar to IM scales where the respondent can either claim or deny the possession of highly desirable or undesirable traits, items of the honesty-humility scale also ask whether one would steal, accept a bribe, or use instrumental flattery (Ashton & Lee, 2009). Thus, self-reported honesty-humility should be particularly prone to self-serving response biases.

The aim of this study was to combine the two approaches of validating self-reports—measures of response bias and other-reports—to examine the validity of purported measures of response distortions. To the extent that the OCT and IM scales can be considered as reflecting response distortions, it is to be expected that both account for deviations between self- and other-ratings in that they are associated with overly favorable self-ratings. In particular, our prime interest lay in those HEXACO dimensions that arguably entail the highest evaluative component, honesty-humility and agreeableness. Whereas the case for honesty-humility seems to be quite obvious, agreeableness appears to be the most evaluative trait in the sense that strong social norms or a high social value is placed on it (Funder, 1995). In line with this reasoning, agreeableness shows the highest self-other discrepancies among the dimensions of the Big Five (Connelly & Ones, 2010). Moreover, when assessed in an application context, individuals assign themselves higher scores than nonapplicants on the HEXACO dimensions of honesty-humility, agreeableness, extraversion, and conscientiousness (Anglim, Morse, de Vries, MacCann, & Marty, 2017).

Method

Participants and procedure

The study has been approved by the local ethics committee and was conducted online in close agreement with contemporary standards of online experimenting (Reips, 2002). Dyad partners took part independently of each other. The first partner forwarded the link to the study's homepage to his or her partner upon completion of the study. Thus, dyad members had no access to their partner's ratings. After providing informed consent, participants completed the overclaiming task, the HEXACO self-report, the BIDR, and the HEXACO observer-report. The study closed with

Table 1. Self-other correlations and intercorrelations of the HEXACO personality factors, impression management (IM), and overclaiming index (OCI).

Factor	<i>M</i>	<i>SD</i>	α	1	2	3	4	5	6	7	8
1. Honesty-Humility	3.37 / 3.38	0.59 / 0.65	.71 / .81	.46*	.14	.01	.34*	-.02	.07	.36*	-.02
2. Emotionality	3.42 / 3.36	0.65 / 0.58	.81 / .80	.05	.55*	-.13	-.10	.13	.02	.11	-.17*
3. Extraversion	3.46 / 3.51	0.57 / 0.57	.80 / .79	.09	-.03	.59*	.12	.01	.07	-.03	.04
4. Agreeableness	3.12 / 3.32	0.53 / 0.63	.73 / .84	.22*	-.21*	.21*	.54*	-.08	.12	.17*	-.11
5. Conscientiousness	3.61 / 3.57	0.58 / 0.70	.81 / .87	-.04	.27*	.06	-.08	.52*	-.14	.05	-.04
6. Openness	3.50 / 3.36	0.59 / 0.59	.74 / .76	.11	-.11	.19*	.10	-.13	.55*	-.07	.05
7. IM	2.79 / —	0.55 / —	.66 / —	.49*	.20*	.05	.24*	.12	-.01	—	—
8. OCI	-.018 / —	0.40 / —	—	-.04	-.19*	-.03	-.03	-.01	.16*	-.17*	—

Note. *N* = 162. *M* = mean (self-ratings / other-ratings); *SD* = standard deviation (self-ratings / other-ratings); α = Cronbach's alpha estimate of internal consistency (self-ratings / other-ratings). Self-other correlations are presented on the diagonal (bold). Intercorrelations between the self-rated (other-rated) factors are presented below (above) the diagonal.

* $p < .05$.

demographic background information and questions regarding the dyad partners' relationship. Note that both IM and overclaiming appear to be stable across administration modes (Dodou & de Winter, 2014; Paulhus & Harms, 2004).

Participants were recruited at a medium-sized German university via mailing lists, social media, and flyers on campus. An a priori power analysis using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) indicated that 150 participants (i.e., 75 dyads) are required to detect a medium-sized effect (corresponding to a standardized regression slope of $\beta = .20$) with a power of .80 given an alpha level of .05. There were no missing data, as participants were technically required to provide a response before proceeding. The drop-out rate was 22.8%. If one dyad member did not complete the study, the dyad was excluded from further analyses. The final sample included 162 individuals forming 81 dyads. Of the participants, 113 (70%) were female and 49 (30%) were male. Age ranged from 19 to 46 ($M = 24.7$, $SD = 3.74$). All participants indicated native or fluent German language skills. Participants received a flat fee of €5 as compensation for completing the study.

The dyads comprised well-acquainted individuals. Mean relationship duration was 57.2 months ($SD = 69.6$). Ninety-one (56%) participants reported being in touch with their dyad partner at least on a daily basis and another 37% at least several times a week. Eighty-three (51%) participants attended the study with a friend; 55 (34%) with their romantic partner; 13 (8%) with a colleague, fellow student, or schoolmate; and 11 (7%) with a relative. Participants indicated on a 10-point scale that they knew their dyad partner rather well ($M = 7.73$, $SD = 1.75$) and that they were quite confident regarding their ratings of the dyad partner's personality ($M = 6.71$, $SD = 1.98$).

Materials

Balanced inventory of desirable responding

The German 20-item version of the BIDR (Musch, Brockhaus, & Bröder, 2002; Paulhus, 1991) assesses self-deceptive enhancement and IM, respectively, with 10 items each. All items were rated on a 5-point Likert scale. Cronbach's alpha of the IM scale was .66, which is in line with previous research (e.g., Borkenau & Zaltauskas, 2009; Musch et al., 2002; Paulhus, 1991).

HEXACO

We used the German version (Moshagen, Hilbig, & Zettler, 2014) of the HEXACO-60 (Ashton & Lee, 2009) to assess the six personality dimensions honesty-humility, emotionality, extraversion, agreeableness, conscientiousness, and openness (10 items each). The measure is available both as a self-report and as an other report form. Items were answered on a 5-point Likert scale. For self-reports, Cronbach's alpha ranged from .71 (honesty-humility) to .81 (emotionality and conscientiousness). For other-reports, internal consistencies ranged from .76 (openness) to .87 (conscientiousness). Thus, internal consistencies were similar to other studies (Ashton & Lee, 2009; Moshagen et al., 2014).

Overclaiming

To assess overclaiming, we used a set of 60 terms spanning three domains (i.e., the arts and philosophy, natural sciences, and social sciences; 20 terms each) allegedly concerning academic and everyday knowledge. Items of each domain comprised 12 targets (factually existing terms; e.g., categorical imperative) and 8 lures (nonexistent terms; e.g., paradox of duplicity). The item set was based on items used in related studies (Dunlop et al., 2017; Musch, Ostapczuk, & Klaiher, 2012).

Item lists were presented in random order, as were the terms within each domain. The measure was introduced as a general knowledge test. Participants were instructed to indicate for each item if they knew the term or not. In most implementations of the OCT, participants are asked to rate their familiarity with each item on a continuous scale. Because the resulting scores are then dichotomized to calculate indexes of response bias (e.g., Paulhus et al., 2003), we opted to directly use a dichotomous scoring procedure.

Results

Descriptive results

As is typical in research comparing self- with other-reports of personality (e.g., Ashton & Lee, 2009; Connelly & Ones, 2010), the correlations between self- and other-ratings were around .50 (see Table 1), ranging from $r = .46$ for honesty-humility to $r = .59$ for extraversion. In the overclaiming task, the mean hit rate was .68 ($SD = .12$) and the mean

Table 2. Linear mixed models predicting impression management (IM)/overclaiming index (OCI) by self- and other-ratings of personality.

Model/estimate	<i>S</i>	<i>O</i>	<i>S + O</i>	<i>S - O</i>	<i>abs</i>	<i>ICC</i>	$\gamma_0(\beta_0)$	$\sigma_{0(\beta_0)}^2$	σ_e^2
Overclaiming									
Honesty-humility	-.04 (.07)	.00 (.05)	-.04 (.06)	-.04 (.11)	.00 (.10)	.20	-0.05 (.18)	.03 (.02)	.13 (.02)*
Emotionality	-.09 (.05) [†]	-.04 (.07)	-.13 (.06)*	-.05 (.10)	-.08 (.13)	.15	0.26 (.21)	.02 (.02)	.13 (.02)*
Extraversion	-.07 (.06)	.07 (.06)	.01 (.05)	-.14 (.11)	.13 (.11)	.20	-0.20 (.18)	.03 (.02)	.13 (.02)*
Agreeableness	.03 (.06)	-.07 (.05)	-.04 (.06)	.10 (.10)	.06 (.13)	.17	-0.06 (.18)	.03 (.02)	.13 (.02)*
Conscientiousness	.01 (.05)	-.02 (.05)	-.02 (.05)	.03 (.09)	.01 (.10)	.19	-0.12 (.20)	.03 (.02)	.13 (.02)*
Openness	.13 (.06)*	-.04 (.06)	.10 (.06) [†]	.17 (.11)	.07 (.13)	.19	-0.53 (.20)*	.03 (.02)	.13 (.02)*
Impression management									
Honesty-humility	.39 (.08)*	.14 (.07)*	.52 (.07)*	.25 (.12)*	-.27 (.13)*	.10	1.03 (.24)*	.03 (.03)	.19 (.04)*
Emotionality	.15 (.07)*	-.02 (.10)	.13 (.10)	.17 (.14)	.04 (.20)	.14	2.33 (.34)*	.04 (.04)	.25 (.05)*
Extraversion	.11 (.08)	-.09 (.08)	.02 (.09)	.20 (.13)	.18 (.16)	.20	2.71 (.32)*	.06 (.03) [†]	.24 (.04)*
Agreeableness	.23 (.08)*	.03 (.07)	.26 (.07)*	.20 (.14)	-.06 (.14)	.18	1.98 (.21)*	.05 (.03)	.23 (.04)*
Conscientiousness	.12 (.09)	-.03 (.07)	.10 (.08)	.15 (.13)	.05 (.13)	.18	2.44 (.30)*	.05 (.03)	.24 (.04)*
Openness	.04 (.07)	-.07 (.08)	-.03 (.08)	.11 (.13)	.08 (.14)	.18	2.89 (.25)*	.05 (.03)	.24 (.04)*

Note. $N = 162$. S = self-rating regression coefficient; O = other rating regression coefficient; $abs = |S - O| - |S + O|$. Presented are the fixed effects for 12 separate models of regressing the measures of response biases (OCI/IM) on the respective self- and other-ratings of the HEXACO dimensions as well as random intercepts and variances. Standard errors are shown in parentheses.

* $p < .05$.

[†] $p < .10$.

false alarm rate was .22 ($SD = .15$). Signal detection theory (Macmillan & Creelman, 2005), based on a loglinear correction of the hit and false alarm rates (Stanislaw & Todorov, 1999), was used to determine the index of discrimination accuracy, $d' = z(H) - z(FA)$, and the location of the criterion, $c = -.5 * [z(H) + z(FA)]$. Because c expresses the tendency to respond “No”, expressed the extent of overclaiming as $OCI = (-1) * c$, so that higher values of the overclaiming index (OCI) correspond to stronger overclaiming. The OCI ranged from -1.24 to 1.04 ($M = -0.18$, $SD = 0.40$), indicating sufficient variance in overclaiming. Participants quite accurately discriminated between targets and lures ($M_{d'} = 1.38$, $SD = 0.46$, range = $0.47-2.4$). Accordingly, OCI and d' were negatively correlated ($r = -.45$, $p < .001$). Similar to overclaiming, the IM scale also exhibited sufficient variation ($M = 2.8$, $SD = 0.55$, range = $1.6-4.2$). Overall, descriptive results concerning IM and hit and false alarm rates (and thus, d' and OCI) were similar to related research (Dunlop et al., 2017; Musch et al., 2012; Paulhus et al., 2003; Zettler et al., 2015).

Unexpectedly, IM and OCI were negatively correlated ($r = -.17$, $p = .027$), so that higher IM was associated with less overclaiming. The OCI showed rather small correlations to the personality dimensions, with the only significant correlations concerning self-rated and other-rated emotionality ($r = -.19$, $p = .017$ and $r = -.17$, $p = .029$) and self-rated openness ($r = .16$, $p = .039$). Similar to previous findings (e.g., Paulhus, 2002), IM was positively correlated with self-rated emotionality ($r = .20$, $p = .012$) and self- and other-rated agreeableness ($r = .24$, $p = .002$ and $r = .17$, $p = .029$). Moreover, IM exhibited positive correlations with both self-rated ($r = .49$, $p < .001$) and other-rated ($r = .36$, $p < .001$) honesty-humility.

Statistical analyses

Given that individuals exhibiting self-favoring response biases inflate their self-ratings to an overly favorable level, other-ratings should offer a more realistic view of these individuals. If IM or the OCT are valid measures of

response biases, they are expected to be positively associated with differences between self- and other-ratings such that a large difference occurs for individuals with a strong tendency to provide self-enhanced responses (and vice versa). To test this assumption, we employed the analysis strategy recently suggested by Humberg et al. (2018). Specifically, we estimated (hierarchical) linear regression models predicting IM or OCI, respectively, by self- and other-ratings of personality separately for each HEXACO factor. A positive regression coefficient of the self-rating in conjunction with a negative coefficient of the other-ratings suggests that self-favoring response distortions are associated with higher self-enhancement (i.e., IM or OCI) scores and, thus, points to the validity of IM and OCI in terms of measuring self-enhancement. To test this difference between these regression coefficients directly, we defined the auxiliary parameter $abs = |b_{self-rating} - b_{other-rating}| - |b_{self-rating} + b_{other-rating}|$ (Humberg et al., 2018). A positive abs parameter indicates that self-other discrepancies are associated with individual differences in IM or OCI scores. In particular, a significantly positive abs parameter in conjunction with ($b_{self-rating} - b_{other-rating}$) being positive indicates that higher self-other discrepancies are associated with higher IM or OCI scores, whereas a positive abs parameter in conjunction with ($b_{self-rating} - b_{other-rating}$) being negative indicates a negative association between IM or OCI scores and self-other discrepancies. If abs is negative, self-other discrepancies are not systematically related to IM or OCI scores. Addressing the dyadic structure in our data, we computed mixed regression models using *Mplus 7* (Muthén & Muthén, 1998–2012). A random intercept was included to allow for variation between dyads. ICCs indicated that a multilevel approach was required (see Table 2).

Testing for self-favoring response biases

The regression results are summarized in Table 2. Neither IM nor the OCI were associated with larger discrepancies between self- and other-ratings as there was no significantly positive abs parameter. Thus, none of the purported

measures of self-favoring response biases could account for such. Apart from that, there was a significantly positive main effect of self-rated openness on overclaiming. Concerning IM, self- and other-rated honesty-humility coefficients were both significantly positive.

Because the dyads were composed of individuals having quite different kinds of relationships, relationship characteristics might have had a confounding influence on target ratings. To control for this, we also ran all the regression models using the various relationship variables (i.e., relationship length, relationship type, contact frequency, contact type, and perceived closeness) as potential moderator variables. However, none of the relationship variables acted as a moderator.

Testing for suppression effects

Following related research (Borkenau & Zaltauskas, 2009; Piedmont et al., 2000), we also tested if IM or OCI act as suppressors; that is, whether the correlation between self- and other-ratings increases when controlling for IM or OCI. To this end, we compared the slopes regressing self- on other-ratings with the slopes regressing the self-ratings residualized for IM and OCI, respectively, on the other-ratings. The regression coefficients using the residualized self-ratings were virtually identical or even significantly smaller (see Appendix A), thereby disconfirming the presence of suppressor effects for any of the HEXACO factors. We also compared the slopes regressing other-ratings on self-ratings with the slopes regressing other-ratings on the residualized self-ratings. Likewise, there were no suppressor effects (see Appendix B for details).

Discussion

IM scales and the OCT share the purpose of measuring self-favoring response biases. Given the present criticism of such measures (Kam et al., 2015; Uziel, 2010; Zettler et al., 2015), we argued that a valid measure of self-favoring response bias should account for discrepancies between self- and other-ratings of personality. To test this assumption, we let well-acquainted dyads rate both their own and their dyad partner's personality. It was expected that measures of response distortions are associated with discrepancies between self- and other-ratings of personality, as they should account for inflated self-ratings in particular for those personality dimensions that are prone to socially desirable responding.

However, we found neither IM nor OCI to account for self-other discrepancies and thus self-enhancement effects not being associated with IM or OCI, respectively. In addition, other parts of the results further question whether IM scales can be interpreted as a measure of response biases. In particular, IM exhibited substantial associations with both self- and other-ratings of honesty-humility. If the positive association between IM and self-reported honesty-humility is interpreted as reflecting that individuals merely claim to possess this positively valued attribute, the respective other-

ratings of honesty-humility should be unrelated (or even negatively related) to self-reported IM. Thus, the results rather suggest that individuals high in IM actually possessed high trait levels of honesty-humility (de Vries et al., 2014; Zettler et al., 2015).

Even though overclaiming occurred in our sample, the OCI failed to account for self-other discrepancies for any considered personality dimension. Thus, the results are in line with research suggesting that overclaiming does not seem to be an adequate measure of self-favoring response bias (Dunlop et al., 2017; Kam et al., 2015). Instead, overclaiming was positively associated with self-rated openness. This rather fits the cognitive exploration account of overclaiming (Dunlop et al., 2017), which suggests that overclaiming might reflect a dispositional tendency for cognitive and aesthetic exploration. Consistent with this account, Dunlop et al. (2017) found that the residuals of self-rated openness controlled for the respective other-ratings were correlated with overclaiming. Correspondingly, Atir, Rosenzweig, and Dunning (2015) observed a link between self-perceived knowledge in a domain and overclaiming in the same domain.

In tandem, the results suggest that neither IM nor OCI can be seen to measure self-serving biases. Moreover, the rather surprising negative association between IM and OCI indicates that these measures could even reflect psychological variables that work in opposite directions. Whereas some previous studies found no relationship between IM and OCI (Dunlop et al., 2017; Mesmer-Magnus et al., 2006), we are not aware of any study showing negative associations. Frankly, the reasons are unclear. Although IM might be regarded as reflecting honesty, the OCI itself was unrelated to this factor and failed to account for self-enhancement. Thus, this pattern of associations should be reexamined in further research.

This study is subject to some limitations. Most important, the approach used in this study to evaluate measures of response biases rests on the assumption that differences between self- and other-reports are at least in part due to self-favoring response biases (e.g., Borkenau & Ostendorf, 1992; Piedmont et al., 2000). Clearly, differences also occur for other reasons, most important as result of different informational foundations (e.g., Connelly & Ones, 2010; Funder, 1995; McAbee & Connelly, 2016; Vazire, 2010). Further, individuals who show strong self-presentation tendencies in an assessment might regularly present themselves in a more positive light toward their dyad partners (Nezlek, Schütz, & Sellin, 2007), so that other-ratings are inflated to a similar extent as the corresponding self-rating (albeit to the best of others' knowledge and beliefs). Indeed, it has been suggested that overlap between self- and other-ratings is, in fact, a result of successful impression management influencing the other-ratings especially for traits with lower visibility such as neuroticism or agreeableness (Connelly & Ones, 2010; Danay & Ziegler, 2011). It should further be noted that other-ratings can also be affected by response biases. For example, other-ratings of personality are influenced by the extent of liking for the target (Leising, Erbs, &

Fritz, 2010; Leising, Ostrovski, & Zimmermann, 2013), so that the rather close relationships between the dyad partners in this study could have led to more socially desirable ratings as well, even though none of the relationship variables acted as moderators. Moreover, we deliberately chose not to rely on a high-demand situation (e.g., an applicant setting) to elicit socially desirable responding. Paulhus (2011) argued that overclaiming reflects response biases when an audience is salient. Although participants were aware of the fact that they would also be rated by their dyad partner, so that an audience might have been implicitly present in the form of their partner, this situation differs from a public assessment. Nevertheless, there was sufficient variation in both IM and overclaiming in our sample, so that some participants evidently did exhibit the two behaviors. Finally, our overclaiming measure relied on dichotomous knowledge ratings of targets and lures that differed from the Likert-type familiarity ratings (which are later dichotomized) as employed, for example, by Paulhus et al. (2003). It might be argued that different scoring procedures elicit different answering processes. However, our item set was based on items that were previously either used with continuous familiarity ratings (Dunlop et al., 2017) or dichotomous knowledge ratings (Musch et al., 2012). Considering that we observed a mean false alarm rate very similar to the ones reported by Paulhus et al. (2003) or Dunlop et al. (2017), the scoring procedure does not seem to have a substantial influence on individuals' decision processes. In line with this reasoning, Atir, Rosenzweig, and Dunning (2015) directly compared using knowledge ratings and familiarity ratings and found no difference between the approaches.

Conclusion

In conclusion, this study indicates that neither IM nor overclaiming should be used as measures of self-favoring response distortions, given that both failed to conclusively account for differences between self- and other-ratings of personality as an indicator of self-enhancement. In contrast, the results are aligned with a cognitive exploration account of overclaiming and with the ideas of considering IM as reflecting interpersonally oriented self-control or honesty rather than a self-serving response bias.

Acknowledgments

The data are available in the Open Science Framework at <https://osf.io/mbje5/>. This research was not preregistered.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft (MO 2830/1-1).

References

Anglim, J., Morse, G., de Vries, R. E., MacCann, C., & Marty, A. (2017). Comparing job applicants to non-applicants using an item-

- level bifactor model on the HEXACO Personality Inventory. *European Journal of Personality*, 31, 669–684. doi:10.1002/per.2120
- Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment*, 91, 340–345. doi:10.1080/00223890902935878
- Atir, S., Rosenzweig, E., & Dunning, D. (2015). When knowledge knows no bounds: Self-perceived expertise predicts claims of impossible knowledge. *Psychological Science*, 26, 1295–1303. doi:10.1177/0956797615588195
- Barber, L. K., Barnes, C. M., & Carlson, K. D. (2013). Random and systematic error effects of insomnia on survey behavior. *Organizational Research Methods*, 16, 616–649. doi:10.1177/1094428113493120
- Bensch, D., Paulhus, D. L., Stankov, L., & Ziegler, M. (in press). Teasing apart overclaiming, overconfidence, and socially desirable responding. *Assessment*. Advance online publication. doi:10.1177/1073191117700268
- Bing, M. N., Klumper, D., Davison, H. K., Taylor, S., & Novicevic, M. (2011). Overclaiming as a measure of faking. *Organizational Behavior and Human Decision Processes*, 116, 148–162. doi:10.1016/j.obhdp.2011.05.006
- Borkenau, P., & Ostendorf, F. (1992). Social desirability scales as moderator and suppressor variables. *European Journal of Personality*, 6, 199–214. doi:10.1002/per.2410060303
- Borkenau, P., & Zaltauskas, K. (2009). Effects of self-enhancement on agreement on personality profiles. *European Journal of Personality*, 23, 107–123. doi:10.1002/per.707
- 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. doi:10.1111/j.1468-2389.2007.00371.x
- Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, 136, 1092–1122. doi:10.1037/a0021212
- Danay, E., & Ziegler, M. (2011). Is there really a single factor of personality? A multitrait approach to the apex of personality. *Journal of Research in Personality*, 45, 560–567. doi:10.1016/j.jrp.2011.07.003
- de Vries, R. E., Zettler, I., & Hilbig, B. E. (2014). Rethinking trait conceptions of social desirability scales: Impression management as an expression of honesty-humility. *Assessment*, 21, 286–299. doi:10.1177/1073191113504619
- Dodou, D., & de Winter, J. (2014). Social desirability is the same in offline, online, and paper surveys: A meta-analysis. *Computers in Human Behavior*, 36, 487–495. doi:10.1016/j.chb.2014.04.005
- Dunlop, P. D., Bourdage, J. S., de Vries, R. E., Hilbig, B. E., Zettler, I., & Ludeke, S. G. (2017). Openness to (reporting) experiences that one never had: Overclaiming as an outcome of the knowledge accumulated through a proclivity for cognitive and aesthetic exploration. *Journal of Personality and Social Psychology*, 113, 810–834. doi:10.1037/pspp0000110
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191. doi:10.3758/BF03193146
- Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, 102, 652–670. doi:10.1037/0033-295X.102.4.652
- Goffin, R. D., & Boyd, A. C. (2009). Faking and personality assessment in personnel selection: Advancing models of faking. *Canadian Psychology/Psychologie Canadienne*, 50, 151–160. doi:10.1037/a0015946
- Hough, L. M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance*, 11, 209–244. doi:10.1080/08959285.1998.9668032
- Humberg, S., Dufner, M., Schönbrodt, F. D., Geukes, K., Hutteman, R., van Zalk, M. H. W., ... Back, M. D. (2018). Enhanced versus simply positive: A new condition-based regression analysis to disentangle effects of self-enhancement from effects of positivity of self-view. *Journal of Personality and Social Psychology* 114, 303–322. doi:10.1037/pspp0000134

- Kam, C., Risavy, S. D., & Perunovic, W. E. (2015). Using over-claiming technique to probe social desirability ratings of personality items: A validity examination. *Personality and Individual Differences, 74*, 177–181. doi:10.1016/j.paid.2014.10.017
- Kandler, C. (2012). Knowing your personality is knowing its nature: The role of information accuracy of peer assessments for heritability estimates of temperamental and personality traits. *Personality and Individual Differences, 53*, 387–392. doi:10.1016/j.paid.2012.01.004
- Kemper, C. J., & Menold, N. (2014). Nuisance or remedy?: The utility of stylistic responding as an indicator of data fabrication in surveys. *Methodology, 10*(3), 92–99. doi:10.1027/1614-2241/a000078
- Kolar, D. W., Funder, D. C., & Colvin, C. R. (1996). Comparing the accuracy of personality judgements by the self and knowledgeable others. *Journal of Personality, 64*, 311–337. doi:10.1111/j.1467-6494.1996.tb00513.x
- Konstabel, K., Aavik, T., & Allik, J. (2006). Social desirability and consensual validity of personality traits. *European Journal of Personality, 20*, 549–566. doi:10.1002/per.593
- Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO Personality Inventory. *Multivariate Behavioral Research, 39*, 329–358. doi:10.1207/s15327906mbr3902_8
- Leising, D., Erbs, J., & Fritz, U. (2010). The letter of recommendation effect in informant ratings of personality. *Journal of Personality and Social Psychology, 98*, 668–682. doi:10.1037/a0018771
- Leising, D., Ostrovski, O., & Zimmermann, J. (2013). “Are we talking about the same person here?” Interrater agreement in judgments of personality varies dramatically with how much the perceivers like the targets. *Social Psychological and Personality Science, 4*, 468–474. doi:10.1177/1948550612462414
- Ludeke, S. G., & Makransky, G. (2016). Does the over-claiming questionnaire measure overclaiming? Absent convergent validity in a large community sample. *Psychological Assessment, 28*, 765–774. doi:10.1037/pas0000211
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.
- McAbee, S. T., & Connelly, B. S. (2016). A multi-rater framework for studying personality: The trait-reputation-identity model. *Psychological Review, 123*, 569–591. doi:10.1037/rev0000035
- McCrae, R. R., & Weiss, A. (2007). Observer ratings of personality. In R. W. Robins, R. C. Fraley, & R. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 259–272). New York, NY: Guilford.
- Mesmer-Magnus, J., Viswesvaran, C., Deshpande, S., & Joseph, J. (2006). Social desirability: The role of over-claiming, self-esteem, and emotional intelligence. *Psychology Science, 48*, 336–356.
- Miller, J. D., Pilkonis, P. A., & Clifton, A. (2005). Self- and other-reports of traits from the five-factor model: Relations to personality disorder. *Journal of Personality Disorders, 19*, 400–419. doi:10.1521/pepi.2005.19.4.400
- Moshagen, M., Hilbig, B. E., & Zettler, I. (2014). Faktorenstruktur, psychometrische Eigenschaften und Messinvarianz der deutschsprachigen Version des 60-Item HEXACO Persönlichkeitsinventars [Factor structure, psychometric properties, and measurement invariance of the German-language version of the 60-item HEXACO personality inventory]. *Diagnostica, 60*(2), 86–97. doi:10.1026/0012-1924/a000112
- Mount, M. K., Barrick, M. R., & Strauss, J. P. (1994). Validity of observer ratings of the Big Five personality factors. *Journal of Applied Psychology, 79*, 272–280. doi:10.1037/0021-9010.79.2.272
- Musch, J., Brockhaus, R., & Bröder, A. (2002). Ein Inventar zur Erfassung von zwei Faktoren sozialer Erwünschtheit [An inventory assessing two factors of socially desirable responding]. *Diagnostica, 48*(3), 121–129. doi:10.1026/0012-1924.48.3.121
- Musch, J., Ostapczuk, M., & Klaiher, Y. (2012). Validating an inventory for the assessment of egoistic bias and moralistic bias as two separable components of social desirability. *Journal of Personality Assessment, 94*, 620–629. doi:10.1080/00223891.2012.672505
- Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Nezlek, J. B., Schütz, A., & Sellin, I. (2007). Self-presentational success in daily social interaction. *Self and Identity, 6*, 361–379. doi:10.1080/15298860600979997
- Oh, I.-S., Wang, G., & Mount, M. K. (2011). Validity of observer ratings of the five-factor model of personality traits: A meta-analysis. *Journal of Applied Psychology, 96*, 762–773. doi:10.1037/a0021832
- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*, 660–679. doi:10.1037/0021-9010.81.6.660
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of social psychological attitudes: Measures of personality and social psychological attitudes* (pp. 17–59). San Diego, CA: Academic Press.
- Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49–69). Mahwah, NJ: Erlbaum.
- Paulhus, D. L. (2011). Overclaiming on personality questionnaires. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 151–164). New York, NY: Oxford University Press.
- Paulhus, D. L., & Harms, P. D. (2004). Measuring cognitive ability with the overclaiming technique. *Intelligence, 32*, 297–314. doi:10.1016/j.intell.2004.02.001
- Paulhus, D. L., Harms, P. D., Bruce, M. N., & Lysy, D. C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of Personality and Social Psychology, 84*, 890–904. doi:10.1037/0022-3514.84.4.890
- Paulhus, D. L., & Holden, R. R. (2010). Measuring self-enhancement: From self-report to concrete behavior. In C. R. Agnew (Ed.), *Then a miracle occurs: Focusing on behavior in social psychological theory and research. Purdue Symposium on Psychological Sciences* (pp. 227–246). Oxford, UK: Oxford University Press.
- Phillips, D. L., & Clancy, K. J. (1972). Some effects of “social desirability” in survey studies. *American Journal of Sociology, 77*, 921–940. doi:10.1086/225231
- Piedmont, R. L., McCrae, R. R., Riemann, R., & Angleitner, A. (2000). On the invalidity of validity scales: Evidence from self-reports and observer ratings in volunteer samples. *Journal of Personality and Social Psychology, 78*, 582–593. doi:10.1037/0022-3514.78.3.582
- Randall, D. M., & Fernandes, M. F. (1991). The social desirability response bias in ethics research. *Journal of Business Ethics, 10*, 805–817. doi:10.1007/BF00383696
- Reips, U.-D. (2002). Standards for Internet-based experimenting. *Experimental Psychology, 49*, 243–256. doi:10.1026/1618-3169.49.4.243
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*, 137–149. doi:10.3758/BF03207704
- Tonković, M., Galić, Z., & Jernei, Ž. (2011). The construct validity of over-claiming as a measure of egoistic enhancement. *Review of Psychology, 18*(1), 13–21.
- Tracy, J. L., Cheng, J. T., Robins, R. W., & Trzesniewski, K. H. (2009). Authentic and hubristic pride: The affective core of self-esteem and narcissism. *Self and Identity, 8*, 196–213. doi:10.1080/15298860802505053
- Uziel, L. (2010). Rethinking social desirability scales: From impression management to interpersonally oriented self-control. *Perspectives on Psychological Science, 5*, 243–262. doi:10.1177/1745691610369465
- 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. doi:10.1037/a0017908
- Vazire, S., & Carlson, E. N. (2011). Others sometimes know us better than we know ourselves. *Current Directions in Psychological Science, 20*, 104–108. doi:10.1177/0963721411402478
- Zettler, I., Hilbig, B. E., Moshagen, M., & de Vries, R. E. (2015). Dishonest responding or true virtue?: A behavioral test of impression management. *Personality and Individual Differences, 81*, 107–111. doi:10.1016/j.paid.2014.10.007

Appendix A. Suppression analysis: Fixed effects of linear mixed models

Fixed effect	H _{self}	E _{self}	X _{self}	A _{self}	C _{self}	O _{self}
Uncontrolled self-ratings						
Other rating	.46 (.07)	.55 (.07)	.59 (.06)	.54 (.07)	.50 (.07)	.55 (.07)
Self-ratings controlled for OCI						
Other rating	.41 (.07)	.53 (.07)	.55 (.07)	.53 (.07)	.52 (.07)	.42 (.07)*
Self-ratings controlled for IM						
Other rating	.30 (.08)*	.54 (.07)	.55 (.07)	.51 (.07)	.52 (.07)	.43 (.07)*

Note. $N = 162$. H_{self} = self-rated honesty-humility; E_{self} = self-rated emotionality; X_{self} = self-rated extraversion; A_{self} = self-rated agreeableness; C_{self} = self-rated conscientiousness; O_{self} = self-rated openness; OCI = overclaiming index; IM = impression management. Uncontrolled self-ratings show the (fixed) slopes for the models regressing the self-rated HEXACO personality factors on the respective other-rated personality factors. Controlled self-ratings show the (fixed) slopes for the models regressing the resulting residualized self-ratings (controlling for OCI and IM, respectively) on the other-ratings. All coefficients are standardized, standard errors in parentheses.

*Estimates differ significantly from the respective uncontrolled estimate at $p < .05$.

Appendix B. Suppression analyses predicting other-ratings by self-report data

Fixed effect	H _{other}	E _{other}	X _{other}	A _{other}	C _{other}	O _{other}
Uncontrolled Self-ratings						
Self-rating	.46 (.07)	.55 (.07)	.59 (.06)	.53 (.07)	.52 (.07)	.55 (.07)
Self-ratings controlled for OCI						
Self-rating	.41 (.07)	.53 (.07)	.55 (.07)	.52 (.07)	.52 (.07)	.42 (.07)*
Self-ratings controlled for IM						
Self-rating	.30 (.08)*	.54 (.07)	.55 (.07)	.50 (.07)	.52 (.07)	.43 (.07)*

Note. $N = 162$. H_{other} = other-rated honesty-humility; E_{other} = other-rated emotionality; X_{other} = other-rated extraversion; A_{other} = other-rated agreeableness; C_{other} = other-rated conscientiousness; O_{other} = other-rated openness; OCI = overclaiming index; IM = impression management. Uncontrolled self-ratings show the (fixed) slopes for the linear mixed models regressing the other-rated HEXACO personality factors on the respective self-rated personality factors. Controlled self-ratings show the (fixed) slopes for the linear mixed models regressing the other-ratings on the resulting residualized self-ratings (controlling for OCI and IM, respectively). All coefficients are standardized, standard errors in parentheses.

*Estimates differ significantly from the respective uncontrolled estimate at $p < .05$.