Running Head: SOCIAL DESIRABILITY AND DOPING

# Substantive and Methodological Considerations of Social Desirability for Doping in Sport

Daniel F. Gucciardi[1], Geoffrey Jalleh[2], and Robert J. Donovan[2]

[1]*School of Physiotherapy and Exercise Science, Curtin University*

[2]*Centre for Behavioral Research in Cancer Control, Curtin University*

Corresponding author: Daniel Gucciardi, School of Physiotherapy and Exercise Science, Curtin University, GPO Box U1987, Perth, Australia, 6845. Phone: +61 8 9266 3653, Fax: +61 8 9266 3699, Email: daniel.f.gucciardi@gmail.com

Book chapter in V. Barkoukis, L. Lazuras, & H. Tsorbatzoudis (2016) (Eds.), *The psychology of doping in sport* (pp. 78-92). New York, NY: Routledge.

**Abstract**

Socially desirable responding (SDR) refers to the propensity for some people to excessively overestimate desirable attributes or actions and underestimate or deny undesirable qualities or behaviors in self-report scales (Paulhus, 2002). Because this construct is represented by a deliberate *or* unintentional act to present oneself in the best possible light for others, particularly for attributes or behaviors that are highly valued or congruent with social norms, it remains an important issue for scholars interested in the psycho-social aspects of doping in sport. The purposes of this chapter are to provide readers with an overview of substantive and methodological considerations of SDR, and assess existing doping research against these issues. To facilitate future research on doping in sport that involves self-reports, we discuss a variety of proactive (e.g., bogus pipeline) and post hoc techniques (e.g., latent variable modeling) that have been proposed to detect and control for SDR, and provide examples of how doping scholars might integrate these methods into their research.

# Substantive and Methodological Considerations of Social Desirability for Doping in Sport

Have you taken illicit drugs? Are you mentally tough? How often do you display aggressive acts towards your opponents in sport? Would you assist a stranger in need of help? These questions and many others like them are central to several academic disciplines within the social sciences. Asking people to respond truthfully to these questions, however, is a complex endeavor given the propensity for some people to excessively overestimate desirable attributes or actions and underestimate or deny undesirable qualities or behaviors on self-report scales (Paulhus, 2002). Referred to as socially desirable responding (SDR), this response distortion has the potential to contaminate the accuracy of peoples' self-reports and therefore the validity of empirical findings. In this chapter, we will provide an overview of the literature on SDR as it pertains to the social science of use of banned performance-enhancing substances or methods in sport, otherwise referred to as doping.

## Substantive Considerations of SDR

Several definitions of social desirability or SDR have appeared over the years (for reviews, see Holden & Passey, 2009; Paulhus, 2002; Uziel, 2010). Although there does not appear to be a universally agreed upon definition of this concept, common themes suggest that SDR is best described as a conscious or unconscious attempt to distort responses by overestimating positive or underestimating negative qualities or behaviors. A key point is that SDR can be either a deliberate *or* unintentional act to present oneself in the best possible light for others (Paulhus, 2002). The former is concerned with a conscious effort to distort responses in order to present oneself in a desirable manner to others (i.e., impression management), whereas the latter is an unconscious tendency to view oneself in a positive light through biased responses

towards socially desirable qualities or behaviors that are considered true by the individual (i.e., self-deception enhancement). In either case, the process results in individuals to some degree distorting their responses to convey a desirable impression on others and therefore influencing their attitudes, beliefs or behavior (Kuncel & Tellegen, 2009).

The issue of whether SDR is best conceptualized as trait (substance) or error (style) variance has been the subject of considerable debate for over 40 years (e.g., Block, 1965; Edwards, 1967). On the one hand, the observed covariation between SDR scales and personality traits such as neuroticism, agreeableness, and conscientiousness (e.g., Ones, Viswesvaran, & Reiss, 1996; Paulhus, 1991) supports an interpretation that SDR represents a substantively meaningful trait (see also, Smith & Ellingson, 2002). To truly understand valid variance that belongs to the scales of substantive interest, scholars have proposed that it is necessary to remove SDR from measures of (un)desirable traits or behaviors (McCrae & Costa, 1983). The dilemma, however, is to accurately delineate what a person's score on an inherently (un)desirable trait or behavior means if the items designed to capture the construct are free from any desirability component (cf. Paunonen & LeBel, 2012). Doping is an inherently undesirable behavior and is therefore a prime example of this challenge because self-reports dominate methodological approaches in this research area (Ntoumanis, Ng, Barkoukis, & Backhouse, 2014). Nevertheless, without estimating and controlling for SDR, speculations will remain as to whether individuals' true level of the trait or behavior, their tendency to distort responses towards the desirable pole, or some combination of both accounts for respondents' high or low scores on an (un)desirable construct (Paunonen & LeBel, 2012).

Conceptualized as a style rather than a substantive trait, SDR is thought to comprise a source of artificial variance (i.e., systematic bias or error variance) and covariance (i.e., common method) attributable to social desirability and not the intended construct, and therefore threatens construct validity (e.g., factor structure) and interpretations of relations among two or more variables (Chan, 2012). If SDR does affect criterion-related validities, it could do so in two primary ways. First, it could moderate the relations among predictor and outcome variables, such that different criterion-related validities would be observed for groups of differing levels of SDR (e.g., low, moderate, high). In other words, high levels of SDR results in a loss of meaningful information or a large degree of error variance. However, this expectation has been questioned in personality assessments (Hough, Eaton, Dunnette, Kamp, & McCloy, 1990) and applicant testing settings (Ones, Viswesvaran, & Schmidt, 1993; see also, McGrath, Mitchell, Kim, & Hough, 2010). Second, SDR could serve as a suppressor variable, such that it inflates the association between two or more variables. In other words, SDR shifts one's 'true' response on the predictor variable towards the desirable end. Meta-analytic data in which zero-order and SDR-partialed correlations between personality and job performance are compared does not support a suppressor explanation (Li & Bagger, 2006; Ones et al., 1996). It is important to note that the limited support for these interpretations regarding criterion-related validities might be a result of the validity of the SDR scales rather than the effect itself (Uziel, 2010). Alternatively, it may be that everyone in a sample engaged in SDR to a similar degree (Rosse, Stecher, Miller, & Levin, 1998).

Both the content and context of the survey can influence the likelihood of SDR (Chan, 2012; McFarland & Ryan, 2000; Paunonen & LeBel, 2012). For example, topics which are sensitive (e.g., criminal offences, drug use) or have a high degree of perceived social appeal

(e.g., mental toughness, altruism) are likely to be more susceptible to SDR than issues in which there is an absence of any clear sensitivity (e.g., use of nutritional supplements) or desirable social norms (e.g., extraversion versus neuroticism). There might also be an elevated degree of SDR in those contexts where respondents can be identified or the consequences of the assessment are important or valued (e.g., job application, talent identification tests). Finally, if people possess a high amount of a desirable attribute or regularly engage in a desirable behavior (i.e., true level is high), then there should be little need for them to engage in SDR to present themselves in the best possible light for others. Simply put, contamination of self-reports from SDR may not always be a major concern with some people, contexts or topics (Chan, 2012; Paunonen & LeBel, 2012; Peidmont, McCrae, Riemann, & Angleitner, 2000).

## Methodological Considerations of SDR

As science is largely concerned with evidence, one requires data demonstrating that self-reports depart from reality to support any allegation that responses are biased or distorted in some way (Paulhus, 2002). Unsurprisingly, scholars have devoted considerable effort towards developing and evaluating ways to assess or detect and, if necessary, control for such distortions in self-reports (Nederhof, 1985; Paulhus, 1991; see also, Tourangeau & Yan, 2007; Krumpal, 2013). In this section, we review both proactive and post hoc approaches to minimizing, detecting and controlling for SDR, and provide examples pertinent to the study of doping in sport. Interested readers are referred to other chapters of this book for detailed overviews of techniques not discussed here including the randomized response technique and implicit measurements.

## Proactive Approaches to Minimizing SDR

***Techniques employed during scale development.*** The social desirability of items (or item desirability) is often a primary focus during scale development. Typically, the aim is to develop items with SDR in mind, such that one identifies and eliminates potentially problematic items during scale construction. Perceived desirability of each item can be obtained by directly asking participants (e.g., Kuncel & Tellegen, 2009; Randall & Fernandes, 1991) or indirectly by correlating items with a presumed measure of SDR (e.g., Kam, 2013). A number of psychometric scales have been developed to assess SDR and published in peer-reviewed outlets. The most commonly employed questionnaires include the Marlowe-Crowne Social Desirability Scale (M-CSDS; Crowne & Marlowe, 1960; see also, Reynolds, 1982; Strahan & Gerbasi, 1972), Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1991), and Social Desirability Scale-17 (SDS-17; Stöber, 2001). Both the M-CSDS and SDS-17 are unidimensional and capture behaviors that are socially approved or desirable yet infrequent, as well as socially disapproved or undesirable but frequent. In contrast, the BIDR measures the dualistic model of SDR (Paulhus, 1991, 2002), including items to capture impression management (i.e., conscious deception to make a favorable impression on others) or self-deceptive enhancement (i.e., unconscious deception to protect self-esteem).

Despite the intuitive appeal, the elimination of items that capture a target concept can result in a reduction in the validity of the scale (Johnson, 2004). To alleviate this concern, researchers might focus their attention on developing items that maximize content validity and minimize SDR through careful attention to the content of items during their development

(Chan, 2012; Holden, 2010). For example, it is important to minimize value-laden content (e.g., "Doping is a trivial issue in my sport") as it can lead to responses based on both the content and tone of the statement. The content validity index (Lynn, 1986) represents one approach that researchers might find useful in making an informed judgment on the representativeness and clarity of items using assessments from expert raters (see also, Polit & Beck, 2006). Nevertheless, some attributes or behaviors are inherently desirable (e.g., mental toughness, altruism) or undesirable (e.g., neuroticism, doping) and therefore it is important to recognize that it might be difficult to minimize SDR through scale development processes for these topics (Paunonen & LeBel, 2012).

***Techniques employed during scale completion*.** Rational techniques are designed to minimize the extent to which individuals respond to items in a desirable fashion through measures that are built into the scale itself (Paulhus, 1991). Researchers can employ forced-choice items, which require respondents to choose between two alternatives with an equal degree of un/desirableness, or single statements that are neutral in social desirability (Nederhof, 1985; see also, Converse et al., 2010). However, a limitation of these techniques is that the researcher's judgment of social desirability may differ from those of participants and therefore SDR may still have an influence on their responses. Demand reduction techniques such as maximizing anonymity and confidentiality, and reminding participants that honest responses are important to obtain useful feedback, are considered effective for minimizing SDR (Chan, 2012; Paulhus & Vazire, 2007). However, the evidence regarding the usefulness of anonymity for SDR is equivocal (e.g., Holden, Magruder, Stein, Sitarenios, & Sheldon, 1999; Lautenschlager & Flaherty, 1990), and there are situations in which the desire to distort one's responses would be expected to be exceptionally high regardless of

such reminders (e.g., job or school related assessments, illegal activities; Paunonen & LeBel, 2012). It is important that this feature of the survey administration process is tested as it pertains to doping research; for example, athletes could be randomized to complete a survey package with or without the traditional instructional set (i.e., honesty, confidentiality, anonymity). Nevertheless, there is evidence that people are more likely to report socially undesirable behaviors or activities when the survey is self-administered rather than interviewer-administered (Tourangeau & Yan, 2007).

Another methodological approach that can be applied during the testing phase is to compare individuals who complete a survey package as per normal (i.e., honest responses) with people who have a strong incentive or motivation to convey a desirable impression on others (e.g., applicants versus non-applicants), or have been instructed to "fake good" on the test (i.e., report the most desirable impression). For example, doping scholars might randomize athletes to complete a survey in a desirable condition (e.g., scholarship application for institute of sport) or as per normal. There is ample evidence that instructions to fake good in experimental settings change mean scores in predictable directions, with social desirability scales particularly susceptible to response distortion under faking conditions (Viswesvaran & Ones, 1999). Such changes in scale scores can have dramatic effects on test validity (e.g., Holden, 2007; Jackson, Wroblewski, & Ashton, 2000). Nevertheless, although people can consciously distort their responses on self-reports to make a favorable impression on others in experimental settings (Ones et al., 1996), it does not necessarily mean they do it to the same degree (or at all) in real-world settings (Hough et al., 1990).

The bogus pipeline technique is designed to maximize motivation to provide accurate self-reports of socially sensitive behaviors by informing participants that dishonest responses can be detected by an objective device such as a polygraph (Jones & Sigall, 1971). Participants are typically connected to equipment and led to believe that the machines assess the truthfulness of peoples' verbal statements through objective psycho-physiological indicators that cannot be consciously manipulated by the respondent. For example, doping scholars could integrate heart rate measures as part of the data collection process, and inform athletes that their pulse (relative to baseline) provides an indication as to whether or not they are telling the truth in their survey responses. One variant of this technique is to check public records (e.g., voting activity; Hanmer, Banks, & White, 2014). Although requiring specialized equipment in most instances (e.g., polygraph, galvanic skin response), this technique has been shown to reduce SDR and encourage honest responses in an attempt to avoid being detected as a liar (Roese & Jamieson, 1993).

Specialized equipment can also offer more objective methods of detecting SDR than self-report scales. Eye-tracking equipment, for example, has been used to gain insight into the response processes during conditions in which individuals are instructed to respond honestly or fake good (van Hooft & Born, 2012). Using a repeated-measures design, van Hooft and Born (2012) found that response latencies are 0.25 sec faster when responding in a socially desirable manner when compared to answering honestly ($d$ = .23). Furthermore, eye tracking data revealed that SDR involved approximately 1 eye fixation less per scale item, greater attention to extreme response options, and more immediate movement to the extreme response option after having read the scale item when compared with the honest condition. These findings suggest that SDR involved a less cognitively complex process of information

retrieval when compared with honest responses, such that there was little emphasis on the retrieval of accurate information relating to self-referenced information but rather a focus on the instructional set and desirability of the item (van Hooft & Born, 2012).

**Post Hoc Approaches to Detecting and Controlling SDR**

A number of approaches have been proposed to estimate or control for the effects of SDR on construct and criterion-related validities (Chan, 2012; Paunonen & LeBel, 2012; Ziegler & Buehner, 2009; for a review, see Podsakoff, MacKenzie, & Podsakoff, 2012). Perhaps the most common approach is to administer a scale presumed to capture SDR alongside tools that target the constructs of substantive focus; people who report high SDR scores are assumed to have distorted their responses to the primary constructs of interest. Typically, researchers statistically test this assumption by ascertaining if SDR moderated or suppressed criterion-related validities. Despite its intuitive appeal, it is ineffective to use self-report measures of SDR to statistically control or correct for SDR bias (Li & Bagger, 2006; Ones et al., 1996) because they typically remove valid variance (Paulhus & Vazire, 2007), particularly when such scales do not adequately capture SDR (Uziel, 2010; see also, Paunonen & LeBel, 2012). Given that SDR scales have meaningful correlations with personality traits such as neuroticism, agreeableness, and conscientiousness (Ones et al., 1996; Paulhus, 1991), it is difficult to conclude that extreme scores on SDR scales represent response bias rather than extreme personalities (Ziegler & Buehner, 2009).

When conceptualized as spurious measurement error or systematic variance that is caused by a person's perception of situational demands or characteristics (Schmidt, Le, & Ilies, 2003),

statistical techniques commensurate with common method variance can be applied to model the effects of SDR (Ziegler & Buehner, 2009). Common method variance is typically represented as a latent variable within a structural equation modeling framework whereby all items or scales distorted by SDR load on their respective trait variable and a latent variable representing common method (for a review, see Podsakoff et al., 2012). Nevertheless, although method variance could be due to SDR (Ziegler & Buehner, 2009), one cannot rule out other explanations such as response sets or styles (e.g., Wetzel, Carstensen, & Böhnke, 2013) or acquiescence (e.g., Rammstedt & Kemper, 2011). The use of marker variables has been recommended as a method by which to provide insight into the nature of the latent variable (Podsakoff et al., 2012). In the case of social desirability, correlating a scale presumed to capture SDR with the latent variable does not necessarily mean this response bias was modeled; rather, it could well mean that some higher-order personality construct was extracted given the substantive overlap evidence in previous research (Ones et al., 1996; Paulhus, 1991). An alternative approach might be the use of bogus items, which are designed on the surface to appear similar to real tasks or behaviors but are actually fake or non-existent (e.g., Dwight & Donovan, 2003); thus, no respondent should report engaging in such fictitious behaviors. An example of a bogus item for athletes might be "How often do you utilize mesotechnic stretching during a warm-up activity?" Bogus items are embedded within the survey and subsequently used as a proxy for SDR. Researchers might also consider the integration of statistical and experimental methodologies. For example, Ziegler and Buehner (2009) demonstrated the usefulness of an approach whereby two groups completed measurements at two time points, with one group instructed to provide honest assessments for both ratings (control group) and the other group asked to "fake good" at the second assessment (experimental group). Common method variance was modeled at both time points

within a latent variable framework and allowed for the separation of trait and SDR variances, alongside the reversal of the effects of SDR on means and covariance structure.

Monte Carlo simulations have also been used to examine the effects of SDR (e.g., Berry & Sackett, 2009; Komor, Brown, Komar, & Robbie, 2008; Paunonen & LeBel, 2012). Rather than estimating or manipulating variables using real participants, Monte Carlo simulations enable researchers to sample means from a given population distribution that represents people's scores on constructs of interest, and assess the performance of the model by varying specific features (e.g., sample size, effect sizes, covariance matrices) that may have bearing on the interpretation of substantive outcomes (Paxton, Curran, Bollen, Kirby, & Chen, 2001). For example, Paunonen and LeBel (2012) varied the amount of SDR to trait scores of a bipolar personality inventory (e.g., honesty-dishonesty) and examined the effects on three levels of criterion-related validity ($\rho_{XY}$ = .20, .40, and .60) in three different sample sizes ($n$ = 90, 180, and 270). They found that there was little impact of SDR on criterion-related validity, even when some respondents' test scores were highly distorted due to response bias. Consistent with meta-analytic data (Hough et al., 1990; McGrath et al., 2010; Ones et al., 1993; Ones et al., 1996), they also found little evidence to support a moderation or suppressor effect of SDR.

## SDR and Doping in Sport: Current Status and Future Directions

Contamination of self-reports from SDR may not always be a major concern with some samples or topics (Chan, 2012; Peidmont et al., 2000). However, doping is a sensitive issue and therefore represents a behavior that individuals might be highly motivated to

intentionally minimize or deny when asked about their involvement in this activity. Thus, the presence of SDR may pose a threat to the validity of findings in which individuals are asked to self-report key variables of interest (e.g., attitudes to doping, engagement in doping). A recent meta-analysis of doping research revealed that scholars have relied primarily on self-report data (Ntoumanis et al., 2014). Despite the potential threats of this response bias to the validity of empirical findings, there have been few attempts to delineate an understanding of the effects of SDR for research on doping in sport. As depicted in Table 1, we located only seven published studies in which the influence of SDR for self-reports of doping-related variables has been considered.

[*insert Table 1 about here*]

Our review of the doping literature with regard to SDR revealed two noteworthy considerations. First, researchers have relied on post hoc approaches to detect and control for SDR in cross-sectional designs, exclusively by the use of presumed SDR scales. This approach is not ideal (Li & Bagger, 2006; Ones et al., 1996) because there is an assumption that they validly capture the SDR construct (Uziel, 2010). Particularly when SDR scales correlate with substantive variables, correcting for this construct typically removes valid variance (Paulhus & Vazire, 2007). Clearly, there is a need for researchers to utilize alternative methodological (e.g., faking paradigm) and statistical techniques (e.g., Monte Carlo simulations) to enhance our understanding of SDR as it pertains to doping research.

Second, there is conflicting evidence in the available literature. With the exception of one null finding (Gucciardi et al., 2010), SDR has been shown to be inversely associated with

doping-related variables (e.g., attitudes, intentions, susceptibility) in adolescent (Barkoukis et al., 2014) and adult Greek athletes (Barkoukis et al., 2011; Lazuras et al., 2010), college athletes (Petróczi, 2007; Petróczi & Nepusz, 2011) and competitive athletes (Whitaker et al., in press). These findings suggest that SDR shares some conceptual overlap with substantively important variables. Yet the evidence regarding the influence of SDR on criterion-related validities is equivocal. Controlling for SDR did not significantly alter the association between doping intentions and variables such as attitudes, subjective norms, and perceived behavioral control (Barkoukis et al., 2014; Lazuras et al., 2010), thereby providing null results with regard to a suppressor effect interpretation. However, SDR has been shown to moderate the relation between attitudes to doping and doping susceptibility (Gucciardi et al., 2011). Similarly, strong and significant relations among study variables (e.g., attitudes, internal deterrence factors) were observed for athletes high in SDR, but these associations were low and non-significant in a low SDR group (Petróczi & Nepusz, 2011). There is also evidence to suggest that the inclusion of SDR in structural equation modeling can enhance the variance explained and model-data fit (Petróczi & Nepusz, 2011). Such inconsistencies in evidence could be clarified through designs that provide insight into cause-and-effect, that is, the manipulation of features related to the content and context of the study. One could compare theoretical sequences under varying conditions where SDR may be minimized or maximized; for example, anonymous surveys excluding all personal information versus assessments in which participants provide such information or responses that increase the likelihood that their data could be identified (e.g., birthdate, linked code for multiple assessment points) (e.g., Smith & Ellingson, 2002).

Validation studies are considered the gold standard for assessing whether or not responses depart from reality because they compare self-reports with the 'true' value from external sources such as medical or administrative records (Krumpal, 2013). Though not directly targeting SDR, one study to date has revealed an interesting methodological finding. In a sample of competitive athletes, evidence of doping behavior from a bioanalysis of hair samples was denied in athletes' self-reported data; moreover, self-reported doping behavior was not reflected in bioanalysis of hair samples (Petróczi et al., 2011). Alternatively, researchers might sample athletes who have and have not been convicted of an anti-doping rule violation from formal records (e.g., Australian Sports Anti-Doping Authority), and ascertain if self-reports converge with these records. There is evidence that the likelihood of SDR is low when the accuracy of item responses is verifiable (Becker & Colquitt, 1992). Nevertheless, there is an implicit assumption in this approach that formal records are reliable estimates of the target behavior and this assumption might not hold with doping in sport. Indeed, a recent probability and cost analysis of worldwide data on positive doping tests from 93 sports revealed that the likelihood of catching a doping athlete is only 2.9% (Hermann & Henneberg, 2013).

## Conclusion

In this chapter, we have reviewed the literature on SDR with regard to its conceptualization, and methods to assess or detect and, if necessary, control for such distortions in self-reports. Although much of the evidence supports an interpretation that SDR represents a substantively meaningful trait (e.g., Ones et al., 1996; Smith & Ellingson, 2002), one cannot completely rule out its influence as a response bias for construct and criterion validity (e.g., Holden, 2007; Viswesvaran & Ones, 1999). A variety of proactive (e.g., bogus pipeline) and post hoc

techniques (e.g., latent variable modeling) have been proposed to detect and control for SDR, yet researchers interested in the psycho-social aspects of doping have not yet capitalized on these methods. Technological (e.g., eye-tracking) and statistical advancements (e.g., Monte Carlo simulations) represent avenues for future research on the effects of SDR for self-reported doping attributes or behaviors. It is our hope that this chapter will provide the foundation upon which to stimulate future research in this important area.

## References

Barkoukis, V., Lazuras, L., & Tsorbatzoudis, H. (2014). Beliefs about the causes of success in sports and susceptibility for doping use in adolescent athletes. *Journal of Sports Sciences, 32,* 212-219. doi: 10.1080/02640414.2013.819521

Barkoukis, V., Lazuras, L., Tsorbatzoudis, H., & Rodafinos, A. (2011). Motivational and sportspersonship profiles of elite athletes in relation to doping behavior. *Psychology of Sport and Exercise, 12,* 205-212. doi: 10.1016/j.psychsport.2010.10.003

Block, J. (1965). *The challenge of response sets*. New York, NY: Appleton-Century-Crofts.

Becker, T. E., & Colquitt, A. L. (1992). Potential versus actual faking of biodata form: An analysis along several dimensions of item type. *Personnel Psychology, 45,* 389-406. doi: 10.1111/j.1744-6570.1992.tb00855.x

Berry, C. M., & Sackett, P. R. (2009). Faking in personnel selection: Tradeoffs in performance versus fairness resulting from two cut-score strategies. *Personnel Psychology, 62,* 833–863. doi:10.1111/j.1744-6570.2009.01159.x

Chan, D. (2009). So why ask me? Are self-report data really that bad? In C. E. Lance and R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences* (pp. 309-336). New York, NY: Routledge.

Chen, P. Y., & Spector, P. E. (1991). Negative affectivity as the underlying cause of correlations between stressors and strains. *Journal of Applied Psychology, 76,* 398–407. doi: 10.1037//0021-9010.76.3.398

Converse, P. D., Pathak, J., Quist, J., Merbedone, M., Gotlib, T., & Kostic, E. (2010). Statement desirability ratings in forced-choice personality measure development: Implications for reducing score inflation and providing trait-level information. *Human Performance, 23,* 323-342. doi: 10.1080/08959285.2010.501047

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24,* 349-354. doi: 10.1037/h0047358

Dwight, S. A., & Donovan, J. J. (2003). Do warnings not to fake reduce faking? *Human Performance, 16,* 1–23. doi: 10.1207/S15327043HUP1601_1

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. New York, NY: Dryden.

Gucciardi, D. F., Jalleh, G., & Donovan, R. J. (2010). Does social desirability influence the relationship between doping attitudes and doping susceptibility in athletes? *Psychology of Sport and Exercise, 11,* 479-486. doi:10.1016/j.psychsport.2010.06.002

Hanmer, M. J., Banks, A. J., & White, I. K. (2014). Experiments to reduce the over-reporting of voting: A pipeline to the truth. *Political Analysis, 22,* 130-141. doi: 10.1093/pan/mpt027

Hermann, A., & Henneberg, M. (2013). *Anti-doping systems in sports are doomed to fail: A probability and cost analysis*. Unpublished manuscript, The University of Adelaide. Retrieved from http://www.adelaide.edu.au/news/binary9861/Doping.pdf.

Holden, R. R. (2007). Socially desirable responding does moderate personality scale validity both in experimental and in nonexperimental contexts. *Canadian Journal of Behavioral Science/Revue canadienne des sciences du comportment, 39,* 184–201. doi: 10.1037/cjbs2007015

Holden, R. R. (2010). Social desirability. In I. B. Weiner & W. E. Craighead (Eds.), *The Corsini Encyclopedia of Psychology* (4th ed., pp. 1-2). Hoboken, NJ: Wiley. doi: 10.1002/9780470479216.corpsy0889

Holden, R. R., & Passey, J. (2010). Social desirability. In M. R. Leary and R. H. Hoyle (Eds.), *Handbook of individual differences in social behavior* (pp. 441-454). New York, NY: Guilford Publications.

Holden, R. R., Magruder, C. D., Stein, S. J., Sitarenios, G., & Sheldon, S. (1999). The effects of anonymity on the Holden Psychological Screening Inventory. *Personality and Individual Differences, 27,* 737-742. doi: 10.1016/S0191-8869(98)00274-8

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75*, 581-595. doi:10.1037/0021-9010.75.5.581

Jackson, D. N., Wroblewski, V. R., & Ashton, M. C. (2000). The impact of faking on employment tests: Does forced choice offer a solution? *Human Performance, 13*, 371–388. doi: 10.1207/S15327043HUP1304_3

Johnson, J. (2004). The impact of item characteristics on item and scale validity. *Multivariate Behavioral Research, 39,* 273–302. doi:10.1207/s15327906mbr3902

Jones, E. E., & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin, 76,* 349–364. doi: 10.1037/h0031617

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

Komar, S., Brown, D. G., Komar, J. A., & Robie, C. (2008). Faking and the validity of conscientiousness: A Monte Carlo investigation. *Journal of Applied Psychology, 93,* 140–154. doi:10.1037/0021-9010.93.1.140

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity, 4,* 2025-2047. doi: 10.1007/s11135-011-9640-9

Kuncel, N. R. & Tellegen, A. (2009). A conceptual and empirical reexamination of the measurement of the social desirability of items: Implications for detecting desirable response style and scale development. *Personnel Psychology, 62,* 201–228. doi: 10.1111/j.1744-6570.2009.01136.x

Lautenschlager, G. J., & Flaherty, V. L. (1990). Computer administration of questions: More desirable or more social desirability? *Journal of Applied Psychology, 75,* 310-314. doi: 10.1037/0021-9010.75.3.310

Lazuras, L., Barkoukis, V., Rodafinos, A., & Tsorbatzoudis, H. (2010). Predictors of doping intentions in elite-level athletes: A social cognition approach. *Journal of Sport & Exercise Psychology, 32,* 694-710.

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. doi: 10.1111/j.1468-2389.2006.00339.x

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research, 35,* 382–385. doi: 10.1097/00006199-198611000-00017

McFarland, L. A, & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology, 85,* 812-821. doi: 10.1037/0021-9010.85.5.812

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. doi: 10.1037/a0019216

McCrae, R. R., & Costa, P. T., Jr. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51,* 882–888. doi: 10.1037/0022-006X.51.6.882

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology, 15,* 263-280. doi: 10.1002/ejsp.2420150303

Ntoumanis, N., Ng, J. Y. Y., Barkoukis, V., & Backhouse, S. (2014). Personal and psychosocial predictors of doping use in physical activity settings: A meta-analysis. *Sports Medicine, 44,* 1603-1624. doi: 10.1007/s40279-014-0240-4

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*(6), 660-679. doi:10.1037/0021-9010.81.6.660

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology, 78,* 679–703. doi: 10.1037/0021-9010.78.4.679

Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen, F. N. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling, 8,* 287–312. Doi: 10.1207/S15328007SEM0802_7

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). San Diego, CA: Academic Press.

Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49–69). Mahwah, NJ: Erlbaum.

Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R.W. Robins, R.C. Fraley & R. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 224–239). New York, NY: Guilford Press.

Paunonen, S. V., & LeBel, E. P. (2012). Socially desirable responding and its elusive effects on the validity of personality assessments. *Journal of Personality and Social Psychology, 103,* 158-175. doi: 10.1037/a0028165

Peidmont, 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.

Petróczi, A. (2007). Attitudes and doping: A structural equation analysis of the relationship between athletes' attitudes, sport orientation and doping behavior. *Substance Abuse Treatment, Prevention, and Policy*, *2*: 34 doi: 10.1186/1747-597X-2-34

Petróczi, A., & Nepusz, T. (2011). Methodological considerations regarding response bias effect in substance use research: Is correlation between the measured variables sufficient? *Substance Abuse Treatment, Prevention, and Policy*, *6*: 1. doi: 10.1186/1747-597X-6-1

Petróczi, A., Uvacsek, M., Nepusz, T., Deshmukh, N., Shah, I., Aimdan, E., … Naughton, D. P. (2011). Incongruence in doping related attitudes, beliefs and opinions in the context of discordant behavioral data: In which measures do we trust? *PLoS ONE 6,* e18804. doi:10.1371/journal.pone.0018804

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63,* 539-569. doi: 10.1146/annurev-psych-120710-100452

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health, 29,* 489–497. doi: 10.1002/nur.20147

Rammstedt, B., & Kemper, C. J. (2011). Measurement equivalence of the Big Five: Shedding further light on potential causes of the educational bias. *Journal of Research in Personality, 45*, 121–125. doi: 10.1016/j.jrp.2010.11.006

Randall, D. M., & Fernandes, M. F. (1991). The social desirability response bias in ethics research. *Journal of Business Ethics, 10,* 805-817. doi: 10.1007/BF00383696

Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne social desirability scale. *Journal of Clinical Psychology, 38,* 119-125. doi: 10.1002/1097-4679(198201)38:1<119::AID-JCLP2270380118>3.0.CO;2-I

Roese, N. J., & Jamieson, D. W. (1993). Twenty years of bogus pipeline research: A critical review and meta-analysis. *Psychological Bulletin, 114,* 363–375.

Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment testing and hiring decisions. *Journal of Applied Psychology, 83,* 634–644. doi:10.1037/0021-9010.83.4.634

Schmidt, F. L., Le, H., & Ilies, R. (2003). Beyond alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual-differences constructs. *Psychological Methods, 8,* 206-224. doi: 10.1037/1082-989X.8.2.206

Smith, D. B., & Ellingson, J. E. (2002). Substance versus style: A new look at social desirability in motivating contexts. *Journal of Applied Psychology, 87*, 211–219. doi: 10.1037//0021-9010.87.2.211

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

Strahan, R., & Gerbasi, K. C. (1972). Short, homogeneous versions of the Marlowe–Crowne social desirability scale. *Journal of Clinical Psychology, 28,* 191–193.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133,* 859-883. doi: 10.1037/0033-2909.133.5.859

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

van Hooft, E. A. J., & Born, M. P. (2012). Intentional respond distortion on personality tests: Using eye-tracking to understand response processes when faking. *Journal of Applied Psychology, 97,* 301-316. doi: 10.1037/a0025711

Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement, 59,* 197–210. doi: 10.1177/00131649921969802

Wetzel, E., Carstensen, C. H., & Böhnke, J. R. (2013). Consistency of extreme response style and nonextreme response style across traits. *Journal of Research in Personality, 47,* 178–189. doi: 10.1016/j.jrp.2012.10.010

Whitaker, L., Long, J., Petróczi, A., & Backhouse, S. H. (in press). Using the prototype willingness model to predict doping in sport. *Scandinavian Journal of Medicine and Science in Sport.* doi: 10.1111/sms.12148

Ziegler, M., & Buehner, M. (2009). Modeling socially desirable responding and its effects. *Educational and Psychological Measurement, 69,* 548–565. doi: 10.1177/0013164408324469

Table 1. *Overview of doping research which have focused on or included a measure of SDR.*

| Reference | Sample | Design | Measure of SDR | Analyses and Results |
|---|---|---|---|---|
| Barkoukis et al. (2011) | 1040 Greek elite, adult athletes ($M_{age}$ = 22.97; *SD* = 6.39; 62.6% males) | Cross-sectional survey | 10-item version of the Marlowe-Crowne Social Desirability Scale (Strahan & Gerbasi, 1972) | • SDR included as a covariate in ANOVA to examine differences in past doping use and intentions according to motivation and achievement goals<br>• SDR had a significant effect on past use and intentions both when groups were formed on self-determined motivation and achievement goals<br>• SDR inversely related to doping intentions ($r$ = -.16, $p$ < .001) |
| Barkoukis et al. (2014) | 309 Greek elite, adolescent athletes ($M_{age}$ = 16.64; *SD* = 1.15; 57.6% males) | Cross-sectional survey | 10-item version of the Marlowe-Crowne Social Desirability Scale (Strahan & Gerbasi, 1972) | • Variance reduction rate (VRR; see Chen & Spector, 1991) used to control for SDR in regression (VRR = [zero-order correlation]$^2$ – [partial correlation] $^2$ / [zero-order correlation]$^2$)<br>• SDR associated with susceptibility for doping use ($r$ = -.25, $p$ <.001)<br>• Correlations between study variables (attitudes, subjective norms, descriptive norms, deception) and doping susceptibility did not change significantly after controlling for SDR |
| Gucciardi et al. (2010) | 224 Australian elite athletes ($M_{age}$ = 22.68; *SD* = 6.70; 61.2% males) | Cross-sectional survey | 16-item Social Desirability Scale-17 (Stöber, 2001) | • SDR unrelated to attitudes to doping ($r$ = -.13, $p$ = .06) or doping susceptibility ($r$ = .10, $p$ = .13)<br>• Indirect effect of attitudes to doping to doping susceptibility via SDR (β = .05, 95% CI = .01 to .11, $p$ = .017) within a structural equation modelling (SEM) framework<br>• Moderation analyses indicated that the strength of the relation between attitudes to doping and doping susceptibility was influenced by SDR (intact for low, less for medium and high SDR) |
| Lazuras et al. (2010) | 750 Greek elite, adult athletes ($M_{age}$ = 25; *SD* = 5.89; 63.9% males) | Cross-sectional survey | 10-item version of the Marlowe-Crowne Social Desirability Scale (Strahan & Gerbasi, 1972) | • VRR employed to examine the influence of SDR<br>• SDR associated with susceptibility for doping intentions ($r$ = -.16, $p$ <.001) and situational temptations ($r$ = -.27, $p$ <.001)<br>• Correlations between study variables (past doping behavior, attitudes, subjective norms, descriptive norms, perceived behavioral control, situational temptation) and doping intentions did not change significantly after controlling for SDR |
| Petróczi (2007) | 199 college athletes ($M_{age}$ = 20.1; *SD* = 1.9; 71.6% males) | Cross-sectional survey | 33-item Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960) | • SDR associated with attitudes to doping ($r$ = -.22, $p$ <.001) |
| Petróczi and Nepusz (2011) | 278 male, college athletes ($M_{age}$ = 20.2; *SD* = 2.15) | Cross-sectional survey | 33-item Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960) | • SDR associated with doping attitudes ($r$ = -.22, $p$ <.001), internal deterrence factors such as moral values and health concerns ($r$ = .14, |

| | | | | |
|---|---|---|---|---|
| | | | | $p$ <.05), and opinion regarding legalizing doping for top athletes ($r$ = -.18, $p$ <.01) or all athletes ($r$ = -.14, $p$ <.05)<br>• Stronger associations between SDR and the study variables for the high SDR group, but non-significant relations for the low SDR group [sample split using cluster analysis]<br>• SEM without SDR was a poor fit with the data and contained a large degree of unexplained variance; the addition of SDR enhance variance explained and model-data fit |
| Whitaker et al. (in press) | 729 competitive athletes ($M_{age}$ = 28.8; $SD$ = 10.1; 63% males) | Cross-sectional survey | 20-item impression management subscale of the Balanced Inventory of Desirable Responding (Paulhus, 1991) | • SDR associated with willingness to dope ($r$ = -.19, $p$ <.001), PES user favorability[#] ($r$ = -.08, $p$ <.05), PES user similarity* ($r$ = -.10, $p$ <.05), nonuser similarity ($r$ = -.08, $p$ <.001), attitudes to doping ($r$ = -.15, $p$ <.001), subjective norms ($r$ = -.13, $p$ <.001), and positive outcome expectancies associated with the use of PES ($r$ = -.15, $p$ <.001) |

*Note*: SDR = socially desirable responding; PES = performance enhancing substances; [#] favorability refers to perceptions of how favorable individuals perceived PES user/nonuser to be; * similarity refers to perceptions of whether or not characteristics perceived to describe a PES user/nonuser also described them.