



UvA-DARE (Digital Academic Repository)

Do impression management and self-deception distort self-report measures with content of dynamic risk factors in offender samples? A meta-analytic review

Hildebrand, M.; Wibbelink, C.J.M.; Verschuere, B.

DOI

[10.1016/j.ijlp.2018.02.013](https://doi.org/10.1016/j.ijlp.2018.02.013)

Publication date

2018

Document Version

Final published version

Published in

International Journal of Law and Psychiatry

License

Article 25fa Dutch Copyright Act

[Link to publication](#)

Citation for published version (APA):

Hildebrand, M., Wibbelink, C. J. M., & Verschuere, B. (2018). Do impression management and self-deception distort self-report measures with content of dynamic risk factors in offender samples? A meta-analytic review. *International Journal of Law and Psychiatry*, 58, 157-170. <https://doi.org/10.1016/j.ijlp.2018.02.013>

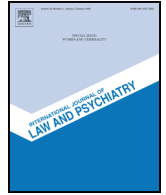
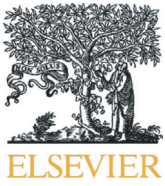
General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)



Do impression management and self-deception distort self-report measures with content of dynamic risk factors in offender samples? A meta-analytic review



Martin Hildebrand^{a,*}, Carlijn J.M. Wibbelink^b, Bruno Verschuere^b

^a Private Practice (www.martinhildebrand.nl), Roermond, The Netherlands

^b Department of Clinical Psychology, University of Amsterdam, The Netherlands

ARTICLE INFO

Article history:

Received 2 June 2017

Received in revised form 20 February 2018

Accepted 20 February 2018

Available online xxxx

Keywords:

BIDR

Impression management

Self-deceptive enhancement

Offenders

Meta-analysis

ABSTRACT

Self-report measures provide an important source of information in correctional/forensic settings, yet at the same time the validity of that information is often questioned because self-reports are thought to be highly vulnerable to self-presentation biases. Primary studies in offender samples have provided mixed results with regard to the impact of socially desirable responding on self-reports. The main aim of the current study was therefore to investigate—via a meta-analytic review of published studies—the association between the two dimensions of socially desirable responding, impression management and self-deceptive enhancement, and self-report measures with content of dynamic risk factors using the Balanced Inventory of Desirable Responding (BIDR) in offender samples. These self-report measures were significantly and negatively related with self-deception ($r = -0.120, p < 0.001; k = 170$ effect sizes) and impression management ($r = -0.158, p < 0.001; k = 157$ effect sizes), yet there was evidence of publication bias for the impression management effect with the trim and fill method indicating that the relation is probably even smaller ($r = -0.07$). The magnitude of the effect sizes was small. Moderation analyses suggested that type of dynamic risk factor (e.g., antisocial cognition versus antisocial personality), incentives, and publication year affected the relationship between impression management and self-report measures with content of dynamic risk factors, whereas sample size, setting (e.g., incarcerated, community), and publication year influenced the relation between self-deception and these self-report measures. The results indicate that the use of self-report measures to assess dynamic risk factors in correctional/forensic settings is not inevitably compromised by socially desirable responding, yet caution is warranted for some risk factors (antisocial personality traits), particularly when incentives are at play.

© 2018 Elsevier Ltd. All rights reserved.

Contents

1.	Introduction	158
2.	The current review	158
3.	Method	159
3.1.	Identification and selection of studies	159
3.2.	Substantive criteria for inclusion	159
3.3.	Coding of studies for the meta-analyses	159
3.4.	Meta-analytic procedures	160
3.4.1.	Publication bias	161
4.	Results	161
4.1.	Study characteristics	161
4.2.	Overall effect sizes	161
4.3.	Publication bias analyses	161
4.4.	Heterogeneity in effect sizes	162
4.5.	Moderator analyses	162
4.5.1.	Impression management	163

* Corresponding author at: Walbreukergraaf 49, 6041 NW, Roermond, The Netherlands.
E-mail address: info@martinhildebrand.nl (M. Hildebrand).

4.5.2.	Self-deceptive enhancement	163
4.6.	Confounding	163
5.	Discussion	166
5.1.	Relation between BIDR impression management and self-deceptive enhancement, and self-report measures with content of dynamic risk factors	166
5.1.1.	Presence of publication bias	167
5.2.	What characteristics moderated the relation between socially desirable responding and self-report measures with content of dynamic risk factors?	167
5.3.	Limitations and directions for future research	167
6.	Conclusions	168
	References	168

1. Introduction

The validity of most psychological measures, including self-reports, is based on the key assumption that subjects are accurate and sincere in their answers (Rogers & Bender, 2003). Socially desirable responding, the tendency to give biased, distorted, and/or overly positive self-descriptions that portray oneself in a way that can make a favorable impression on others (Paulhus, 2002), poses a significant threat to the validity of these measures. Socially desirable responding has long been identified as a potential contaminate of self-report information, particularly in forensic settings, where there often is a strong motivation to present oneself in a positive way. Also, it has been suggested that individuals with antisocial personality disorder—a disorder that is overrepresented in forensic settings—are more likely than others to present themselves in a favorable light in the context of clinical or personality assessment (e.g., American Psychiatric Association, 2013). A positive assessment might lead to favorable outcomes, such as special privileges, entry into a treatment program, parole, or early release (e.g., Benedict & Lanyon, 1992), making a formal assessment of socially desirable responding an important component of a clinical interview or evaluation.

The most common defense against socially desirable response distortion is the use of scales designed to assess the individual's tendency to give overly positive self-descriptions. Scores on these scales have been used to identify suspicious protocols that may be discarded (flagging possible invalid responding), to adjust scores on personality self-reports to account for a desirability response bias (statistical control), to examine convergent and/or divergent validity (score validation), and serve as dependent variables in controlled experiments designed to highlight situations most likely to elicit SDR (outcome assessment) (Vispoel & Kim, 2014; Vispoel & Tao, 2013; also Tan & Grace, 2008).¹

Over the years, a number of instruments have been developed to detect socially desirable responding, including the Edwards (1957) Social Desirability Scale, the Eysenck Lie scale (Eysenck & Eysenck, 1964), the Minnesota Multiphasic Personality Inventory Lie Scale (Hathaway & McKinley, 1951), the Social Desirability Index (SDI; Hofstee, 2003), and the Marlowe–Crowne Social Desirability scale (MCSDS; Crowne & Marlowe, 1960, 1964). In most of these measures, socially desirable responding was conceptualized as a *unitary* construct. Paulhus (1984, 1991), however, argued that measures of social desirability assess two relatively distinct components or factors which he termed impression management and self-deceptive enhancement. Impression management refers to the deliberate distortion of responses with the aim of making a favorable impression on others. This form of socially desirable responding is sometimes described as lying or faking. Self-deception, on the other hand, refers to the tendency of an “unconscious positive bias in item responses with the aim of protecting positive self-esteem” (Stöber et al., 2002, p. 371) and is closely related to narcissism; it is a

deception that is consciously believed and so deeply rooted in a person's belief system that he or she can remain unaware of it and unaffected by situational demands. It is assumed (e.g., Vispoel & Tao, 2013) that impression management typically represents a more serious threat to the validity of questionnaire results than self-deceptive enhancement, because it can represent willful distortion of information.

The Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1984, 1988) consists of 20 impression management items (e.g., “I always obey laws, even if I'm unlikely to get caught”), assessing deliberate attempts to impress the test result user, and 20 self-deception items (“Once I've made up my mind, other people can seldom change my opinion”), assessing unrealistic but honestly held positive self-descriptions. Probably because it is designed to tap impression management as well as self-deception, the BIDR is one of the most widely used instruments to detect socially desirable responding (Li & Bagger, 2007). The BIDR shows satisfactory internal consistency, adequate test–retest, and convergent and discriminant validity (Li & Bagger, 2006).

2. The current review

In forensic settings, self-report provides an important source of information yet at the same time the validity of that information has been doubted because of possible respondents engagement in socially desirable responding. In our view, it is especially important to understand whether (and if so, how) socially desirable responding affects self-report measures with content related to dynamic risk factors (i.e., self-reports intended to contribute to the identification of dynamic risk factors), because dynamic risk factors, also commonly known as criminogenic needs, are characteristics statistically related to recidivism that can (in principle) change and when changed, are expected to result in a decrease in recidivism (Andrews et al., 1990; Andrews & Bonta, 2010). Scholars generally agree that accurate identification of dynamic risk factors is essential as they present the best candidates for intervention. Given the importance of dynamic risk factors, it is essential that self-reports used to gather information on these factors are in fact reliable indicators. Primary studies in offender samples have provided mixed results regarding the impact of socially desirable responding on self-reports. Seifert, Boulas, Huss, and Scalora (2015) examined the degree to which institutionalized sex offenders exhibit response bias on two self-report measures of sexual fantasies and found significant negative correlations with the MCSDS ($r = -0.412$ and -0.316), whereas Keown, Gannon, and Ward (2010) did not find a significant correlation between the BIDR impression management scale and self-reported offense-supportive beliefs in a (small) sample of child sex offenders ($r = -0.076$). There are several reviews on the effect of the BIDR on criterion validity in healthy or clinical samples (Huang, 2013; Li & Bagger, 2006; Perinelli & Gremigni, 2016), but none that focus on the relation with self-report measures with content of dynamic risk factors of the BIDR scales in offender samples has been published. We therefore conducted a meta-analysis to examine the impact of BIDR impression management and self-deception on self-report measures used in order to contribute to the identification of dynamic risk factors in offenders. Since one of the

¹ Alternative approaches to self-reports in forensic assessment that try to also address these problems include implicit measures (e.g., Schmidt, Banse, & Imhoff, 2015), phallometry (e.g., Marshall & Fernandez, 2001), crime scene behavior (e.g., Lehmann, Goodwill, Hanson, & Dahle, 2014), startle-eyeblink response (e.g., Patrick, Bradley, & Lang, 1993), or neuroimaging (e.g., Meijer & Verschuere, 2017).

difficulties in selecting dynamic risk factors is that many putatively dynamic risk factors have been proposed, we relied on meta-analytic studies to identify the dynamic risk factors that are correlated across studies with offending.² Thus, we focused on the relationship of the BIDR scales impression management and self-deceptive enhancement with (1) dynamic risk factors (in principle open to change), (2) as identified through meta-analytic research, and (3) measured through self-report measures used in order to contribute to the identification of dynamic risk factors. Based on the results of the meta-analytic studies, the following dynamic risk factor domains were considered suitable for inclusion in our meta-analytic review: Antisocial cognition (i.e., attitudes, values, beliefs, rationalizations favorable to crime), Antisocial personality pattern/traits (e.g., impulsive, anger problems, low on self-control), Substance misuse/abuse, and Deviant sexual interests.³

3. Method

3.1. Identification and selection of studies

The Medline, PsychARTICLES, PsychINFO, PubMed, ScienceDirect, and Web of Science databases were searched to August 31, 2016 for the combination of terms ('Balanced Inventory of Desirable Responding OR BIDR) and (offender OR forensic) to find potential studies for inclusion. Studies were included if they were published in peer reviewed English language journals. A total of 242 unique records were found, 167 of which were excluded because the BIDR was not used and/or no forensic sample was used. The reference lists of the remaining 75 papers were examined in order to identify additional studies that did not emerge in the initial database search (resulting in 12 additional potential papers). Finally, four additional potential usable papers were identified through other sources. Altogether, we identified 91 potential usable articles, that were read to determine eligibility.

3.2. Substantive criteria for inclusion

We included articles if they reported on the correlation between at least one of the BIDR scales and at least one self-report measure with content of antisocial cognition, antisocial personality pattern/traits, or substance misuse. We retained 19 publications for inclusion. Fig. 1 shows the flowchart of the study review and selection process (a list of the 72 excluded studies is available from the first author upon request).

² We based our selection of dynamic risk factors/constructs on the results of the following meta-analytic studies (see Appendix A on <https://osf.io/ec8dn/> for the complete references of these studies): Bennett, Holloway, and Farrington (2008), Bonta, Blais, and Wilson (2013, 2014), Bonta, Law, and Hanson (1998), Dowden and Brown (2002), Foran and O'Leary (2008), Gendreau, Goggin, and Smith (2002), Gendreau, Little, and Goggin (1996), Hanson and Bussière (1998), Hanson and Morton-Bourgon (2004, 2005), Hawes, Boccaccini, and Murie (2013), Helmus, Hanson, Babchishin, and Mann (2013), Hemphill, Hare, and Wong (1998), Jolliffe and Farrington (2004), Leistico, Salekin, DeCoster, and Rogers (2008), Mann, Hanson, and Thornton (2010), Moore et al. (2008), Rosenfeld (2004), Salekin, Rogers, and Sewell (1996), van Langen, Wissink, van Vugt, van der Stouwe, and Stams (2014), Walters (2003a, 2003b), and Yu, Geddes, and Fazel (2012).

³ Antisocial personality pattern/traits, Antisocial cognition, and Substance misuse belong to Andrews and Bonta's "Central Eight" risk factors most predictive of general offending (together with History of criminal behavior, Procriminal companions, Education/Employment, Family/Marital, Leisure/Recreation). Since no self-report measure had content related to Procriminal companions, Education/Employment, Family/Marital, and Leisure/Recreation (all dynamic risk factors) these factors were not included in our meta-analysis. History of criminal behavior was not included because it is not a dynamic risk factor. Unfortunately, we had to exclude Deviant sexual interests, a well-established dynamic risk factor for sexual offending, since only one study in our meta-analysis fulfilled the inclusion criteria mentioned above (for more information on this, see Section 3.3, second paragraph, where we describe the coding of the outcome characteristic).

3.3. Coding of studies for the meta-analyses

We recorded study characteristics that potentially can influence the effect of impression management and/or self-deceptive enhancement on self-report measures with content of dynamic risk factors. Study characteristics were grouped into outcome, predictor, sample, and publication bias characteristics (the publication bias characteristics are described in 3.4.1 below). Complete Tables showing the coding of all moderator variables for each study are available on <https://osf.io/ec8dn/> (impression management: Appendix B; self-deception: Appendix C).

The outcome characteristic relates to the measurement of dynamic risk factors. We investigated whether type of dynamic risk factor (i.e., antisocial cognition, antisocial personality pattern/traits, substance misuse, or deviant sexual interests) may play a role as a moderator. In the 91 potential articles for inclusion the BIDR scales have been correlated with 68 different self-report measures, assessing a wide variety of (personality) constructs, including, for example, depression, coping, self-esteem, and criminal attitudes. Study authors (MH and BV) coded whether or not (yes/no) these self-report measures had content related to one of the four selected dynamic risk factors (92.6% agreement between the authors; $\kappa = 0.845$). Errors and discrepancies were discussed and corrected, and agreement was reached among the authors. Of the 68 self-report measures, 44 had content related to (at least) one of the four selected dynamic risk factors (an overview of the excluded self-report measures is available from the first author upon request). Next, study authors coded whether these 44 self-report measures had content related to the selected risk factor domains. If a self-report measure consisted of various subscales, which was usually the case, the respective subscales were (also) sorted in the appropriate risk domain.⁴ In this way, the authors coded 103 subscales (91.1% agreement between the two authors with regard to the appropriate dynamic risk factor; $\kappa = 0.835$). Again, errors and discrepancies were discussed and corrected, and agreement was reached among the authors. The result of this procedure was that 100 subscales of 44 different questionnaires (reported in 23 articles) were considered suitable for inclusion and coded as either tapping primarily into antisocial cognition, antisocial personality pattern/traits, substance misuse, or deviant sexual interests.⁵ Authors were contacted by email to obtain relevant information on the association between BIDR impression management and/or self-deception and self-report measures with content of dynamic risk factors that was not provided in these 23 articles.⁶ As a result of requesting this information, we had to exclude another eight subscales of six different questionnaires (reported in four articles). Also, at this stage, it became clear that we had to exclude deviant sexual interests (and therefore four more subscales were excluded), since only one study in our meta-analysis fulfilled the inclusion criteria.

The three predictor characteristics relate to the features of the BIDR scales. The first is the use of the original version of the BIDR or a translated version. The second is the use of the original versus a modified or short BIDR version, classified into the following four categories for the impression management scale: (1) IM_20 (original 20-item version), (2) IM_17 (17-item version), (3) IM_8 (8-item version), and (4) IM_5

⁴ Subscales without content related to dynamic risk factors (e.g., validity scales) were excluded.

⁵ Examples of self-report measures that we decided to have content related to the selected dynamic risk factors are the Criminal Attribution Inventory (antisocial cognition), the Impulsivity Scale (antisocial personality pattern/traits, scales of the Multiphasic Sex Inventory (deviant sexual interests), and the Texas Christian University Drug Screen I (substance misuse).

⁶ For example, when the magnitude of (some of) the correlations was not reported, or correlations were not reported for offenders only, authors were requested to obtain the relevant correlation sizes. We contacted the first authors of seven publications to provide additional data. Two authors responded positively to this request. The others stated that the requested data were no longer available (one), that they were willing to send us the data but failed to do so after two reminder emails (two), or did not respond at all (two).

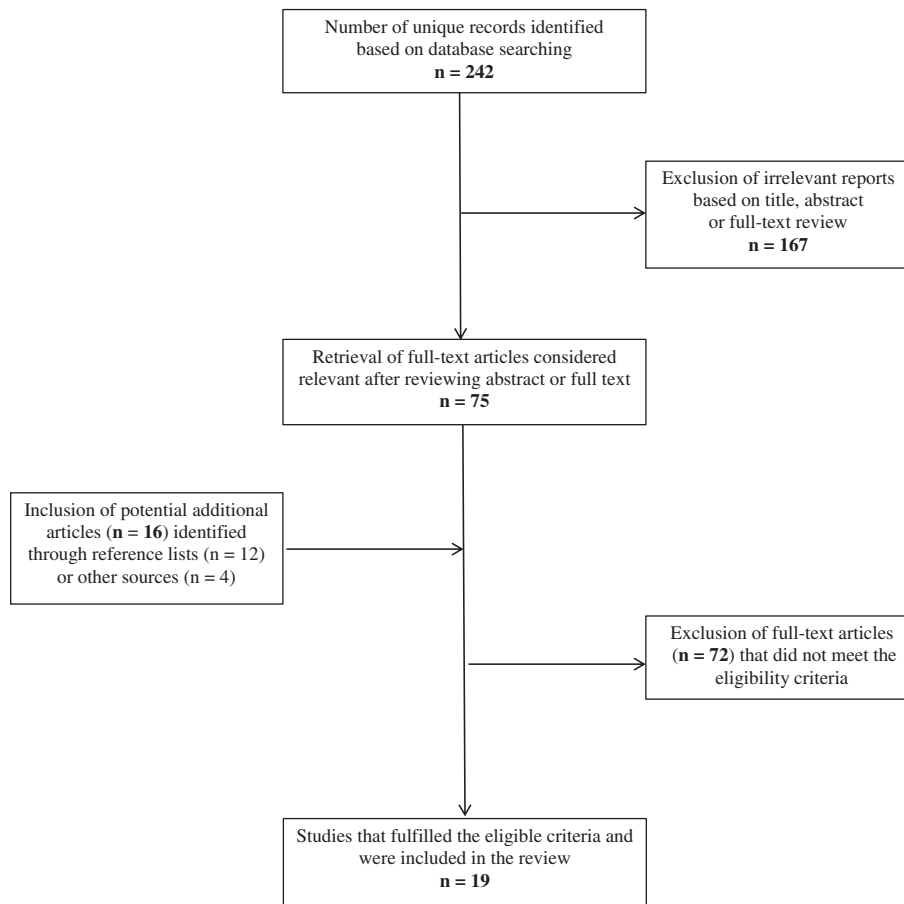


Fig. 1. Flowchart showing the literature search process and the selection of studies included in the review.

(5-item version); and the following four categories for the self-deceptive enhancement scale: (1) SDE_20 (original 20-item version), (2) SDE_15 (15-item version), (3) Denial of the Negative (DN; 11-item subscale of the self-deceptive enhancement scale, and (4) Over Confident Rigidity (OCR; 9-item subscale of the self-deceptive enhancement scale.⁷ The third is the observed internal consistency of the impression management and self-deceptive enhancement scales.

We also included seven sample characteristics. First, gender of the sample was coded as either (1) exclusively male, (2) exclusively female, or (3) mixed. Second, we assessed the moderating impact of sample age (Kroner & Weekes, 1996; Mathie & Wakeling, 2011). Third, we examined educational background, coded as the average years of schooling and also the proportion of offenders that did not complete high school. (Unfortunately, since most studies provided limited to no information on educational background, we were forced to exclude this potential moderator.) Fourth, we examined whether effects differed depending on the setting of the study. Setting was coded as (1) incarcerated/institutionalized (“in”) or (2) community/probation (“out”). Fifth, we included offenders’ type of offense, classified into four categories: (1) violent non-sexual offense, (2) sexual offense, (3) domestic offense, and (4) other (i.e., sample includes offenders with different types of offenses, or offenses other than violent, sexual, or domestic offenses). Sixth, we investigated whether (perceived) personal implications (i.e., incentives; Tourangeau & Yan, 2007) may play a role as a moderator. Studies were coded as (1) no personal implications (i.e., data used for research purposes only), (2) possible implications (e.g., responses used for treatment but incentives are unclear), and

(3) implications (e.g., responses may have consequences for parole). Seventh, cross-cultural differences, operationalized as the location of the study and coded as (1) North America, (2) Europe, or (3) Oceania were examined.

Inter-rater agreement was based on 10 studies that were randomly selected and scored by two coders (MH and CW). The intraclass correlation coefficient was used for continuous variables and Kappa for categorical variables. The inter-rater reliability for continuous variables proved to be perfect, with interclass correlations of 1.00 for age, year of publication, sample size, reliability, and the effect size value. Kappa’s for the categorical variables were at least satisfactory, ranging from 0.73 for personal implications (incentives) to 1.00 for geographic location, setting, offense type, BIDR version, impression management/self-deceptive enhancement version, and gender.

3.4. Meta-analytic procedures

Two separate meta-analyses were conducted: one assessed the relation between impression management and self-report measures with content of dynamic risk factors, and one assessed the relation between self-deceptive enhancement and these self-reports. For the effect size, correlation coefficient r was computed using formulas from Lipsey and Wilson (2001), and recoded into Fisher z -values (Lipsey & Wilson, 2001), since the variance of the correlation coefficient depends strongly on the correlation (Borenstein, Hedges, Higgins, & Rothstein, 2009). After the analyses, the mean Fisher z -values were transformed back into correlation coefficients for interpretation. For the moderator analyses, continuous variables were centered around its mean and categorical variables were converted to dummy variables through binary coding. We checked for outliers by searching for extreme values of

⁷ See Kroner and Weekes (1996).

the effect sizes (Fisher z -values $< -3.29 SD$ or $> 3.29 SD$; Tabachnick & Fidell, 2007). One effect size was identified with a z value < -3.29 (impression management). To reduce the impact of this outlier, the effect size was substituted by the nearest observed effect size within the $-3.29 SD < x < 3.29 SD$ range (i.e., 0.76).

The data of most studies resulted in multiple effect sizes per study. Consequently, the assumption of independent effect sizes, is violated. In order to deal for this dependency, we applied a three-level random effects model (Cheung, 2014; van den Noortgate, López-López, Marín-Martínez, & Sánchez-Meca, 2013). By using a three-level random effects model, multiple effect sizes derived from the same study can be included, generating maximum information and statistical power.⁸ A three-level random effect model accounts for three sources of variance: sampling variance (level 1), variance between effect sizes from the same study (level 2), and variance between studies (level 3). We performed one-tailed log-likelihood-ratio-tests to determine whether significant variation was present between effect sizes from the same study (level 2) and/or between studies (level 3). If there was evidence for heterogeneity in effect sizes, we conducted moderator analyses by extending the model with characteristics of studies or effect sizes. For models including moderators, an omnibus test was conducted, which tests the null hypothesis that the group mean effect sizes are equal. The Knapp and Hartung (2003) was applied to control for Type I error rates. Analyses were performed with the metafor package (Viechtbauer, 2010) for the R environment (version 3.2.1; R Core Team, 2015), using guidelines formulated by Assink and Wibbelink (2016). All model parameters were estimated by using the restricted maximum likelihood procedure.

3.4.1. Publication bias

A common problem in conducting a meta-analysis is that some studies may not have been published due to unfavorable or non-significant results. This may result in an overestimation of the true effect size, also known as the publication bias (Borenstein et al., 2009) or file drawer bias (Rosenthal, 1995). To test for the file drawer bias, funnel plot asymmetry was examined using Egger's test (Egger, Smith, Schneider, & Minder, 1997). The funnel plot is a scatter plot of the distribution of each effect size on the horizontal axis against its precision (reciprocal of the standard error). In case this test was statistically significant, the trim and fill procedure was performed, which corrects for funnel plot asymmetry by imputing missing effect sizes (Duval & Tweedie, 2000). Note that imputing non-existing effect sizes into a meta-analysis is controversial and that we should not rely on the results of imputed effect sizes for the final conclusions (Sutton, Duval, Tweedie, Abrams, & Jones, 2000).

In addition, three other potential indicators of publication bias were explored. First, we examined whether there are differences between studies that assessed the impact of socially desirable responding on self-report measures with content of dynamic risk factors as a primary goal versus studies that reported on this relation as a secondary goal or side issue (cf. Suchotzki, Verschuere, van Bockstaele, Ben-Shakhar, & Crombez, 2017). It can be argued that publication bias is especially present in studies focusing primarily on the relation between socially desirable responding and self-report measures with content of dynamic risk factors because null-findings in these studies may be

difficult to get published, whereas side issue studies do not rely as much on the statistical significance of the results of the relation between socially desirable responding and self-report measures with content of dynamic risk (i.e., publication of the study does not depend on side issue findings, results can be considered as by-catch). Second, since in many research areas a decline in effect sizes over the years can be observed—the so-called “decline effect” (Schooler, 2011)—we explored whether there was an influence of publication year. Third, since previous findings suggest that the effect sizes reported in meta-analyses are negatively correlated with study sample sizes (Ioannidis, 2005; Kühberger, Fritz, & Scherndl, 2014), we also investigated the effect of sample size.⁹

4. Results

4.1. Study characteristics

The current meta-analysis included 19 articles reporting on 17 independent studies and 327 effect sizes (for impression management, $k = 157$; for self-deceptive enhancement, $k = 170$). Eighteen articles (16 independent studies) reported correlations between impression management and self-report measures with content of dynamic risk factors, and 16 articles (14 independent studies) reported correlations between self-deception and these self-reports. In total, 88 (sub)scales of 38 different questionnaires with content of selected dynamic risk factors were included (see Appendix D on <https://osf.io/ec8dn/> for a list of the references of these 38 questionnaires). Table 1 presents a description of the studies included in the meta-analysis. In Table 2 we report the (subscales of) self-report measures per dynamic risk factor. Most of the self-report measures had content related to antisocial cognition ($N = 26$; 63 scales in total), followed by antisocial personality pattern/traits ($N = 10$; 19 scales), and substance misuse ($N = 6$; 6 scales).¹⁰

4.2. Overall effect sizes

The overall mean effect size of the relation between impression management and self-report measures with content of dynamic risk was $r = -0.158$, $p < 0.001$ ($k = 157$ effect sizes), indicating that higher scores on impression management are associated with lower scores on these self-report measures (Table 3). This effect size was small according to Cohen's (1988) convention regarding magnitude of correlation size, with r 's between 0.10 and 0.30 considered “small”, those between 0.30 and 0.50 “medium” and those from 0.50 “large” (but see Gignac & Szodorai, 2016). The overall mean effect size of the relation between self-deceptive enhancement and self-report measures with content of dynamic risk was $r = -0.120$, $p < 0.001$ ($k = 170$ effect sizes), indicating that as scores on self-deception increase, scores on these self-reports decrease (Table 4). This effect size can also be considered small.

4.3. Publication bias analyses

We examined possible publication bias by testing funnel plot asymmetry for studies on impression management and self-deceptive enhancement. The standard normal deviate was regressed against the estimate's precision (Egger et al., 1997). For studies on self-deception,

⁸ An important requirement in a traditional meta-analysis is that there is no dependency between effect sizes. However, there are now different approaches for including multiple effect sizes derived from the same study (generating maximum information and statistical power) and dealing with the resulting dependency between effect sizes (e.g., Borenstein et al., 2009). One of these approaches is applying a three-level structure to a meta-analytic model, a strong method for dealing with dependency between effect sizes (e.g., Assink & Wibbelink, 2016). A three-level meta-analytic model accounts for the hierarchical structure in the data (multiple effect sizes within studies). If, for example, one study results in 15 effect sizes and another in only one effect size, the study with 15 effect sizes will not contribute 15 times as much as the study with one effect size to the calculation of the mean effect size (Van den Noortgate et al., 2013).

⁹ Meta-analytic investigations sometimes require a minimum sample size for inclusion because it is known that studies with small sample sizes are more susceptible to publication bias and tend to overestimate effect sizes (e.g., Turner, Bird, & Higgins, 2013). Specifically, many meta-analytic studies have used samples sizes of 10 and more as inclusion criterion (Suchotzki et al., 2017). We checked our data, and none of the studied we included had a sample size < 10 , so no studies needed to be excluded based upon this criterion.

¹⁰ The numbers of the questionnaires add up to 42 (not 38) because four self-report measures have subscales with content of Antisocial cognition and subscales with content of Antisocial personality pattern/traits (also Table 2).

Table 1
Articles Included in the meta-analysis.

Article	Year	N	Sex	Offense type	BIDR scales	Self-report measure ^a
Barrowcliffe & Gannon	2015	18 ^b	Male	Firesetters	IM	FAS, FPS, FSS, IFS
Bowen & Gilchrist	2006	120	Male	Domestic	IM, SDE	ABI, IBWB, NAS
Dalbert & Filke	2007	100	Male	Mixed	IM_8	Justice judgments, personal BJW, STAXI
Davis et al.	2014	1747	Male	Mixed	IM, SDE	BASA, CBQ, DASES, DRIE, DTCQ
Dutton & Hemphill	1992	50	Male	Domestic	IM, SDE	CTS, MAI, PMWI
Ilycan et al.	2015	95 ^b	Male	Domestic	IM, SDE	PPI-SF
Keown et al.	2010	34 ^c	Male	Sex	IM	Molest scale
Kroner, Boer et al.	2004	176	Male	Mixed	IM_17, DN, OCR	RSAS
Kroner, Mills et al.	2004	230 ^b	Male	Mixed	IM_17, DN, OCR	CRAI
Mann et al.	2007	481	Male	Sex	IM, SDE	SWCH
Mathie & Wakeling	2011	1730	Male	Sex	IM, SDE	CSQ, Deceitful women, dissipation-rumination ES, HI, IRI, IS, MSI, SWCH
Mills & Kroner	2005	172 ^d	Male	Violent	IM_5, SDE_15	MCAA-B
Mills & Kroner	2006	172	Male	Violent	IM, SDE	MCAA-B
O'Brien & Webster	2007	75 ^e	Male	Sex	IM, SDE	IBAQ
Richards & Pai	2003	312	Mixed	Mixed	IM, SDE	TCUDS I
Seto & Barbaree	1993	18 ^b	Male	Sex	IM, SDE	QMEE
Simourd & Mamuzza	2000	146	Male	Violent	SDE	AQ, HIQ, NAS
Visu-Petra et al.	2008	50 ^b	Mixed	Violent	IM, SDE	CSS-M
Wakeling & Barnett	2011	936	Male	Sex	IM, SDE	Dissipation-rumination

Note. BIDR = Balanced Inventory of Desirable Responding. ABI = Abusive Behaviour Inventory (Shepard & Campbell, 1992). AQ = Aggression Questionnaire (Buss & Perry, 1992). BASA = Beliefs About Substance Abuse Questionnaire (Beck, Wright, Newman, & Liese, 1993). CBQ = Cravings Belief Questionnaire (Beck et al., 1993). CRAI = Criminal Attribution Inventory (Kroner & Mills, 2003). CSQ = Children and Sex Questionnaire (Beckett, 1987). CSS-M = Criminal Sentiments Scale-Modified (Shields & Simourd, 1991). CTS = Conflict Tactics Scale (Straus, 1979). DASES = Drug Avoidance Self-Efficacy Scale (Martin, Wilkinson, & Poulos, 1995). Deceitful Women = Deceitful Women Scale (ISMG, 1995). DN = Denial of the negative. Dissipation-Rumination = Dissipation-Rumination Questionnaire (Capara, 1986). DRIE = Drinking/Drug Related Locus of Control (Donovan & O'Leary, 1978). DTCQ = Drug Taking Confidence Questionnaire (Annis, Turner, & Sklar, 1997). ES = Entitlement to Sex (Hanson, Gizzarelli, & Scott, 1994). FAS = Fire Attitude Scale (Muckley, 1997). FPS = Fire Proclivity Scale (Gannon & Barrowcliffe, 2012). FSS = Fire Setting Scale (Gannon & Barrowcliffe, 2012). HI = Hypermasculinity Inventory (Mosher & Sirken, 1984). HIQ = Hostile Interpretations Questionnaire (Mamuzza & Simourd, 1997). IBAQ = Internet Behaviour and Attitudes Questionnaire (O'Brien & Webster, 2007). IBWB = Inventory of Beliefs about Wife Beating (Saunders, Lynch, Grayson, & Linz, 1987). IFS = Identification with Fire Scale (Gannon, Ó Ciardha, & Barnoux, 2011). IM = Impression Management scale. IRI = Interpersonal Reactivity Index (Davis, 1980). IS = Impulsivity Scale (Eysenck & Eysenck, 1978). Justice Judgments (Dalbert & Filke, 2007; Otto & Dalbert, 2005). MAI = Multidimensional Anger Inventory (Siegel, 1986). MCAA-B = Measures of Criminal Attitudes and Associates Part B (Mills & Kroner, 1999). Molest Scale = Bumby's (Rape and) Molest Scale (Bumby, 1996). MSI = Multiphasic Sex Inventory (Nichols & Molinder, 1984). NAS = Novaco Anger Scale (Novaco, 1994). OCR = Overconfident rigidity. Personal BJW = Personal Belief in a Just World (Dalbert, 1999). PMWI = Psychological Maltreatment of Women Inventory (Tolman, 1989). PPI-SF = Psychopathic Personality Inventory-Short Form (Lilienfeld & Hess, 2001). QMEE = Questionnaire Measure of Emotional Empathy (Mehrabian & Epstein, 1972). RSAS = Rape Supportive Attitude Scale (Kroner, Boer, & Mills, 2004). SDE = Self-deception enhancement scale. STAXI = State-Trait Anger Expression Inventory (Spielberger, 1988). SWCH = Sex With Children is Justifiable Questionnaire (Marshall, 1995). TCUDS I = Texas Christian University Drug Screen I (Simpson & Knight, 1998).

^a Only the self-report measures that are included in our meta-analysis are mentioned here. Due to space limitations and because of readability issues subscales of the self-report measures are not mentioned here. These can be found in Table 2.

^b The study also included non-offenders but only the number of offenders are mentioned here.

^c The study also included 33 offender controls, of which no correlation between BIDR scales and self-report measure were reported.

^d The study also included a sample of 97 offenders of which no correlations between BIDR scales and self-report measure were calculated.

^e The total sample consisted of 123 internet sexual offenders, of which 75 participants had BIDR data.

the intercept did not significantly deviate from zero, $t(168) = -0.29$, $p = 0.769$. However, the intercept significantly deviated from zero for studies on impression management, $t(155) = -2.56$, $p = 0.010$. This suggests that there was an indication for publication bias for studies on impression management. The trim and fill analysis showed an overall mean effect size for impression management of $r = -0.071$, $p < 0.001$, based on 16 independent studies and 186 effect sizes. This indicates that the true relation between impression management and self-report measures with content of dynamic risk factors may be somewhat smaller than what we observed ($r = -0.158$, $p < 0.001$), but also that it is still significant and negative.

Studies that reported on the relation between the BIDR scales and self-report measures with content of dynamic risk as a primary goal did not report significantly higher effect sizes than studies that reported on this relation as a side issue.

We did find a significant effect for publication year. For both impression management and self-deception, smaller effect sizes were found for more recently published studies (i.e., decline effect).

Finally, sample size significantly moderated the effect size of the relation between self-deception (but not impression management) and self-report measures with content of dynamic risk factors: larger effect sizes were found when the sample size decreased. The results of the analyses of the three additional measures of publication bias are presented in Table 3 (impression management) and 4 (self-deceptive enhancement).

4.4. Heterogeneity in effect sizes

The three-level approach allowed assessing the heterogeneity between studies (i.e., between-study heterogeneity), as well as heterogeneity between effect sizes from the same study (i.e., within-study heterogeneity). With regard to the heterogeneity in effect sizes for impression management, we found significant variation between studies, $\sigma^2 = 0.020$, $\chi^2(1) = 39.46$, $p < 0.001$, as well as between effect sizes within studies, $\sigma^2 = 0.035$, $\chi^2(1) = 3294.21$, $p < 0.001$. Regarding the heterogeneity in effect sizes for self-deceptive enhancement, we also found significant variation between studies, $\sigma^2 = 0.008$, $\chi^2(1) = 4.25$, $p = 0.039$, as well as between effect sizes within studies, $\sigma^2 = 0.032$, $\chi^2(1) = 3344.10$, $p < 0.001$.

4.5. Moderator analyses

In addition to the publication bias moderators described above, we conducted moderator analyses to identify possible outcome, predictor, and sample characteristics that could moderate the relationship between impression management and self-deceptive enhancement on the one hand and self-report measures with content of dynamic risk factors on the other hand. The results of these analyses are presented in Table 3 (impression management) and 4 (self-deception). Below, we describe the results of the moderating variables of which the omnibus test statistics was significant ($p < 0.05$).

Table 2

Included self-report measures with content of dynamic risk factors (N = 38) grouped per dynamic risk factor.

	Included (sub)scale of self-report measure (N = 88)
<i>Antisocial cognition</i>	
Aggression questionnaire	Hostility
Bumby's (rape and) molest scale	Total score
Children and sex questionnaire	Cognitive distortions; emotional congruence with children
Criminal attribution inventory	Psychopathology; personal; victim; alcohol; society; randomness
Criminal sentiments scale-modified	Attitudes toward law, courts, and police; tolerance for law violations; Identification with criminal others
Deceitful women scale	Total score
Dissipation-rumination questionnaire	Total score
Entitlement to sex	Total score
Fire attitude scale	Total score
Fire proclivity scale	Total score
Hostile interpretations questionnaire	Total score; hostility towards authority figures; Hostility in close interpersonal relationships; hostility in distant interpersonal relationships; hostility in work relationships; hostility in stranger interactions; overgeneralization; attribution; personal responsibility; hostile reaction; external blame
Hypermasculinity inventory	Danger as exciting; violence as manly; callous sex attitudes
Identification with fire scale	Total score
Internet behaviour and attitudes questionnaire	Distorted thinking; self-management; total (attitude items)
Interpersonal reactivity index	Fantasy; Empathic concern; perspective taking; personal distress
Inventory of beliefs about wife beating	Sympathy for battered women scale
Justice judgments	Justice of legal proceedings; interpersonal justice; Justice of prison affairs
Measures of criminal attitudes and associates Part B	Total score (criminal attitudes); violence; entitlement; antisocial intent; associated
Multidimensional anger inventory	Range of anger-eliciting situations; hostile outlook; anger-in
Multiphasic sex inventory	Lie scales (child molest lie, rape lie, incest lie, exhibitionism lie); cognitive distortion/immaturity; justifications
Novaco anger scale	Cognitive
Personal belief in a just world	Total score
Psychological maltreatment of women inventory	Dominance/isolation
Questionnaire measure of emotional empathy	Total score
Rape supportive attitude scale	Total score
Sex with children is justifiable questionnaire	Total score; harmless sex with children; provocative sexual children
<i>Antisocial personality pattern/traits</i>	
Abusive behaviour inventory	Psychological abuse; physical abuse
Aggression questionnaire	Total score; physical aggression; verbal aggression; anger
Conflict tactics scale	Verbal aggression; violence
Fire setting scale	Total score
Impulsivity scale	Total score
Multidimensional anger inventory	Anger-arousal; anger-out
Novaco anger scale	Part B; arousal; behavioral
Psychological maltreatment of women inventory	Emotional/verbal abuse
Psychopathic personality inventory-short form	Fearless dominance; impulsive antisociality
State-trait anger expression inventory	Anger-out
<i>Substance misuse</i>	
Beliefs about substance abuse questionnaire	Total score
Cravings belief questionnaire	Total score
Drug avoidance self-efficacy scale	Total score
Drinking/drug related locus of control	Total score
Drug taking confidence questionnaire	Total score
Texas Christian University drug screen I	Total score

Note. The Aggression Questionnaire, Multidimensional Anger Inventory, Novaco Anger Scale, and Psychological Maltreatment of Women Inventory have subscales with content of Antisocial cognition and subscales with content of Antisocial personality pattern/traits, and are therefore grouped under both risk factors.

4.5.1. Impression management

The outcome characteristic affected the relationship between impression management and self-report measures with content of dynamic risk factors (Table 3). Significant negative effect sizes were found for antisocial cognition ($r = -0.119$) and antisocial personality pattern/traits ($r = -0.317$), but not for substance misuse ($r = -0.123$). Moreover, the effect size for antisocial personality pattern/traits was significantly larger compared to the effect size for antisocial cognition. Also, the effect size for the relation between impression management and self-report measures with content of dynamic risk factors was affected by one sample characteristic, personal implications (incentives), with significant negative associations between impression management and these self-report measures when there were incentives and when there were no incentives, whereas no significant association was found when there were possible incentives. Additionally, the effect size for incentives was significantly larger compared to no incentives and compared to possible incentives. None of the predictor characteristics moderated the association between impression management and self-report measures with content of dynamic risk factors.

4.5.2. Self-deceptive enhancement

Nor the outcome characteristic as the predictor characteristics moderated the association between self-deceptive enhancement and self-report measures with content of dynamic risk factors (Table 4). However, we did find that one sample characteristic, the setting of the study, significantly moderated the association between self-deception and self-report measures with content of dynamic risk factors. Significant negative effect sizes between self-deception and these self-report measures were found when offenders were incarcerated/institutionalized as well as when they were in the community or on probation. The effect size for incarcerated/institutionalized offenders was significantly smaller than the effect size for offenders in the community/on probation.

4.6. Confounding

In order to check for dependencies between the significant moderators, we calculated Cramer's V for categorical variables, Pearson's r for continuous variables, and Spearman's ρ for a categorical and continuous variable to qualify the strength of the observed associations.

Table 3
Results for the moderators for impression management.

Moderator variables	# Studies	# ES	Mean r	β_0 (95% CI)	β_1 (95% CI)	Omnibus test	p-value	Variance level 2 ^a	Variance level 3 ^b
Impression management	16	157	-0.158***	-0.159 (-0.244; -0.074)***				0.035***	0.020***
<i>Outcome characteristic</i>									
Dynamic risk factor	16	157				F(2,154) = 6.232	0.002**	0.032***	0.019***
AC (RC)	12	109	-0.119*	-0.120 (-0.214; -0.027)*					
APP	7	20	-0.317***	-0.328 (-0.457; -0.199)***	-0.208 (-0.325; -0.091)***				
SM	2	28	-0.123	-0.124 (-0.338; 0.091)	-0.003 (-0.237; 0.231)				
<i>Predictor characteristics</i>									
BIDR: translation	16	157				F(1,155) = 2.224	0.138	0.035***	0.019**
No translation (RC)	14	149	-0.179***	-0.181 (-0.269; -0.093)***					
Translation	2	8	0.018	0.018 (-0.231; 0.266)	0.199 (-0.065; 0.463)				
BIDR: version	16	157				F(3,153) = 0.296	0.829	0.035***	0.022***
IM (RC)	13	122	-0.175***	-0.177 (-0.278; -0.079)***					
IM_17	2	24	-0.079	-0.079 (-0.310; 0.152)	0.098 (-0.153; 0.350)				
IM_8	1	5	-0.086	-0.086 (-0.433; 0.260)	0.091 (-0.270; 0.451)				
IM_5	1	6	-0.207+	-0.210 (-0.445; 0.024)+	-0.033 (-0.258; 0.191)				
Reliability BIDR scales	5	65	-0.186*	-0.188 (-0.357; -0.019)*	-0.945 (-3.046; 1.155)	F(1,63) = 0.809	0.372	0.032***	0.024**
<i>Sample characteristics</i>									
Age	15	153	-0.162**	-0.163 (-0.264; -0.063)**	-289 ^c (-0.011; 0.010)	F(1,151) = 0.003	0.957	0.034***	0.023***
Gender	16	157				F(1,155) = 0.071	0.790	0.035***	0.021***
Male (RC)	13	146	-0.162***	0.163 (-0.258; -0.069)***					
Mixed	3	11	-0.129	0.130 (-0.360; 0.100)	0.033 (-0.215; 0.282)				
Setting	16	157				F(1,155) = 1.110	0.294	0.035***	0.019***
Community/probation (RC)	4	19	-0.242*	-0.247 (-0.433; -0.061)**					
Incarcerated/institutionalized	12	138	-0.135**	-0.136 (-0.230; -0.042)**	0.111 (-0.097; 0.319)				
Offense type	16	157				F(3,153) = 0.362	0.781	0.035***	0.021***
Mixed (RC)	6	60	-0.128*	-0.129 (-0.258; -424 ^c)*					
Violent	3	24	-0.147	-0.148 (-0.351; 0.055)	-0.019 (-0.256; 0.218)				
Sex	6	58	-0.141*	-0.142 (-0.282; -0.002)*	-0.013 (-0.180; 0.155)				
Domestic	3	15	-0.251*	-0.257 (-0.466; -0.048)*	-0.127 (-0.373; 0.118)				
Personal implications (incentives)	16	157				F(2,154) = 4.183	0.017*	0.035***	0.011**
No incentives (RC)	10	62	-0.160**	-0.161 (-0.256; -0.067)***					
Possible incentives	4	72	-0.057	-0.057 (-0.180; 0.065)	0.104 (-0.051; 0.259)				
Incentives	2	23	-0.354***	-0.370 (-0.545; -0.195)***	-0.209 (-0.408; -0.010)*				
Continent of publication	16	157				F(2,154) = 1.197	0.305	0.035***	0.018***
North America (RC)	8	93	-0.211***	-0.214 (-0.323; -0.106)***					
Europe	7	63	-0.084	-0.084 (-0.214; 0.047)	0.131 (-0.039; 0.300)				
Oceania	1	1	-0.076	-0.076 (-0.653; 0.500)	0.138 (-0.449; 0.725)				
<i>Publication bias characteristics</i>									
Intention	16	157				F(1,155) = 0.001	0.974	0.035***	0.021***
Primary goal (RC)	7	112	-0.159*	-0.160 (-0.275; -0.044)**					
Side issue	10	45	-0.156*	-0.157 (-0.278; -0.035)*	-0.003 (-0.157; 0.162)				
Year of publication	16	157	-0.065+	-0.065 (-0.134; 0.004)+	0.021 (0.012; 0.031)***	F(1,155) = 19.731	<0.001***	0.035***	0.005
Sample size	16	157	-0.203**	-0.206 (-0.346; -0.066)**	-0.48 ^c (-1.54 ^c ; 0.58 ^c)	F(1,155) = 0.794	0.374	0.034***	0.023***

Note. # studies = Number of independent studies. # ES = Number of effect sizes. Mean r = Mean effect size (r). β_0 = Mean Fisher z. β_1 = Estimated regression coefficient. CI = Confidence interval. AC = Antisocial cognition. APP = Antisocial personality pattern/traits. IM = Impression management. RC = Reference category. SM = Substance misuse.

^a Variance between the effect sizes from the same study.

^b Variance between studies.

^c 0.000 removed in Table.

+ $p < 0.10$.

* $p < 0.05$.

** $p < 0.01$.

*** $p < 0.001$.

Table 4
Results for the moderators for self-deceptive enhancement.

Moderator variables	# Studies	# ES	Mean <i>r</i>	β_0 (95% CI)	β_1 (95% CI)	Omnibus test	<i>p</i> -value	Variance level 2 ^a	Variance level 3 ^b
Self-deceptive enhancement	14	170	−0.120***	−0.121 (−0.185; −0.057)***				0.032***	0.008*
<i>Outcome characteristic</i>									
Dynamic risk factor	14	170				<i>F</i> (2,167) = 1.858	0.159	0.031***	0.007
AC (RC)	11	125	−0.108**	−0.108 (−0.178; −0.037)**					
APP	5	18	−0.213***	−0.216 (−0.330; −0.101)***	−0.108 (−0.223; 0.007) ⁺				
SM	2	27	−0.086	−0.086 (−0.237; 0.064)	0.021 (−0.145; 0.188)				
<i>Predictor characteristics</i>									
BIDR: translation	14	170				<i>F</i> (1,168) = 1.560	0.213	0.032***	0.008*
No translation (RC)	13	167	−0.128***	−0.129 (−0.194; −0.065)***					
Translation	1	3	0.074	0.074 (−0.241; 0.388)	0.203 (−0.118; 0.524)				
BIDR: version	14	170				<i>F</i> (3,166) = 0.549	0.650	0.032***	0.010**
SDE (RC)	12	117	−0.124**	−0.125 (−0.203; −0.046)**					
SDE_15	1	5	−0.088	−0.088 (−0.308; 0.132)	0.037 (−0.181; 0.255)				
DN	2	23	−0.154 ⁺	−0.155 (−0.327; 0.016) ⁺	0.031 (−0.220; 0.158)				
OCR	2	25	−0.075	−0.075 (−0.246; 0.095)	0.049 (−0.139; 0.237)				
Reliability BIDR scales	3	62	−0.204*	−0.207 (−0.409; −0.004)*	0.359 (−1.838; 1.120)	<i>F</i> (1,60) = 0.235	0.629	0.042***	0.026*
<i>Sample characteristics</i>									
Age	13	167	−0.124***	−0.125 (−0.198; −0.053)***	0.001 (−0.006; 0.008)	<i>F</i> (1,165) = 0.084	0.772	0.030***	0.008*
Gender	14	170				<i>F</i> (1,168) = 0.169	0.681	0.032***	0.009*
Male (RC)	12	164	−0.125***	−0.126 (−0.197; −0.055)***					
Mixed	2	6	−0.079	−0.079 (−0.293; 0.136)	0.047 (−0.179; 0.273)				
Setting	14	170				<i>F</i> (1,168) = 7.736	0.006**	0.033***	0.001
Community/probation (RC)	3	15	−0.266***	−0.273 (−0.392; −0.153)***					
Incarcerated/institutionalized	11	155	−0.096***	−0.096 (−0.134; −0.058)***	0.177 (0.051; 0.302)**				
Offense type	14	170				<i>F</i> (3,166) = 1.633	0.184	0.032***	0.005
Mixed (RC)	4	61	−0.098*	−0.098 (−0.185; −0.010)*					
Violent	3	26	−0.043	−0.043 (−0.169; 0.084)	0.055 (−0.098; 0.208)				
Sex	6	68	−0.126**	−0.127 (−0.217; −0.038)**	−0.030 (−0.142; 0.083)				
Domestic	3	15	−0.243***	−0.248 (−0.393; −0.104)***	−0.151 (−0.320; 0.018) ⁺				
Personal implications (incentives)	14	170				<i>F</i> (2,167) = 0.073	0.930	0.032***	0.013*
No incentives (RC)	8	76	−0.110*	−0.110 (−0.216; −0.004)*					
Possible incentives	4	72	−0.121 ⁺	−0.122 (−0.253; 0.008) ⁺	−0.012 (−0.180; 0.156)				
Incentives	2	22	−0.151	−0.152 (−0.342; 0.038)	−0.042 (−0.259; 0.175)				
Continent of publication	14	170				<i>F</i> (1,168) = 0.066	0.798	0.032***	0.010*
North America (RC)	9	116	−0.114**	−0.115 (−0.199; −0.031)**					
Europe	5	54	−0.133*	−0.134 (−0.255; −0.013)*	−0.019 (−0.166; 0.128)				
<i>Publication bias characteristics</i>									
Intention	14	170				<i>F</i> (1,168) = 0.124	0.725	0.032***	0.010**
Primary goal (RC)	7	134	−0.129**	−0.130 (−0.214; −0.046)**					
Side issue	8	36	−0.106 ⁺	−0.106 (−0.215; −0.003) ⁺	0.024 (−0.109; 0.157)				
Year of publication	14	170	−0.065*	−0.065 (−0.116; −0.015)*	0.011 (0.004; 0.018)**	<i>F</i> (1,168) = 10.315	0.002**	0.032***	0.001
Sample size	14	170	−0.222***	−0.226 (−0.350; −0.102)***	−111 ^c (−208 ^c ; −014 ^c)*	<i>F</i> (1,168) = 5.139	0.025*	0.030***	0.017**

Note. # studies = Number of independent studies. # ES = Number of effect sizes. Mean *r* = mean effect size (*r*). β_0 = Mean Fisher *z*. β_1 = Estimated regression coefficient. CI = Confidence interval. AC = Antisocial cognition. RC = Reference category. APP = Antisocial personality pattern/traits. SM = Substance misuse. SDE = Self-deceptive enhancement. DN = Denial of the negative. OCR = Over confident rigidity.

^a Variance between the effect sizes from the same study.

^b Variance between studies.

^c 0.000 removed in Table.

⁺ *p* < 0.10.

* *p* < 0.05.

** *p* < 0.01.

*** *p* < 0.001.

The dependencies between the moderators varied from small to large. For impression management, a large association was found between dynamic risk factor and year of publication ($\rho = 0.67, p < 0.001$) and small associations were found between dynamic risk factor and personal implications (Cramer's $V = 0.15, p < 0.001$) and personal implications and year of publication ($\rho = 0.03, p < 0.001$). For self-deceptive enhancement, we found a large association between sample size and year of publication ($r = 0.59, p < 0.001$) and small associations between setting and sample size ($\rho = 0.19, p < 0.001$) and setting and year of publication ($\rho = 0.11, p < 0.001$).

5. Discussion

Self-report remains an important source of information for assessing risk of recidivism in offenders. The main aim of the current study was a meta-analytic investigation of the association between socially desirable responding as measured by BIDR impression management and BIDR self-deceptive enhancement and self-report measures used in order to contribute to the identification of dynamic risk factors in offender samples.

5.1. Relation between BIDR impression management and self-deceptive enhancement, and self-report measures with content of dynamic risk factors

We found a small, yet significant, negative association between impression management ($r = -0.158$; $r = -0.07$ after correction for publication bias) and self-deceptive enhancement ($r = -0.120$) on the one hand and self-report measures with content of dynamic risk factors on the other. One way to interpret this main finding is that impression management and self-deception lowered dynamic risk scores, hence that self-report measures with content of dynamic risk factors did not succeed in countering response bias. Alternatively, low socially desirable responding scores may also reflect true variance in personality to the extent that it reflects a disregard of social conventions rather than response bias (Furnham, 1986; Piedmont, McCrae, Riemann, & Angleitner, 2000). Offenders, especially antisocial or psychopathic offenders, are known to understand but violate social norms and conventions (Glenn, Iyer, Graham, Koleva, & Haidt, 2009), and may be less inclined to answer affirmatively on social desirability items. A low score on socially desirable responding measures may thus tap into the (antisocial) personality rather than reflect response bias.

The question whether socially desirable responding scales measure response bias that is primarily determined by situational factors or an intrinsic, more permanent response style that is associated with personality traits, has a long history in psychology (McGrath, Mitchell, Kim, & Hough, 2010; Ones, Viswesvaran, & Reis, 1996; Piedmont et al., 2000). If socially desirable responding reflects a systematic response bias, then arguably methods to reduce or eliminate the bias are warranted, arguably improving the predictive validity of offender self-reports for assessment of risk or treatment outcome. Alternatively, if the tendency to portray oneself in a favorable light is an enduring quality rather than an invalidating response bias, then it should be measured and evaluated as a substantive variable in itself rather than corrected for or eliminated. Evidence from studies using offender samples has tended to support the latter view, i.e., that socially desirable responding measures a disposition or personality trait rather than a response bias (Mills & Kroner, 2006). According to this view, socially desirable responding may be inversely correlated with (self-report measures used in order to contribute to the identification of) dynamic risk factors (as is evident from the results of our meta-analysis) but is not related to recidivism risk directly. For example, Mills, Loza, and Kroner (2003) examined the relation between the BIDR and self-reported criminal risk variables in a sample of incarcerated violent offenders during a criminal risk assessment process. Additionally, an actuarial risk assessment instrument was used to objectively assess risk level. Impression management

and self-deception were significantly and inversely related to self-reported risk level. Low scores on impression management were also associated with a greater likelihood of actuarially assessed risk for criminal behavior. Instead of a response bias, socially desirable responding measures may tap an enduring disposition such that those scoring low on impression management and/or self-deception are more criminally oriented and the more likely to reoffend. Furthermore, Stevens, Tan, and Grace (2016) found that attempts to “correct” for socially desirable responding using a variety of procedures had virtually no impact on the predictive validity for recidivism (violent, sexual, general), suggesting that socially desirable responding does not pose a threat to the predictive validity of dynamic risk assessment via self-reports of child molesters. Because removal of socially desirable responding variance does not improve the correlation of self-report dynamic risk measures with recidivism, researchers have suggested that socially desirable responding scales like the BIDR or MCSDS may actually be measuring a personality trait, and not a response bias that when removed increases the accuracy of risk assessment (cf. Mills et al., 2003; Mills & Kroner, 2006). This view is consistent with a recent reinterpretation of socially desirable responding by Uziel (2010a, 2010b), who suggested that instead of response bias, measures of socially desirable responding (especially of impression management) should be regarded as interpersonally oriented self-control. In his view, people scoring high on impression management are as well behaved as they profess to be, i.e., they tend to be nice people who get along well with others, practice great self-control, and are conventional. To further help distinguish whether high scores on impression management reflect genuine agreeableness, conscientiousness, and virtue or a conscious motivation to exaggerate one's standing on these traits in order to be viewed favorably, Davis, Thake, and Weekes (2012) examined the crimes for which a large sample ($N = 11,370$) of offenders was incarcerated. They argued that if impression management is a proxy for trait agreeableness, conscientiousness, or virtue, the crimes for which the offenders have been committed should be less morally reprehensible than those committed by offenders who do not possess these traits (i.e., a negative relation of impression management with severity of crime). Alternatively, if impression management reflects a motivation to faking good, to the extent that offenders who have committed the most heinous crimes have the most to gain by appearing to possess these qualities of goodness, impression management scores should be positively correlated with crime severity. The results indicate that offenders with high impression management scores are less likely to project antisocial attitudes (cf. Mills & Kroner, 2005, 2006) but at the same time are more likely than those scoring low to have committed the most morally despicable crimes (e.g., homicide, pedophilia, incest), and are more likely to receive longer sentences. However, contrary to claims (e.g., Ones et al., 1996; Uziel, 2010a, 2010b) that those scoring high on impression management are as pleasant, agreeable and conscientious as they claim to be, according to Davis et al. (2012), offenders “scoring high on impression management are not as nice as they would have us believe, yet they may be highly motivated to present themselves as such. They say the right things, but do not always do the right things. People scoring high on impression management are highly motivated to convey an image of themselves as agreeable, dutiful, and virtuous, but this is not who they are” (Davis et al., 2012, p. 31). Whether socially desirable responding is best viewed as a response bias or personality trait, our findings do suggest that impression management scales in isolation should not be used in research to discard or adjust self-report measures used in order to contribute to the assessment of dynamic risk factors.

The small effect sizes we found in the current meta-analysis also provide some evidence that the measures used with offender populations can be accurate and valid. Our findings should help to allay concerns that “self-reports are highly vulnerable to self-presentation biases” (Hanson & Bussière, 1998, p. 349), and “inherently unreliable” (Center for Sexual Offender Management, 2007, p. 5), and that “there is no reason to assume that a suspected criminal or a prison inmate

will reveal anything of real clinical significance about himself on a questionnaire” (Hare, 1985, p. 157). Our findings underscore that the validity of self-report measures with content of dynamic risk factors is not inevitably compromised by the propensity of offenders to present themselves in a favorable light (see also Ray et al., 2013). Importantly, however, the findings are largely based on individuals who completed the measures in research settings, with no clear incentives for response distortions.

5.1.1. Presence of publication bias

There were indications of publication bias: The funnel plot asymmetry suggested that there was publication bias for studies on impression management, and publication year significantly moderated the association for both impression management and self-deceptive enhancement (decline effect), and larger effect sizes were found for self-deception in smaller samples. Correcting for publication bias using the trim and fill method indicated that the true relation between impression management and self-report measures with content of dynamic risk factors may be somewhat smaller than what we observed, but also that it is still significant and negative, indicating that the overestimation may not be substantial. We want to emphasize, however, that the trim and fill approach should only be seen as a method for sensitivity analysis rather than actually finding the values of missing effect sizes (Duval & Tweedie, 2000). Another indication that the overestimation may not be that substantial was that there was no significant differences between studies that assessed the impact of social desirability on self-report measures with content of dynamic risk factors as a primary goal versus studies that reported on this relation as a side issue. Finally, we sought out only peer-reviewed published studies for our meta-analysis. We cannot rule out the possibility that the significant effect sizes found in the present study are partly due to this form of publication bias. The only way to rule out this possibility is to thoroughly sample and identify “gray” literature, such as unpublished doctoral dissertations (but see Ferguson & Brannick, 2012). Future research may profit from extending analyses by including associations between socially desirable responding (i.e., impression management and self-deception) and self-report indices of dynamic risk factors in unpublished studies.

Although publication bias is common in many research areas, it may have especially important consequences in correctional or forensic settings. Getting unbiased estimates of effect sizes is essential to make informed decisions as to whether or not measures of social desirability should be implemented in these contexts, as well as to determine the weight that should be given to their results (e.g., by clinicians, professionals responsible for risk assessments). We encourage researchers to embrace some of the recently proposed solutions (Munafò et al., 2017), including two-stage reviews (the first stage for the design and the second for the results), pre-analysis plans (Casey, Glennerster, & Miguel, 2012), pre-registering studies (Pashler & Wagenmakers, 2012), and incentives to not bury insignificant results in file drawers (Franco, Malhotra, & Simonovits, 2014), such as creating high-status publication outlets for these studies (Franco et al., 2014).

5.2. What characteristics moderated the relation between socially desirable responding and self-report measures with content of dynamic risk factors?

The impact of impression management and self-deceptive enhancement on self-report measures with content of dynamic risk factors varied considerably between studies. Two moderators for impression management were identified: (type of) dynamic risk factor and incentives. With regard to type of dynamic risk factor, significant negative effect sizes were found if antisocial cognition and antisocial personality pattern/traits were measured, but not if substance misuse was measured. Interestingly, the effect size for self-report measures with content of antisocial personality pattern/traits was significantly larger compared

to the effect size for antisocial cognition. A possible explanation for this finding is that self-report measures used in order to contribute to the identification of antisocial personality traits, compared to questionnaires used to measure antisocial cognition, have highly transparent items, so that it is fairly obvious to the offender as to what the test is measuring, and therefore what the socially acceptable responses might be. For example, the Aggression Questionnaire (AQ; Buss & Perry, 1992) which is commonly used to assess different components of aggression of offenders, includes items such as ‘If somebody hits me, I hit back’. Using transparent measures makes it relatively easy for offenders to minimize or deny problematic behavior, and to exaggerate any positive or pro-social traits. An alternative explanation for this finding may be that the more antisocial the offender, the more willing he is to answer honestly to questions relating to antisocial behavior. Such questions might be sufficiently threatening to elicit a response bias from law-abiding individuals, they may not be for offender convicted of more serious crimes. It has also been argued that offenders might admit to relatively minor offenses to establish a foundation of credibility that they could later exploit for impression management about more major issues (Mills & Kroner, 2005).

We also found that people respond differently on questionnaires when they believe that their answers may have personal implications. We found significant negative associations between impression management and self-report measures with content of dynamic risk factors when there were clear incentives and when there were no incentives, but no significant association was found for possible incentives. The effect size for incentives ($r = -0.354$) was significantly larger compared to no incentives ($r = -0.160$) and possible incentives ($r = -0.057$). This finding suggests that socially desirable responding (impression management) is possibly better viewed as a response bias than as a personality trait. However, since only two studies were found where offenders clearly had incentives, caution should be exercised when interpreting these findings, and future research should re-examine our results as the literature grows. Nevertheless, our findings emphasize the importance for future studies to report clearly whether there are incentives and, if so, what these incentives are.

Finally, we were able to identify one moderator for self-deception: type of setting. Perhaps counter-intuitively, the effect size for offenders in the community/on probation ($r = -0.266$) appeared to be significantly larger than the effect size for incarcerated/ institutionalized offenders ($r = -0.096$). We have no real explanation for the finding that offenders in the community/on probation tend to deny their socially unfavorable qualities to themselves more than incarcerated offenders do, and hence present unwittingly a more desirable picture of their behaviors and beliefs when answering self-report measures with content of dynamic risk factors. Maybe, the fact that community offenders (in general) tend to be lower risk offenders with higher incentives (violating conditions can send them to jail) contribute to stronger effect sizes here. However, since only three studies consisted of offenders in the community/on probation, here too, findings should be interpreted with some caution.

5.3. Limitations and directions for future research

In addition to the previously discussed drawbacks relating to publication bias, a number of limitations need to be highlighted. First, this meta-analysis focused only on studies using the BIDR, since the BIDR is the only measure of socially desirable responding to tap both impression management and self-deceptive enhancement. Yet our conclusions may not apply to other measures of socially desirable responding. That said, we don’t have evidence that the use of other socially desirable responding measures in offender samples would lead to (substantial) different results (e.g., Peersen, Sigurdsson, Gudjonsson, & Gretarsson, 2004; Stevens et al., 2016). Second, given a large number of putatively dynamic risk factors have been proposed, we focused on dynamic factors as identified through meta-analytic

research. Unfortunately a well-established dynamic risk factor, procriminal companions, could not be included since none of the potential relevant articles for inclusion in our meta-analysis had used self-report measures with content related to this factor. This was also the case for education/employment, family/marital, and leisure/recreation. In addition, we had to exclude deviant sexual interests, since only one study in our meta-analysis fulfilled the inclusion criteria. Future research may be able to include (self-report measures with content of) these (or other) proposed dynamic risk factors, which may lead to different results. Third, some might correctly notice that the relationship with recidivism have not been demonstrated (yet) for all self-report measures (e.g., Fire Proclivity Scale, Identification with Fire Scale; Bumby Scales) included in our review. One can argue that, if you want to subsume different self-report measures under the concept of dynamic risk measures, these self-reports need to show criterion validity (i.e., relationship with recidivism). We think that view is too strict. It would imply that for each individual measure there needs to be meta-analytic evidence that the individual measure (1) has been shown to relate to recidivism, (2) can change, (3) and when changed results in reduced risk of recidivism; and the individual measure would else need to be excluded, even when it is clear that the measure has in fact content of the respective dynamic risk factor (as is the case for all the self-report measures included in our review). This does not detract from the fact that research on the relationship between self-report measures intended to contribute to the identification of dynamic risk factors, which are often routinely used on treatment programs (Mathie & Wakeling, 2011), and recidivism is clearly warranted. Fourth, although we used a careful procedure to select the self-report measures that we considered eligible (see Section 3.3), and interrater reliabilities with regard to (1) whether or not (yes/no) the self-report measures used in the selected studies had content related to one of the selected dynamic risk factors and (2) allocation of the measures to the selected dynamic risk factor were very good ($\kappa = 0.845$ and 0.835 , respectively), we cannot rule out the possibility that a direct link of some of the measures with a specific dynamic risk factor (e.g., impulsivity) may be too upfront. However, the selected dynamic risk factors, especially antisocial personality pattern/traits and antisocial cognition, are rather broad constructs, which reduces the chance of misclassification. For example, antisocial personality patterns/traits includes different aspects, e.g., impulsivity, anger problems, low self-control, whereas antisocial cognition includes attitudes, values, beliefs, rationalizations favorable to crime. Fifth, most of the included studies consisted of male offenders only. Therefore, our findings may not be representative of female offenders. Because females may make more efforts to create a positive impression (He et al., 2015), future research could focus on female offenders. Sixth, all of the (subscales of the) self-report measures with content of dynamic risk factors that were included in the meta-analysis were examined regardless of their psychometric properties. It has been argued (Mathie & Wakeling, 2011) that the correlations between the BIDR and self-report measures with poor psychometric properties may not be meaningful since the questionnaires themselves were found to have low reliability estimates. This possibility should be further explored in future research. Seventh, data of several moderators were based on a limited number of studies and effect sizes. Regarding the size of the data set used in multilevel meta-analytic models, it is rather difficult to determine what the minimum number of effect sizes and studies should be (Assink & Wibbelink, 2016). Statistical power increases as the number of effect sizes and studies in the data set increases. Consequently, findings based on small numbers of effect sizes and studies should be interpreted with caution. Last, in order to examine the unique influence of the variables that were identified as significant moderators in the univariate analyses, we planned to conduct multiple moderator analyses (cf. Hox, 2010; also Assink & Wibbelink, 2016). Unfortunately, since many of our moderator analyses were based on a small number of effect sizes, which means that

the statistical power to detect unique contributions of moderating variables in multiple moderator models is low (Assink et al., 2015), it was not possible to reliably examine multiple moderator models.

6. Conclusions

The above mentioned limitations notwithstanding, based on the present results, we would advance the conclusion that previous concerns that the use of self-report measures in correctional/forensic settings is inevitably compromised by socially desirable responding may have been somewhat overstated. We found a significant, but only small association between impression management and self-deceptive enhancement and questionnaires with content of dynamic risk factors. Hence, generally speaking, we did not find strong evidence for the claim that these questionnaires cannot be used accurately and valid with offenders. This finding, coupled with their ease of administration and lack of reliance on corroborative data (e.g., institutional records, criminal files), suggests that the administration of self-reports may be an efficient supplement to clinical assessments, including structured risk assessments, for measuring dynamic risk factors in offender settings. At the same time, however, our findings indicate a stronger effect for some measures (e.g., antisocial personality) and for some circumstances (e.g., incentives, setting).

References*

- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: American Psychiatric Association.
- Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct* (5th ed.). New Providence, NJ: LexisNexis Matthew Bender.
- Andrews, D. A., Zinger, I., Hoge, R. D., Bonta, J., Gendreau, P., & Cullen, F. T. (1990). Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology*, 28, 369–404. <https://doi.org/10.1111/j.1745-9125.1990.tb01330.x>.
- Assink, M., van der Put, C. E., Hoeve, M., de Vries, S. L. A., Stams, G. J. J. M., & Oort, F. J. (2015). Risk factors for persistent delinquent behavior among juveniles: A meta-analytic review. *Clinical Psychology Review*, 42, 47–61. <https://doi.org/10.1016/j.cpr.2015.08.002>.
- Assink, M., & Wibbelink, C. J. M. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology*, 12, 154–174. <https://doi.org/10.20982/tqmp.12.3.p154>.
- *Barrowcliffe, E. R., & Gannon, T. A. (2015). The characteristics of un-apprehended firesetters living in the UK community. *Psychology, Crime & Law*, 21, 836–853. <https://doi.org/10.1080/1068316X.2015.1054385>.
- Benedict, L. W., & Lanyon, R. I. (1992). An analysis of deceptiveness: Incarcerated prisoners. *Journal of Addictions & Offender Counseling*, 13, 23–32. <https://doi.org/10.1002/j.2161-1874.1992.tb00217.x>.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, England: Wiley.
- *Bowen, E., & Gilchrist, E. (2006). Predicting dropout of court-mandated treatment in a British sample of domestic violence offenders. *Psychology, Crime & Law*, 12, 573–587. <https://doi.org/10.1080/10683160500337659>.
- Buss, A., & Perry, M. (1992). The Aggression Questionnaire. *Journal of Personality and Social Psychology*, 63, 452–459. <https://doi.org/10.1037/0022-3514.63.3.452>.
- Casey, K., Glennerster, R., & Miguel, E. (2012). Reshaping institutions: Evidence on aid impacts using a preanalysis plan. *Quarterly Journal of Economics*, 127, 1755–1812. <https://doi.org/10.1093/qje/qje027>.
- Center for Sexual Offender Management (2007). *The importance of assessment in sex offender management: An overview of key principles and practices*. Silver Spring, MD: Author.
- Cheung, M. W. L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, 19, 211–229. <https://doi.org/10.1037/a0032968>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24, 349–354. <https://doi.org/10.1037/h0047358>.
- Crowne, D. P., & Marlowe, D. (1964). *The approval motive*. New York, NY: Wiley.
- *Dalbert, C., & Filke, E. (2007). Belief in a personal just world, justice judgment, and their functions for prisoners. *Criminal Justice and Behavior*, 34, 1516–1527. <https://doi.org/10.1177/0093854807306835>.

* References marked with an asterisk indicate studies included in the meta-analysis.

- *Davis, C. G., Doherty, S., & Moser, A. E. (2014). Social desirability and change following substance abuse treatment in male offenders. *Psychology of Addictive Behaviors*, 28, 872–879. <https://doi.org/10.1037/a0037528>.
- Davis, C. G., Thake, J., & Weekes, J. R. (2012). Impression managers: Nice guys or serious criminals? *Journal of Research in Personality*, 46, 26–31. <https://doi.org/10.1016/j.jrp.2011.11.001>.
- *Dutton, D. G., & Hemphill, K. J. (1992). Patterns of socially desirable responding among perpetrators and victims of wife assault. *Violence and Victims*, 7, 29–39.
- Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56, 455–463. <https://doi.org/10.1111/j.0006-341X.2000.00455.x>.
- Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. New York, NY: Dryden.
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315, 629–634. <https://doi.org/10.1136/bmj.315.7109.629>.
- Eysenck, H. J., & Eysenck, S. B. G. (1964). *The manual of the Eysenck Personality Inventory*. London, England: University of London Press.
- Ferguson, C. J., & Brannick, M. T. (2012). Publication bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods*, 17, 120–128. <https://doi.org/10.1037/a0024445>.
- Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345, 1502–1505. <https://doi.org/10.1126/science.1255484>.
- Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences*, 7, 385–400. [https://doi.org/10.1016/0191-8869\(86\)90014-0](https://doi.org/10.1016/0191-8869(86)90014-0).
- Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, 102, 74–78. <https://doi.org/10.1016/j.paid.2016.06.069>.
- Glenn, A. L., Iyer, R., Graham, J., Koleva, S., & Haidt, J. (2009). Are all types of morality compromised in psychopathy? *Journal of Personality Disorders*, 23, 384–398. <https://doi.org/10.1521/pedi.2009.23.4.384>.
- Hanson, R. K., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, 66, 348–362. <https://doi.org/10.1037/0022-006X.66.2.348>.
- Hare, R. D. (1985). A checklist for the assessment of psychopathy in criminal populations. In M. H. Ben-Aron, S. J. Hucker, & C. D. Webster (Eds.), *Clinical criminology: The assessment and treatment of criminal behavior* (pp. 157–168). Toronto, Canada: M & M Graphics.
- Hathaway, S. R., & McKinley, J. C. (1951). *The MMPI manual*. New York, NY: Psychological Corporation.
- He, J., van de Vijver, F. J., Espinosa, A. D., Abubakar, A., Dimitrova, R., Adams, B. G., ... Villieux, A. (2015). Socially desirable responding enhancement and denial in 20 countries. *Cross-Cultural Research*, 49, 227–249. <https://doi.org/10.1177/1069397114552781>.
- Hofstee, W. K. B. (2003). Structures of personality traits. In I. B. Weiner, T. Millon, & M. J. Lerner (Eds.), *Handbook of Psychology. Personality and social psychology*, Vol. 5. (pp. 231–254). Hoboken, NJ: Wiley.
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. New York, NY: Routledge.
- Huang, C. (2013). Relation between self-esteem and socially desirable responding and the role of socially desirable responding in the relation between self-esteem and performance. *European Journal of Psychology of Education*, 28, 663–683. <https://doi.org/10.1007/s10212-012-0134-5>.
- Ioannidis, J. P. (2005). Contradicted and initially stronger effects in highly cited clinical research. *JAMA*, 294, 218–228. <https://doi.org/10.1001/jama.294.2.218>.
- *Ilycan, S., Sommer, J. M., Kini, S., & Babcock, J. C. (2015). Collateral report of psychopathy: Convergent and divergent validity of the psychopathic personality inventory-short form. *The Journal of Forensic Psychiatry & Psychology*, 26, 476–492. <https://doi.org/10.1080/14789949.2015.1018926>.
- *Keown, K., Gannon, T. A., & Ward, T. (2010). What's in a measure? A multi-method study of child sexual offenders' beliefs. *Psychology, Crime & Law*, 16, 125–143. <https://doi.org/10.1080/10683160802622022>.
- Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, 22, 2693–2710. <https://doi.org/10.1002/sim.1482>.
- *Kroner, D. G., Boer, D. P., & Mills, J. F. (2004). Explaining rape-supportive attitudes among rapists. *American Journal of Forensic Psychology*, 22, 65–76.
- *Kroner, D. G., Mills, J. F., Yessine, A. K., & Hemmati, T. (2004). The generalized instructional set of the Criminal Attribution Inventory (CRAI): Socially desirable responding and faking. *International Journal of Offender Therapy and Comparative Criminology*, 48, 360–372. <https://doi.org/10.1177/0306624X04263452>.
- Kroner, D. G., & Weekes, J. R. (1996). Balanced Inventory of Desirable Responding: Factor structure, reliability, and validity with an offender sample. *Personality and Individual Differences*, 21, 323–333. [https://doi.org/10.1016/0191-8869\(96\)00079-7](https://doi.org/10.1016/0191-8869(96)00079-7).
- Kühberger, A., Fritz, A., & Scherndl, T. (2014). Publication bias in psychology: A diagnosis based on the correlation between effect size and sample size. *PLoS One*, 9(9), e105825. <https://doi.org/10.1371/journal.pone.0105825>.
- Lehmann, R. J. B., Goodwill, A. M., Hanson, R. K., & Dahle, K. -P. (2014). Crime scene behaviors indicate risk-relevant propensities of child molesters. *Criminal Justice and Behavior*, 41, 1008–1028. <https://doi.org/10.1177/0093854814521807>.
- Li, A., & Bagger, J. (2006). Using the BIDR to distinguish the effects of impression management and self-deception on the criterion validity of personality measures: A meta-analysis. *International Journal of Selection and Assessment*, 14, 131–141. <https://doi.org/10.1111/j.1468-2389.2006.00339.x>.
- Li, A., & Bagger, J. (2007). The Balanced Inventory of Desirable Responding (BIDR): A reliability generalization study. *Educational and Psychological Measurement*, 67, 525–544. <https://doi.org/10.1177/0013164406292087>.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- *Mann, R., Webster, S., Wakeling, H., & Marshall, W. (2007). The measurement and influence of child sexual abuse supportive beliefs. *Psychology, Crime & Law*, 13, 443–458. <https://doi.org/10.1080/10683160601061141>.
- Marshall, W. L., & Fernandez, Y. M. (2001). Phallometry in forensic practice. *Journal of Forensic Psychology Practice*, 1, 77–87. https://doi.org/10.1300/J158v01n02_06.
- *Mathie, N. L., & Wakeling, H. C. (2011). Assessing socially desirable responding and its impact on self-report measures among sexual offenders. *Psychology, Crime & Law*, 17, 215–237. <https://doi.org/10.1080/10683160903113681>.
- McGrath, R. E., Mitchell, M., Kim, B. H., & Hough, L. (2010). Evidence for response bias as a source of error variance in applied assessment. *Psychological Bulletin*, 136, 450–470. <https://doi.org/10.1037/a0019216>.
- Meijer, E. H., & Verschuere, B. (2017). Deception detection based on neuroimaging: Better than the polygraph? *Journal of Forensic Radiology and Imaging*, 8, 17–21. <https://doi.org/10.1016/j.jofri.2017.03.003>.
- *Mills, J. F., & Kroner, D. G. (2005). An investigation into the relationship between socially desirable responding and offender self-report. *Psychological Services*, 2, 70–80. <https://doi.org/10.1037/1541-1559.2.1.70>.
- *Mills, J. F., & Kroner, D. G. (2006). Impression management and self-report among violent offenders. *Journal of Interpersonal Violence*, 21, 178–192. <https://doi.org/10.1177/0886260505282288>.
- Mills, J. F., Loza, W., & Kroner, D. G. (2003). Predictive validity despite social desirability: Evidence for the robustness of self-report among offenders. *Criminal Behaviour and Mental Health*, 13, 140–150. <https://doi.org/10.1002/cbm.536>.
- Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., ... Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1, 0021. <https://doi.org/10.1038/s41562-016-0021>.
- *O'Brien, M. D., & Webster, S. D. (2007). The construction and preliminary validation of the Internet Behaviours and Attitudes Questionnaire (IBAQA). *Sexual Abuse: Journal of Research and Treatment*, 19, 237–256. <https://doi.org/10.1007/s1194-007-9057-0>.
- Ones, D. S., Viswesvaran, C., & Reis, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, 81, 660–679. <https://doi.org/10.1037/0021-9010.81.6.660>.
- Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7, 528–530. <https://doi.org/10.1177/1745691612465253>.
- Patrick, C. J., Bradley, M. M., & Lang, P. J. (1993). Emotion in the criminal psychopath: Startle reflex modulation. *Journal of Abnormal Psychology*, 102, 82–92. <https://doi.org/10.1037/0021-843X.102.1.82>.
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46, 598–609. <https://doi.org/10.1037/0022-3514.46.3.598>.
- Paulhus, D. L. (1988). *Assessing self-deception and impression management in self-report: The balanced inventory of desirable responding*. Unpublished manual Vancouver, Canada: University of British Columbia.
- 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 psychology attitudes* (pp. 17–59). New York, NY: 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 education measurement* (pp. 49–69). Mahwah, NJ: Erlbaum.
- Peersen, M., Sigurdsson, J. F., Gudjonsson, G. H., & Gretarsson, S. J. (2004). Predicting re-offending: A 5-year prospective study of Icelandic prison inmates. *Psychology, Crime & Law*, 10, 197–204. <https://doi.org/10.1080/10683160310001614789>.
- Perinelli, E., & Gremigni, P. (2016). Use of social desirability scales in clinical psychology: A systematic review. *Journal of Clinical Psychology*, 72, 534–551. <https://doi.org/10.1002/jclp.22284>.
- 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. <https://doi.org/10.1037/0022-3514.78.3.582>.
- R Core Team (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing (Retrieved from <http://www.R-project.org/>).
- Ray, J. V., Hall, J., Rivera-Hudson, N., Poythress, N. G., Lilienfeld, S. O., & Morano, M. (2013). The relation between self-reported psychopathic traits and distorted response styles: A meta-analytic review. *Personality Disorders: Theory, Research, and Treatment*, 4, 1–14. <https://doi.org/10.1037/a0026482>.
- *Richards, H. J., & Pai, S. M. (2003). Deception in prison assessment of substance abuse. *Journal of Substance Abuse Treatment*, 24, 121–128. [https://doi.org/10.1016/S0740-5472\(02\)00356-2](https://doi.org/10.1016/S0740-5472(02)00356-2).
- Rogers, R., & Bender, S. D. (2003). Evaluation of malingering and deception. In I. B. Weiner, & A. M. Goldstein (Eds.), *Handbook of Psychology. Forensic psychology*, Vol. 11. (pp. 109–129). Hoboken, NJ: Wiley.
- Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin*, 118, 183–192. <https://doi.org/10.1037/0033-2909.118.2.183>.
- Schmidt, A. F., Banse, R., & Imhoff, R. (2015). Indirect measures in forensic contexts. In T. M. Ortner, & F. J. R. van de Vijver (Eds.), *Behavior-based assessment in psychology: Going beyond self-report in the personality, affective, motivation, and social domains* (pp. 173–194). Göttingen: Hogrefe.
- Schooler, J. (2011). Unpublished results hide the decline effect. *Nature*, 470, 437. <https://doi.org/10.1038/470437a>.
- Seifert, K., Boulas, J., Huss, M. T., & Scalora, M. T. (2015). Response bias on self-report measures of sexual fantasies among sexual offenders. *International Journal of Offender Therapy and Comparative Criminology*, 61, 269–281. <https://doi.org/10.1177/0306624X15593748>.

- *Seto, M. C., & Barbaree, H. E. (1993). Victim blame and sexual arousal to rape cues in rapists and nonoffenders. *Annals of Sex Research*, 6, 167–183. <https://doi.org/10.1007/BF00849559>.
- *Simourd, D. J., & Mamuja, J. M. (2000). The Hostile Interpretations Questionnaire: Psychometric properties and construct validity. *Criminal Justice and Behavior*, 27, 645–663. <https://doi.org/10.1177/0093854800027005006>.
- Stevens, C. D. J., Tan, L., & Grace, R. C. (2016). Socially desirable responding and psychometric assessment of dynamic risk in sexual offenders against children. *Psychology, Crime & Law*, 22, 420–434. <https://doi.org/10.1080/1068316X.2015.1120868>.
- Stöber, J., Dette, D. E., & Musch, J. (2002). Comparing dichotomous and continuous scoring of the balanced inventory of desirable responding. *Journal of Personality Assessment*, 78, 370–389. https://doi.org/10.1207/S15327752JPA7802_10.
- Suchotzki, K., Verschuere, B., van Bockstaele, B., Ben-Shakhar, G., & Crombez, G. (2017). Lying takes time: A meta-analysis on reaction time measures of deception. *Psychological Bulletin*, 143, 428–453. <https://doi.org/10.1037/bul0000087>.
- Sutton, A. J., Duval, S. J., Tweedie, R. L., Abrams, K. R., & Jones, D. R. (2000). Empirical assessment of effect of publication bias on meta-analyses. *British Medical Journal*, 320, 1574–1577. <https://doi.org/10.1136/bmj.320.7249.1574>.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson/Allyn & Bacon.
- Tan, L., & Grace, R. C. (2008). Social desirability and sexual offenders: A review. *Sexual Abuse: A Journal of Research and Treatment*, 20, 61–87. <https://doi.org/10.1177/1079063208314820>.
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133, 859–883. <https://doi.org/10.1037/0033-2909.133.5.859>.
- Turner, R. M., Bird, S. M., & Higgins, J. P. (2013). The impact of study size on meta-analyses: Examination of underpowered studies in Cochrane reviews. *PLoS One*, 8, e59202. <https://doi.org/10.1371/journal.pone.0059202>.
- Uziel, L. (2010a). Rethinking social desirability scales from impression management to interpersonally oriented self-control. *Perspectives on Psychological Science*, 5, 243–262. <https://doi.org/10.1177/1745691610369465>.
- Uziel, L. (2010b). Look at me, I'm happy and creative: The effect of impression management on behavior in social presence. *Personality and Social Psychology Bulletin*, 36, 1592–1602. <https://doi.org/10.1177/0146167210386239>.
- van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods*, 45, 576–594. <https://doi.org/10.3758/s13428-012-0261-6>.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36, 1–48. <https://doi.org/10.18637/jss.v036.i03>.
- Vispoel, W. P., & Kim, H. Y. (2014). Psychometric properties for the Balanced Inventory of Desirable Responding: Dichotomous versus polytomous conventional and IRT scoring. *Psychological Assessment*, 3, 878–891. <https://doi.org/10.1037/a0036430>.
- Vispoel, W. P., & Tao, S. (2013). A generalizability analysis of score consistency for the balanced inventory of desirable responding. *Psychological Assessment*, 25, 94–104. <https://doi.org/10.1037/a0029061>.
- *Visu-Petra, G., Borlean, C., Chendran, L., & Buş, I. (2008). An investigation of antisocial attitudes, family background and moral reasoning in violent offenders and police students. *Cognition, Brain, Behavior*, 12, 143–159.
- *Wakeling, H. C., & Barnett, G. (2011). Measuring grievance thinking in sexual offenders: The revised Dissipation-Rumination Scale. *Journal of Sexual Aggression*, 17, 273–289. <https://doi.org/10.1080/13552600.2010.495422>.