Personality and Individual Differences 81 (2015) 107–111

Contents lists available at ScienceDirect

Personality and Individual Differences

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

Dishonest responding or true virtue? A behavioral test of impression management ☆,☆☆

Ingo Zettler [a,*], Benjamin E. Hilbig [b,c], Morten Moshagen [d], Reinout E. de Vries [e]

[a] University of Copenhagen, Denmark
[b] University of Mannheim, Germany
[c] Max Planck Institute for Research on Collective Goods, Germany
[d] University of Kassel, Germany
[e] VU University Amsterdam, Netherlands

ARTICLE INFO

Article history:
Received 14 April 2014
Received in revised form 12 September 2014
Accepted 1 October 2014
Available online 27 October 2014

Keywords:
Cheating
Honesty
Impression management
Social desirability
Socially desirable responding

ABSTRACT

Impression management or social desirability scales have been used widely to assess and control for self-favoring biases in self-reports, both in low and high demand situations. Recently, however, substantive interpretations of impression management scores have surfaced, including the simple but troubling proposition that high scores in impression management scales actually reflect honesty rather than dishonest responding. In line with findings indicating that respondents answer to personality questionnaires rather accurately in typical low demand situations, we herein suggest that high impression management scores indeed reflect true virtues rather than dishonesty under such conditions. We found support for this idea by replicating previous correlations between impression management scores and virtue-related basic personality traits (including honesty–humility), and additionally provided conclusive behavioral evidence: We linked scores on an impression management scale administered under typical low demand condition to behavior in an incentivized, anonymous cheating task. The results clearly indicate that low scores in impression management are associated with more cheating. That is, high – and not low – scores on the impression management scale of the Balanced Inventory of Desirable Responding are aligned with more virtuous, honest behavior.

1. Introduction

On countless occasions, researchers across disciplines and practitioners from various applied fields rely on self-report questionnaires to assess peoples' traits, states, thoughts, emotions, and behavior. Whereas this practice is generally well-accepted, doubts remain regarding an unconditional interpretation of questionnaire-based test scores. Among other things, it has been argued that socially desirable responding affects self-reports in questionnaires, and, in turn, the validity of their interpretation (e.g. Ben-Porath, 2013; Paulhus, 1991). For instance, this implies that people who tend to present themselves in an overly positive light may receive higher test scores concerning positively connoted constructs (e.g. conscientiousness) – not only as compared to their "true" level, but also as compared to others who try to provide accurate self-descriptions and who actually have similar or even higher levels of the respective construct (e.g. more conscientiousness). Consequently, it has been suggested to assess and ultimately control for socially desirable responding, defined as "the tendency to give overly positive self-descriptions" (Paulhus, 2002, p. 50).

One early and straightforward approach has revolved around the use of scales intended to measure such response tendencies, i.e. self-favoring biases in self-reports. Among these scales, that have been (and are still sometimes) referred to as impression management, lie, or social desirability scales, the Balanced Inventory of Desirable Responding (BIDR, later labeled Paulhus Deception Scales; Paulhus, 2002) has been used frequently. The key idea behind the BIDR is that there are two different types of socially desirable responding, captured by two subscales, named Self-Deceptive Enhancement (SDE) and Impression Management (IM). Originally, SDE was understood to capture rather unconscious self-distortions, whereas IM was understood to capture more conscious self-distortions. However, several research findings have suggested that this distinction (unconscious vs. conscious) was premature as, for instance, SDE test scores were also susceptible

☆ The work reported here in was supported by grants from the German Research Foundation (DFG) to the first (ZE 948/1-1) and second (HI 1600/1-1) author.
☆☆ This article is a Special issue article – "Young researcher award 2014".
* Corresponding author at: Department of Psychology, University of Copenhagen, Øster Farimagsgade 2A, 1353 København K, Denmark. Tel.: +45 35324850.
E-mail address: ingo.zettler@psy.ku.dk (I. Zettler).

http://dx.doi.org/10.1016/j.paid.2014.10.007
0191-8869/

to faking instructions and, in turn, cannot be considered a result of unconscious processes alone (e.g. Pauls & Crost, 2004). Also, there are substantial correlations between SDE and IM, respectively, and personality traits (e.g. Lönnqvist, Paunonen, Tuulio-Henriksson, Lönnqvist, & Verkasalo, 2007). In light of these and other findings, the interpretation of SDE and IM has been adapted over the years. For instance, Paulhus (2002) related SDE to an egoistic bias, particularly referring to "a self-deceptive tendency to exaggerate one's social and intellectual status" (p. 63), and IM to a moralistic bias, particularly referring to "a self-deceptive tendency to deny socially-deviant impulses and claim sanctimonious, "saint-like" attributes" (p. 64). However, despite this more nuanced view on SDE and IM suggesting that both can be considered as a tendency to exaggerate one's responses in a desirable manner, many researchers and practitioners have focused exclusively on IM (and similar questionnaires) when attempting to control for socially desirable responding.

In IM and similar measures, respondents self-report on items that refer either to presumably rare but socially desirable attributes (e.g. "I never swear.") or to (reverse coded) presumably frequent but socially undesirable attributes (e.g. "I sometimes litter."). The key idea behind these scales is that respondents motivated to present themselves in a better-than-justified light will agree to the former and disagree with the latter. In other words, it is assumed that respondents with higher scores presumably underclaim undesirable and overclaim desirable attributes. Correspondingly, higher scores have been suggested to reflect a more pronounced tendency towards socially desirable responding. In line with this idea, researchers and practitioners have treated high scores in IM and similar measures as problematic. In particular, it has been implied that once an individual has high scores in such measures, scores on any other self-report measure (e.g. assessing conscientiousness) may also be biased by socially desirable responding.

IM and similar measures have been used widely both in high demand situations (such as personnel selection, cf. Goffin & Christiansen, 2003) and low demand situations such as personality research. One prominent example of the latter is the development of new instruments for which it has been a common approach to exclude items that correlate positively with IM or a similar measure (cf. Kam, 2013). This approach is assumed to rid one's instrument of items that are strongly affected by socially desirable responding. In turn, after exclusion, the final instrument is presumed to be less susceptible for socially desirable responding overall. Clearly, this approach can have fundamental consequences for the final instrument. For example, Stokes and Cooper (2001) used this procedure and excluded items that correlated highly positively with IM when adjusting their instrument. After the exclusion of some items, they observed rather unexpected relations between one of their adjusted (post-exclusion) scales and external criteria. In the end, they even dropped this adjusted scale from the final inventory. In other words, the assumption that high scores in IM serve as a sufficiently pure measure of socially desirable responding has strongly affected the development of instruments in personality research and beyond (e.g. Ferris et al., 2005) – even to the extent of dropping an entire subscale.

There are several other examples of how research results and conclusions have been affected by IM or similar measures administered in low demand situations and most share the approach of interpreting high scores in these measures as reflecting socially desirable responding. Examples span across describing gender differences (e.g. Bernardi, 2006), exploring racial identity development (e.g. Abrams & Trusty, 2004), or testing differences between online vs. lab assessments (e.g. Risko, Quilty, & Oakman, 2006).

Importantly, however, one might question whether respondents actually engage in socially desirable responding in low demand situations, i.e. when there is no obvious reason to present oneself in an overly positive light. Indeed, research on basic personality traits in terms of the Big Five has found that in a high demand situation (applicant testing) respondents tend to ascribe themselves higher levels in socially desirable traits (agreeableness, conscientiousness, and emotionality stability) as compared to a low demand situation (Detrick, Chibnall, & Call, 2010). Also, findings by Pauls and Crost (2005) comprising both the Big Five traits and the BIDR scales suggested that respondents tend to adapt their self-descriptions to the assumed external demands when faking instructions are introduced – as compared to their response behavior in low demand situations. For example, respondents had higher IM scores when instructed to respond like an applicant whose hiring decision is based on the test results, as compared to their IM scores in a low demand situation. Stated simply, such research findings suggest that socially desirable responding in personality questionnaires occurs in high demand situations in particular, but not, at least not to the same degree, in anonymous, low demand situations. This raises the question what IM and similar questionnaires actually measure when administered in low demand situations.

A potential answer to the question what high scores in IM and similar measures reflect in low demand situations might be found in research suggesting that these measures do not (only) represent response styles, but also carry substantive trait-like information. The key idea behind this is that IM items can be taken literally, thus implying that people respond to IM items in exactly the same way as to any other personality-descriptive items. In this case, IM scales would thus reflect true virtue rather than dishonest responding – at least when strong incentives for impression management are absent (i.e. in low demand situations) – as higher scores are based on stronger agreement with socially desirable attributes. Clearly, this severely questions the practice of using high scores in IM to control for socially desirable responding.

Recently, two approaches have adopted the idea of ascribing substantive interpretations to IM and similar measures. Specifically, Uziel (2010) provided a seminal review of research on IM and similar measures, concluding that high IM scores do not represent social desirability, but interpersonally oriented self-control – a positively valenced trait. Uziel (2014) also showed that self- and peer-reports of IM correlate rather strongly with each other ($r = .44$), further contradicting the view that IM captures a mere response style. Most importantly, across two studies, IM correlated with self- and observer-rated trait self-control ($.26 \leqslant r \leqslant .53$), substantiating that "IM scales measure substantive content associated with self-control aimed at social adaptation" (Uziel, 2014, p. 200).

Another view was taken by De Vries, Zettler, and Hilbig (2014) who also concluded that high scores in IM do not represent a mere response bias, but rather carry trait information of a positive valence. But based on their overall impression of the items, De Vries et al. did not consider self-control as the core construct behind IM scores, but suggested that they are an expression of honesty in particular. For instance, an individual strongly agreeing with an item such as "I always obey laws, even if I'm unlikely to get caught." is assumed to be especially virtuous (rather than overclaiming). Correspondingly, De Vries et al. found that IM scores correlated positively with honesty–humility ($.32 \leqslant r \leqslant .62$) and, to a lesser degree, agreeableness ($.15 \leqslant r \leqslant .35$) and conscientiousness ($.19 \leqslant r \leqslant .35$) as conceptualized in HEXACO Model of Personality (Ashton & Lee, 2007).

Clearly, the approach of De Vries et al. (2014) can be aligned with the view by Uziel (2010, 2014). Both consider high IM scores to be an expression of positively valenced virtue-related traits. Moreover, the HEXACO factors honesty–humility, agreeableness, and conscientiousness all positively correlate with self-control (De Vries & Van Gelder, 2013). The only difference is that De

Vries et al. (2014) suggested high IM scores to reflect honesty in particular, whereas Uziel (2010, 2014) suggested that they especially represent self-control. However, the finding by De Vries et al. (2014) that high IM scores are a blend of agreeableness, conscientiousness, and especially honesty–humility is very much aligned with the idea that high IM scores refer to virtue-related content (cf. e.g. Paulhus, 2002; Pauls & Stemmler, 2003).

In light of these arguments, it is straightforward to assume that – in low demand situations – IM and similar measures are answered like other personality questionnaires and, thus, that high scores in these measures primarily reflect true virtue. Indeed, the studies conducted by De Vries et al. (2014) and Uziel (2010, 2014) all administered IM in low demand situations. Correspondingly, the traditionally implied and still commonly used approach to interpret high IM scores as reflecting socially desirable responding would be entirely misguided. However, the empirical basis for this suggestion is, as yet, rather inconclusive (e.g. De Vries et al., 2014, relied on correlations between personality reports only). Thus, we not only aim to replicate the findings by De Vries et al. (2014) in a large community sample, but also, more importantly, to provide a conclusive behavioral test by linking IM scores to actual dishonest (cheating) behavior: If high IM scores actually reflect true virtue in general and honesty in particular, they must be associated with less cheating behavior.

2. Study 1

First, we aimed to replicate the finding by De Vries et al. (2014) that IM positively correlates with virtue-related traits, namely honesty–humility, agreeableness, and conscientiousness, in a large sample.

2.1. Procedure and participants

Participants were recruited via a German web-portal that hosts links to all kinds of tests and questionnaires (http://www.tested-ich.de). Therein, we advertised the present study as a scientific study on the structure of peoples' personality. People over the age of 18 were allowed to participate and a total of 673 (66% female), aged 18–88 ($M = 27$, $SD = 11$) years, completed the study voluntarily. This sample size yields satisfactory statistical power ($1 - \beta = .80$) to detect even a small effect ($r = .10$) in a t-test for correlation coefficients (with $\alpha = .05$). As an incentive for participation, participants were promised feedback on their personality at the end of the study.

2.2. Methods

After providing consent and demographic information, participants filled out the German version of the HEXACO-60 (cf. Ashton & Lee, 2009; Moshagen, Hilbig, & Zettler, 2014) in order to assess the six factors of the HEXACO Model of Personality, followed by the German version of the BIDR version 6 (cf. Musch, Brockhaus, & Bröder, 2002; Paulhus, 1991). For both inventories, an answer scale ranging from 1 (strongly disagree) to 5 (strongly agree) was presented. In line with common recommendations (Stöber, Dette, & Musch, 2002), we used continuous (instead of dichotomous) scoring for the BIDR, but note that highly similar findings were obtained by using dichotomous scores.

2.3. Results and discussion

Means, standard deviations, alpha reliability estimates (Cronbach's alpha), and correlations between all measures are presented in Table 1, indicating satisfactory alpha reliabilities for all scales ($.62 \leqslant \alpha \leqslant .82$). Corroborating the findings by De Vries et al. (2014), honesty–humility ($r = .55$), agreeableness ($r = .34$), and conscientiousness ($r = .29$) all correlated positively with IM. Moreover, in a multiple regression analysis predicting IM from all six HEXACO factors, only these three factors had a significant effect and the coefficient of honesty–humility was clearly largest (honesty–humility: $\beta = .44$, agreeableness: $\beta = .24$, conscientiousness: $\beta = .21$, all $p < .05$). Thus, high scores in IM were quite strongly aligned with high scores in honesty–humility and, though to a lesser degree, higher scores in agreeableness and conscientiousness. These findings are aligned with the hypothesis that high IM scores indeed represent a blend of moralistic characteristics, especially honesty, when administered under low demand conditions. To conclusively test this interpretation, we next conducted an experiment in which dishonest behavior served as the strong, incentive-compatible behavioral criterion.

3. Study 2

3.1. Procedure and participants

The aim of Study 2 was to critically test whether low or high scores of IM are actually associated with dishonest behavior. To this end, we linked low-demand self-reports in IM to responses in an incentivized cheating task. Participants were recruited by a professionally managed online-panel in order to be heterogeneous concerning sex, age, and educational background. Out of 180 participants who started the study, a total of 153 completed it (15% drop-out-rate). From these, 19 (12%) were further excluded because they always selected the same answer category and/or required less than 2 s per item on average on one or several of the questionnaires. Thus, our final sample comprised 134 participants (52% female), aged 18–81 ($M = 45$, $SD = 17$) years, with the majority (55%) in employment and only 4% students. All respondents completed the study for a flat fee and the chance to win an additional monetary incentive in the cheating task.

3.2. Methods

As in Study 1, participants first provided consent and demographic information. Next, they filled out the German version of the BIDR version 6 as well as other questionnaires in order to avoid bringing the IM items into the focus of attention. As a criterion, we considered response behavior in an incentivized task, variants of which have been used widely in the behavioral ethics literature (e.g. Hilbig & Hessler, 2013). More specifically, participants were asked to take a coin, and to choose one target side (i.e. heads or tails). Next, they were instructed to toss the coin exactly twice and to note the number of successes (defined as coin tosses resulting on the target side). They were informed that having obtained less than two successes (in the two tosses) would not yield an additional monetary incentive; by contrast, reporting exactly two successes in the two coin tosses incurred an additional monetary incentive of €5.00 (about USD 7.00) on top of their flat-fee for participation. Importantly, participants were reminded that nobody was able to detect whether they reported their true outcome or not and that the statistical probability of two successes is 25% (in doing so, we intended to avoid that participants refrained from cheating due to a false belief that two successes might be unlikely and thus suspicious). Importantly, of course, this design did not allow us to assess cheating on an individual basis. However, it does allow for a straightforward interpretation of aggregate patterns, given that the statistical baseline probability of winning (i.e. 25%) is known conclusively. In particular, any scale measuring virtue and honesty must negatively predict the probability of winning

Table 1
Descriptive statistics, Cronbach's alphas, and intercorrelations of variables (Study 1).

	Variable	*M*	*SD*	1.	2.	3.	4.	5.	6.	7.	8.
1.	Honesty–Humility	3.40	0.70	(.77)							
2.	Emotionality	3.14	0.65	.10*	(.78)						
3.	Extraversion	3.16	0.68	−.06	−.17*	(.82)					
4.	Agreeableness	3.09	0.60	.26*	−.06	.04	(.77)				
5.	Conscientiousness	3.32	0.63	.18*	.08*	.06	−.04	(.79)			
6.	Openness to experience	3.60	0.58	.04	−.04	.11*	−.06	.06	(.69)		
7.	Self-deceptive enhancement	2.96	0.48	.08*	−.37*	.38*	.04	.25*	.10*	(.62)	
8.	Impression management	2.73	0.60	.55*	.10*	−.02	.34*	.29*	.03	.22*	(.69)

Note. $N = 673$. Cronbach's alpha in parentheses.
* $p < .05$.

in the coin-toss task, with high scores corresponding to a success-rate that matches the baseline-probability and low scores being associated with more frequent alleged wins.

3.3. Results and discussion

As in Study 1, the alpha reliability of IM ($M = 2.88$, $SD = .62$, $\alpha = .74$) was satisfactory. In the coin-toss task, 51 participants (38% of the sample) claimed to have obtained two successes, thus winning the additional incentive. This proportion is well-above the statistical baseline probability of 25% ($z = 3.5$, $p < .001$), indicating that cheating indeed occurred. Most importantly, we linked IM to the reported outcome of the coin-toss task. Given that the latter represents a binary outcome (coded 1 for winning and 0 otherwise), we used logistic regression analyses. The results indicated that a higher probability of winning was predicted by low scores in IM (*odds-ratio* = .56, $\chi^2(1) = 3.9$, $p < .05$), corresponding to a small to moderate effect size (by the conventions of Rosenthal, 1996). The corresponding pattern is displayed in Fig. 1 which reveals that whereas high scores in IM were actually in line with the statistical baseline probability of winning (implying honest reports in the coin-toss task), low scores in IM were associated with an unreasonably high probability of winning (given the statistical baseline) and must therefore be associated with cheating in the coin-toss task. Consequently, the findings indicate that, when administered in a low demand situation, high scores on IM actually reflect honesty, thereby contradicting the traditional interpretation of IM scores. In fact, the finding strongly supports the view that high scores in IM are linked to true virtue, and especially honesty, when administered in low demand situations.

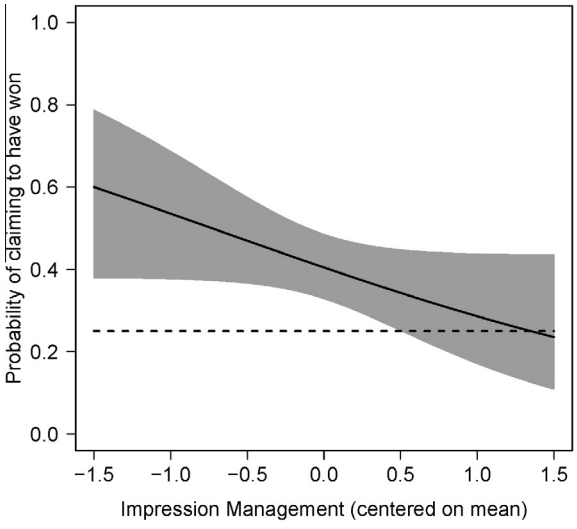

Fig. 1. Predicted probability of claiming two successes in exactly two coin tosses (winning) from a logistic regression on impression management (Study 2). The confidence band represents the 95% confidence interval. The dotted line depicts the baseline probability of winning (25%).

4. General discussion

IM or similar measures have been introduced with the aim to assess a respondent's tendency to distort answers in questionnaires in order to appear more favorably than is actually justified. Given this important implication, they have been used frequently, both in research (Kam, 2013) and practice (Goffin & Christiansen, 2003). However, the exact content of such measures is still a topic of discussion (cf. also the development of the Lie (L) scale of the Minnesota Multiphasic Personality Inventory; e.g. Tellegen & Ben-Porath, 2008/2011). Indeed, high scores in such scales have traditionally been interpreted as reflecting socially desirable responding with striking consequences such as excluding items from inventories or participants from studies. However, doubts have surfaced whether high scores in IM or similar scales actually reflect socially desirable responding – especially in low demand situations in which respondents have no strong incentive to overclaim desirable attributes or underclaim undesirable ones. Thus, in recent years, substantive interpretations of IM measures have been suggested, although the exact content of such scales is still insufficiently understood.

Herein, we investigated the content of such scales by combining two lines of research. On the one hand, recent suggestions imply that IM or similar scales should be treated like personality questionnaires, such that high scores actually represent a blend of true virtues – and not dishonest responding. Specifically, we expected to confirm the suggestion by De Vries et al. (2014) that high IM scores signify honesty in particular. On the other hand, it has been convincingly suggested that most respondents fill in personality questionnaires rather accurately when these are administered under low demand conditions (e.g. Detrick et al., 2010). In combining these arguments, we hypothesized that – under low demand conditions – IM functions like any personality measure, and, in turn, assesses true virtues (especially honesty).

We found support for this conjecture. First, we replicated the findings by De Vries et al. (2014), showing that higher scores in IM are aligned with higher scores in agreeableness, conscientiousness, and especially honesty–humility. More conclusively, we linked IM scores to response behavior in a fully incentivized, anonymous cheating task. Results showed that low scores in IM were related to more cheating behavior whereas high scores in IM were compatible with the statistically expected probability of winning, and, by implication, honest responding. Clearly, this uproots the interpretation of high scores in IM in low demand situations as indicators of dishonest responding. Indeed, exactly the opposite seems to be the case: When administered in low demand situations, high scores in IM – that have been found to be positively correlated to other social desirability measures (e.g. Peterson et al., 2003; Stöber, 2001) – can be taken to reflect true virtues, especially

honesty. By implication, personality researchers should thoroughly rethink the still common practice of excluding items from inventories or participants from studies based on high IM and, more generally, treat conclusions carefully if they are based on interpreting high IM as socially desirable responding.

Whereas our behavioral evidence substantiates a re-interpretation of IM in low demand situations, we cannot conclude whether similar or different mechanisms are at work in high demand situations. Although high scores on IM clearly reflect a blend of positively connoted traits such as honesty, agreeableness, conscientiousness, or self-control in low demand situations, the results need not generalize to high demand conditions. Arguably, in the latter case, high IM scores might represent a mixture of respondents, some of whom may well be honestly reporting virtuous characteristics, while others (strongly) engage in socially desirable responding. In any case, more research aiming to tease apart honest from socially desirable responding in high demand conditions is needed.

As a potential limitation of our investigation, one could argue that positive correlations between IM, honesty–humility, agreeableness, and conscientiousness reflect an artifact of the influence of a common self-presentation bias influencing responses to positively valenced items. However, De Vries et al. (2014) found a similar pattern of correlations across rating sources (self- and observer-reports for both IM and the HEXACO factors) which mitigates this concern. Also, the alternative explanation of a common self-presentation bias cannot account for the negative association between IM and cheating found herein, and a conceptually equivalent result, namely that honesty–humility is negatively correlated with actual cheating behavior (Hershfield, Cohen, & Thompson, 2012). Thus, our results call for caution in the use of IM scales with the original interpretation. Indeed, under low demand conditions, a reversal of how corresponding scores are commonly interpreted seems to be justified.

References

Abrams, L., & Trusty, J. (2004). African Americans' racial identity and socially desirable responding: An empirical model. *Journal of Counseling & Development, 82*, 365–374. http://dx.doi.org/10.1002/j.1556-6678.2004.tb00322.x.

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review, 11*, 150–166. http://dx.doi.org/10.1177/1088868306294907.

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. http://dx.doi.org/10.1080/00223890902935878.

Ben-Porath, Y. S. (2013). Self-report inventories: Assessing personality and psychopathology. In J. R. Graham & J. Naglieri (Eds.), *Handbook of psychology* (2nd ed.. *Assessment psychology* (2nd ed.) (Vol. 10, pp. 622–644). New York, NY: Wiley.

Bernardi, R. A. (2006). Associations between Hofstede's cultural constructs and social desirability response bias. *Journal of Business Ethics, 65*, 43–53. http://dx.doi.org/10.1007/s10551-005-5353-0.

De Vries, R. E., & Van Gelder, J.-L. (2013). Tales of two self-control scales: Relations with Five-Factor and HEXACO traits. *Personality and Individual Differences, 54*, 756–760. http://dx.doi.org/10.1016/j.paid.2012.12.023.

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. http://dx.doi.org/10.1177/1073191113504619.

Detrick, P., Chibnall, J. T., & Call, C. (2010). Demand effects on positive response distortion by police officer applicants on the revised NEO Personality Inventory. *Journal of Personality Assessment, 92*, 410–415. http://dx.doi.org/10.1080/00223891.2010.497401.

Ferris, G. R., Treadway, D. C., Kolodinsky, R. W., Hochwarter, W. A., Kacmar, C. J., Douglas, C., et al. (2005). Development and validation of the Political Skill Inventory. *Journal of Management, 31*, 126–152. http://dx.doi.org/10.1177/0149206304271386.

Goffin, R. D., & Christiansen, N. D. (2003). Correcting personality tests for faking: A review of popular personality tests and an initial survey of researchers. *International Journal of Selection and Assessment, 11*, 340–344. http://dx.doi.org/10.1111/j.0965-075X.2003.00256.x.

Hershfield, H. E., Cohen, T. R., & Thompson, L. (2012). Short horizons and tempting situations: Lack of continuity to our future selves leads to unethical decision making and behavior. *Organizational Behavior and Human Decision Processes, 117*, 298–310. http://dx.doi.org/10.1016/j.obhdp.2011.11.002.

Hilbig, B. E., & Hessler, C. M. (2013). What lies beneath: How the distance between truth and lie drives dishonesty. *Journal of Experimental Social Psychology, 49*, 263–266. http://dx.doi.org/10.1016/j.jesp.2012.11.010.

Kam, C. (2013). Probing item social desirability by correlating personality items with Balanced Inventory of Desirable Responding (BIDR): A validity examination. *Personality and Individual Differences, 54*, 513–518. http://dx.doi.org/10.1016/j.paid.2012.10.017.

Lönnqvist, J.-E., Paunonen, S., Tuulio-Henriksson, A., Lönnqvist, J., & Verkasalo, M. (2007). Substance and style in socially desirable responding. *Journal of Personality, 75*, 291–322. http://dx.doi.org/10.1111/j.1467-6494.2006.00440.x.

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*, 86–97. http://dx.doi.org/10.1026/0012-1924/a000112.

Musch, J., Brockhaus, R., & Bröder, A. (2002). Ein Inventar zur Erfassung von zwei Faktoren sozialer Erwünschtheit (An inventory for the assessment of two factors of social desirability). *Diagnostica, 48*, 121–129. http://dx.doi.org/10.1026//0012-1924.48.3.121.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *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.

Pauls, C. A., & Crost, N. W. (2004). Effects of faking on self-deception and impression management scales. *Personality and Individual Differences, 37*, 1137–1151. http://dx.doi.org/10.1016/j.paid.2003.11.018.

Pauls, C. A., & Crost, N. W. (2005). Cognitive ability and self-reported efficacy of self-presentation predict faking on personality measures. *Journal of Individual Differences, 26*, 194–206. http://dx.doi.org/10.1027/1614-0001.26.4.194.

Pauls, C. A., & Stemmler, G. (2003). Substance and bias in social desirability responding. *Personality and Individual Differences, 35*, 263–275. http://dx.doi.org/10.1016/S0191-8869(02)00187-3.

Peterson, J. B., DeYoung, C. G., Driver-Linn, E., Séguin, J. R., Higgins, D. M., Arseneault, L., et al. (2003). Self-deception and failure to modulate responses despite accruing evidence of error. *Journal of Research in Personality, 37*, 205–223. http://dx.doi.org/10.1016/S0092-6566(02)00569-X.

Risko, E. F., Quilty, L. C., & Oakman, J. M. (2006). Socially desirable responding on the web: Investigating the candor hypothesis. *Journal of Personality Assessment, 87*, 269–276. http://dx.doi.org/10.1207/s15327752jpa8703_08.

Rosenthal, J. A. (1996). Qualitative descriptors of strength of association and effect size. *Journal of Social Service Research, 21*, 37–59. http://dx.doi.org/10.1300/J079v21n04_02.

Stöber, J. (2001). The Social Desirability Scale-17 (SDS-17). Convergent validity, discriminant validity, and relationship with age. *European Journal of Psychological Assessment, 17*, 222–232. http://dx.doi.org/10.1027//1015-5759.17.3.222.

Stöber, J., Dette, D. E., & Musch, J. (2002). Comparing continuous and dichotomous scoring of the balanced inventory of desirable responding. *Journal of Personality Assessment, 78*, 370–389. http://dx.doi.org/10.1207/S15327752JPA7802_10.

Stokes, G. S., & Cooper, L. A. (2001). Content/construct approaches in life history form development for selection. *International Journal of Selection and Assessment, 9*, 138–151. http://dx.doi.org/10.1111/1468-2389.00170.

Tellegen, A., & Ben-Porath, Y. S. (2008/2011). *MMPI-2-RF (Minnesota Multiphasic Personality Inventory-2 Restructured Form): Technical manual*. Minneapolis, MN: University of Minnesota Press.

Uziel, L. (2010). Rethinking social desirability scales: From impression management to interpersonally oriented self-control. *Perspectives on Psychological Science, 5*, 243–262. http://dx.doi.org/10.1177/1745691610369465.

Uziel, L. (2014). Impression management ('lie') scales are associated with interpersonally oriented self-control, not other-deception. *Journal of Personality, 82*, 200–212. http://dx.doi.org/10.1111/jopy.12045.