Received: 13 February 2023 | Revised: 28 January 2024 | Accepted: 7 February 2024

DOI: 10.1111/ijsa.12465

RESEARCH ARTICLE

International Journal of Selection and Assessment | WILEY

Faking good and bad on self-reports versus informant-reports of Dark Triad personality

Sarah A. Walker[1] | Carolyn MacCann[2]

[1]School of Education, Durham University, Durham, UK

[2]School of Psychology, University of Sydney, Camperdown, NSW, Australia

Correspondence
Sarah A. Walker, School of Education, Durham University, Durham DH1 1TA, UK.
Email: sarah.a.walker@durham.ac.uk

Abstract

Research consistently demonstrates that people can distort their responses on self-report personality tests. Informant-reports (where a knowledgeable informant rates a target's personality) can be used as an alternative to self-ratings. However, there has been little research on the extent to which informants can distort their responses on personality tests (or their motives for response distortion). The current study examines the effects of experimentally induced response distortion on self- and informant-reports of the Dark Triad. The participants ($N = 834$ undergraduates) completed Dark Triad measures in a 2×3 between-person design crossing format (self- vs. informant-report [imagined friend]) with instruction condition (answer honestly, look good, or look bad). "Look good" effects were significant for both self-reports ($d = -1.22$ to 1.42) and informant-reports ($d = -1.35$ to 0.62). "Look bad" effects were also significant for both self-reports ($d = -0.56$ to 3.58) and informant-reports ($d = -0.55$ to 3.70). The Five Factor Machiavellianism Inventory results were opposite to hypotheses, but Dirty Dozen Machiavellianism results were as expected. We conclude that people can distort Dark Triad scores for themselves (self-report) and on behalf of someone else (informant-report). We discuss the relevance of our findings for self- and informant-report assessment in applied contexts.

KEYWORDS

Dark Triad, informant-report, instructed faking, observer-report, self-report

Practitioner Points

- The study demonstrates that informants can distort their responses on Dark Triad personality domains, highlighting the need for scrutiny in informant-reports and self-reports.
- By manipulating instructions to "look good" and "look bad," the study found significant effects of faking on both self-reports and informant-reports of the Dark Triad domains.
- Results of this study emphasize the importance of considering response distortion in the assessment of the Dark Triad, especially in high-stakes real-world contexts, to ensure accurate score interpretation.

1 | INTRODUCTION

Although self-report ratings scales are the most frequent method of assessing personality, their accuracy has often been criticized (Paulhus & Vazire, 2007). One major criticism is that self-ratings allow the test-taker to engage in response distortion. That is, test-takers may attempt to make themselves look better than they really are (to "look good") or worse than they really are (to "look bad"), resulting in inaccurate measurement (Edwards, 1957; Paulhus, 2002). Meta-analyses show that test-takers can distort their responses when instructed to do so, resulting in large to very large mean score changes to both Big Five personality traits and Dark Triad personality trait constellations (Viswesvaran & Ones, 1999; Walker et al., 2022). The Big Five represents five personality domains (openness, conscientiousness, extraversion, agreeableness, and neuroticism; Viswesvaran & Ones, 1999) and the Dark Triad represents three personality trait constellations (narcissism, Machiavellianism, and psychopathy; Paulhus & Williams, 2002). Dark Triad assessment has gained popularity following the highly publicized scrutiny surrounding organizational misconduct (Van Scotter & Roglio, 2020), police misconduct (Semrad & Scott-Parker, 2020), the abuse and murder of prisoners by armed forces and CIA operatives (Bartone, 2010) right through to the Dark Triad predicting workplace outcomes including job performance, citizenship behavior, counterproductive workplace behavior, and leadership styles (Spain et al., 2014). As assessment of the Dark Triad and use of informant-report scales become more prevalent in organizations interested in screening out individuals possessing maladaptive characteristics during the selection process (Spain et al., 2014), confidence in these measures to provide accurate and reliable assessment is paramount. This research provides a starting point for researchers to consider the practical implications of informant-report response distortion, and mitigate the potential for faking in informant responses when designing studies and interpreting results.

One suggested solution to response distortion on self-report rating scales is to use informant-reports (Kim et al., 2019; Vazire, 2006). Prior meta-analytic findings show satisfactory agreement between self- and informant-reported personality scores, emphasizing the benefits of using both self- and informant-reports to measure personality (Connelly & Ones, 2010; Kim et al., 2019). However, it is possible that informants might also distort their responses. Although prior work has laid important groundwork for identifying response biases on self-report scales, there exists a considerable gap in understanding how Dark Triad dimensions are captured using informant-report scales. As yet, there has been little to no empirical examination of informant response distortion. That is, no study has investigated the extent to which informants try to make someone else "look good" or "look bad" on subclinical Dark Triad rating scales, and only one prior study has measured informant distortion to make someone else "look good" on the Big Five (König et al., 2017). This study provides the first examination of the extent to which people can distort their responses on informant-report ratings of the Dark Triad personality dimensions providing a foundation to explore response distortion on informant-report ratings scales.

1.1 | Instructed faking

Despite the validity of self-reported information (Holden & Passey, 2010), there has been substantial debate about the accuracy of self-report scales, which represents an ongoing challenge for researchers (Hogan & Foster, 2016). These concerns typically relate to the impact of response sets, biases, and styles on the validity of self-reported noncognitive data. Not all test-takers respond accurately. Some may distort responses unconsciously, whereas others may intentionally distort their responses when motivated to do so (Edwards, 1957; Paulhus & Vazire, 2007). Ego-protective or self-serving biases tend to arise when an individual's sense of self is threatened. An individual may endorse items to enhance or protect their self-perceptions (Rodman et al., 2017).

Instructed faking paradigms are a standard experimental method used to examine the extent to which individuals can distort their responses in a socially desirable way on personality measures. Individuals may intentionally present themselves more favorably than they really are (distorting to look good) or less favorably than they really are (distorting to look bad, also referred to as malingering; Arthur et al., 2010; Donaldson & Grant-Vallone, 2002; Furnham, 1990; Rogers & Bender, 2018; Rogers et al., 2003). For example, a job applicant may try to enhance their positive qualities to look like a good fit for the job (Birkeland et al., 2006). In contrast, test-takers may try to make themselves look worse than they really are in some contexts. For example, in a forensic setting, an offender may inflate their negative qualities when being assessed for competency to stand trial (Mills & Kroner, 2005).

Although there is substantial literature on response bias and psychopathy more generally (see Ray et al., 2013 for a review), the vast majority of this research has focused on identifying such biases using validity scales (sometimes called "lie scales" or "response distortion scales"), often in a clinical context (Paulhus, 2002; Ray et al., 2013; Sleep et al., 2017). For example, measures such as the clinical Minnesota Multiphasic Personality Inventory have a "built-in" validity scale to detect socially desirable responding. Examining response distortion through validity scales is quite different to experimental paradigms manipulating the stakes of the assessment to study such distortion (i.e., instructed faking paradigms). Although validity scales are commonly researched, instructed faking research is much less common. In fact, a recent meta-analysis located only two, four, and 13 instructed faking studies for narcissism, Machiavellianism, and psychopathy respectively (Walker et al., 2022).

Given the limited research on faking in the context of the Dark Triad, research examining the Honesty-Humility factor of the HEXACO model (Lee & Ashton, 2014) may offer additional insight. Honesty-Humility evaluates characteristics such as fairness, sincerity, and modesty, has been shown to be negatively related to the Dark Triad traits (Lee & Ashton, 2014). Instructed Faking research on the HEXACO suggests individuals may modify their responses under faking conditions (MacCann, 2013). Similarly, Dunlop et al. (2015) showed a positive association between faking and the modesty facet of Honesty-Humility, suggesting that more modest individuals may

be more likely to feel pressure to present themselves in a positive light. Given the demonstrated relationship between the Dark Triad and the HEXACO Honesty-Humility dimension, it is reasonable to expect similar response distortion may occur further indicating that people can fake on measures of the Dark Triad. The accuracy of self-report rating scales relies on the test-taker's truthful responding and the assumption of accurate self-perception (Klonsky, Jane, et al., 2002; Paulhus & Vazire, 2007).

An alternate method of obtaining information about a person is through informant-reports. Informant-reports are assessments in which another person (e.g., work colleague, friend, or parent) rates a target (Vazire & Carlson, 2010). How well an informant knows the target and how much the informant likes the target can impact the informant's reporting (Hollander, 1956; Leising et al., 2010). Indeed, early users of informant-reports raised concerns about potential "friendship effects," suggesting informants with a close relationship to the target may respond favorably about them (Hollander, 1956). Nevertheless, the nature of inaccuracy may depend on the source of the report. When self-report ratings are more favorable than informant-reports, these discrepancies may reflect self-deceptive enhancement, in which the individual genuinely has an inflated self-perception. Conversely, when informant-reports are more favorable, the inaccuracy may by influenced by "friendship effects" such that an informant's relationship with the target positively biases their evaluation.

Such "friendship effect" concerns are often dismissed with the suggestion that informants are not motivated to distort their responses as the consequences of their responses do not directly impact them (Leising et al., 2010; Vazire, 2006). However, the degree to which informants like the target inevitably impacted their ratings of the target (Hollander, 1956; Klonsky, Oltmanns, et al., 2002; Leising et al., 2010). Recent findings support this view. For example, Beckmann et al. (2020) found that informant-ratings are more favorable than self-ratings for all Big Five traits (i.e., higher for extraversion, agreeableness, conscientiousness, and openness but lower for neuroticism). Effects were larger for nonwork informants (romantic partners, friends, and/or family members) than for work informants (direct reports, supervisors and/or work peers), consistent with a "friendship effect." Similarly, Grös et al. (2007) found informants tended to rate their friends more favorably compared to self-reports, especially on evaluative traits.

In contrast, self-ratings were found to be more favorable than informant ratings across all Big Five personality traits (Mount et al., 1994) and emotional intelligence (Walker & MacCann, 2023). Informant ratings were more favorable for well-known, rather than lesser-known targets, consistent with a "friendship effect." Similarly, Clifton et al. (2005), Miller et al. (2005), and Sleep et al. (2017) found that self-reported ratings of narcissism were more favorable (i.e., lower narcissism) than informant ratings. Psychopathy was also rated more favorably (i.e., lower psychopathy) when self-reported compared to when rated by an informant (Miller et al., 2011). Taken together, these findings indicate that self-ratings are likely to be more favorable than informant-ratings on all Dark Triad trait constellations. That is, informant-ratings will be generally higher for narcissism, Machiavellianism, and psychopathy domains and facets.

1.2 | The Dark Triad

The Dark Triad comprises narcissism, Machiavellianism, and psychopathy (Paulhus & Williams, 2002). Although conceptually distinct from each other, there are empirically overlapping characteristics such as ethical, moral, and socially deviant behavior as well as an interpersonally exploitative demeanor with goal-focused manipulation of others' emotions for personal gain. Given the Dark Triad predicts a range of problematic behaviors in the workplace, for example (for a review, see Spain et al., 2014), it is unsurprising interest in the Dark Triad assessment has gained popularity for its use of screening out people possessing maladaptive traits (Spain et al., 2014).

1.2.1 | Narcissism

As a subclinical personality domain, narcissism covers a spectrum from mild to extreme (Miller & Campbell, 2008) and comprises two facets: grandiose and vulnerable narcissism. Grandiose narcissism is characterized by grandiosity, self-confidence, and exploitation of others with a tendency to rely on self-internal validation (Dickinson & Pincus, 2003; Zhang et al., 2017). When threatened, people with high grandiose narcissism may blame and devalue others while refusing to acknowledge their own weaknesses. Vulnerable narcissism is characterized by grandiose fantasies, oscillations between self-love and self-loathing, a fragile sense of self, a reliance on external validation for self-esteem maintenance, and hypersensitivity to negative feedback (Wink, 1991). When threatened, people with high vulnerable narcissism become defensive and resentful, and may show aggressive outbursts toward others (Wink, 1991).

1.2.2 | Machiavellianism

Subclinical Machiavellianism is derived from the philosophical writings of Niccolo Machiavelli, a political advisor to the Medici family in 16th century Firenze (Christie & Geis, 1970). Machiavellianism is characterized by goal-focused manipulative and callous social interactions, including the use of long-term strategic planning to delay gratification for better rewards in the future, questionable morals, and a cold, cynical world-view (Christie & Geis, 1970; Furnham et al., 2013).

1.2.3 | Psychopathy

Subclinical psychopathy is characterized by superficial charm, pathological lying, and lack of empathy, conscience, and remorse (Cleckley, 1951; Hare, 2003), existing on a continuum in the wider population (Berg et al., 2013). Psychopathy measures, such as the Levenson's Self-report Psychopathy Scale (Levenson et al., 1995), were developed to measure the two-factor structure of psychopathy proposed by Karpman (1941). This two-factor structure comprises two related but distinct factors differing in their etiology and symptomology: primary and secondary

psychopathy (Hare, 2003). Both facets are typified by indifference to one's own and others' emotions, underpinned by an antagonistic interpersonal style (Miller & Lynam, 2012). More recently, a three-factor approach to psychopathy (as measured by Levenson's Self-report Psychopathy Scale) has been shown to better capture the egocentric, callous, and antisocial characteristics of psychopathy better than the original two-factor approach (Brinkley et al., 2008; Garofalo et al., 2019; Sellbom, 2011). Egocentricity relates to interpersonally manipulative and antagonistic characteristics associated with perceived low social responsibility (Christian & Sellbom, 2016; Sellbom, 2011), callousness relates to lacking empathy and remorse and is associated with cold-heartedness, lack of remorse, and low empathy (Anderson et al., 2013; Sellbom, 2011), whereas antisocial is related to impulsivity, and antisocial behavior (Brinkley et al., 2008; Sellbom, 2011).

1.3 | Instructed faking on self- and informant-reports

The current study aims to examine how much people distort their responses on self and informant-ratings of the Dark Triad under simulated "high stakes" conditions of obtaining or avoiding employment. We use a 2 × 3 between-person design crossing the target (oneself vs. an imagined friend) with the motivational condition, or stakes of the assessment (no stakes vs. wanting a desirable job vs. wanting to avoid an undesirable job). Using the job simulation condition ensures there is an element of choice over the extent to which people will distort their responses under simulated high-stakes conditions (as opposed to asking the participants explicitly to "fake").

Our expectations for the extent to which people can distort their responses on self-report scales are based on prior meta-analytic work examining faking on personality questionnaires (see Viswesvaran & Ones [1999] for a review). Large effects of faking-good have been demonstrated on all Big Five personality domains such that when instructed to, people can increase their positive attributes and decrease negative attributes on self-report scales.

Our expectations regarding the effects of response distortion on informant-report scales are largely based on prior findings demonstrating that informants inflate their target's positive characteristics (Leising et al., 2010). Similarly, assessing psychopathology reported by self- and informant-reports indicated that informants were also prone to overreporting symptomology as measured by validity scales (Quilty et al., 2018; Webber et al., 2022). As such, we expect informants can distort their informant-reported responses on Dark Triad measures when instructed to do so.

1.4 | Hypotheses

All hypotheses were pre-registered at http://aspredicted.org/blind.php?x=ke3wc3

Hypothesis 1. Individuals will fake on all measures of the Dark Triad. There will be a significant mean difference on all Dark Triad scores for "honest" as compared to "look good" and "look bad" conditions. We expect all Dark Triad scores to be higher in the "look bad" condition (Hypothesis 1a) and lower in the "look good" condition (Hypothesis 1b) as compared to the "answer honestly" condition.

Hypothesis 2. Self-reports will be more favorable than informant-reports. Specifically, all Dark Triad scores will be significantly *lower* for self-reports than informant-reports.

Hypothesis 3. There will be an interaction between instruction condition and rater (i.e., self or informant),[1] such that people will fake more for others than for themselves. That is, differences between the "look good" and "answer honestly" conditions will be larger for informant- than self-reports (indicating that people asked to "look good" for their friend are likely to distort their responses to a larger degree, as compared to people asked to "look good" for themselves; Hypothesis 3a) and the difference between "look bad" and "answer honestly" will similarly be larger for informant- than self-reports (Hypothesis 3b).

2 | METHOD

2.1 | Participants

Undergraduate psychology students (convenience sample) at the first author's university (N = 834; 557 female, 256 male, 1 nonbinary) took part in the 10-min study during an in-class activity and as such, did not receive any incentives for taking part. Participants were aged between 17 and 55 years (M = 27.46; SD = 3.07). Additional participant information (including exclusion based on preregistered criteria) can be found in Supporting Information materials at https://osf.io/eb49g/?view_only=28d4eceb865049558c8b1818596cb684. Study data are available in Supporting Information materials.

2.2 | Sample size

This study used a convenience sample of all students enrolled in a particular course. As such, the sample size was driven by practicalities rather than power analysis to determined minimum sample size. However, a power analysis based on a target effect size of d = 0.40 (f = 0.20) for a 3 × 2 analysis of variance (ANOVA) showed that a sample of 390 would be required to achieve 95% power (N = 244 for 80% power). This target effect size was based on the smallest effect in a meta-analysis of instructed faking on the Dark Triad (Walker et al., 2022). We therefore have adequate power to test hypotheses.

2.3 | Materials

Pathological Narcissism Inventory-Brief (Schoenleber et al., 2015) is a 28-item scale measuring grandiose narcissism (e.g., "I can usually talk

my way out of anything"), vulnerable narcissism (e.g., "I often hide my needs for fear that others will see me as needy and desperate"), and total narcissism. Items are rated on a 6-point scale ranging from "not at all like me" to "very much like me".

Levenson's Self-Report Psychopathy Scale (LRSP; Levenson et al., 1995) is a 19-item scale assessing egocentricity (e.g., "Success is based on survival of the fittest; I am not concerned about the losers"), callousness (e.g., "I make a point of trying not to hurt others in pursuit of my goals", and antisocial (e.g., "I find myself in the same kinds of trouble time after time") psychopathy dimensions (Christian & Sellbom, 2016). Items are rated on a 4-point scale ranging from "disagree strongly" to "agree strongly".

The Five Factor Machiavellianism Inventory (FFMI; Collison et al., 2018) assesses Machiavellianism with 52 items based on an expert-derived trait profile of the five factor model of personality (e.g., "I will go out of my way to help other people", "I like to carefully consider the consequences before I make a decision", "I work hard to pursue my goals"). It should be noted that although the FFMI was developed to assess normal variations in personality traits, its design may not effectively capture the upper extremes of Machiavellianism. Items are rated on a 5-point scale from "disagree strongly" to "agree strongly".

The Dirty Dozen (Jonason & Webster, 2010) is a 12-item scale measuring the Dark Triad personality dimensions on a 5-point Likert scale (not well to extremely well). There are four narcissism items (e.g., "I tend to seek prestige or status"), four Machiavellianism items (e.g., "I tend to manipulate others to get my way"), and four psychopathy items (e.g., "I tend to be callous or insensitive").

2.4 | Procedure

All participants completed a demographic questionnaire. Participants were then randomly assigned to one of six conditions using Qualtrics randomization: self/fake-bad, self/fake-good, self/honest, informant/fake-bad, informant/fake-good, informant/honest. In the "informant-report" conditions, participants were asked to think of a peer of the same sex and age (not a romantic partner). There were several reasons for this instruction including avoiding potential emotional biases romantic partners may feel toward their significant other. By asking participants to think of a peer of the same sex and age (but not a romantic partner), the study aims to control those extraneous factors. Additionally, these instructions promote comparability between groups, simplify interpretation by limiting the types of relationships participants could consider, and uniformity of conditions ensuring observed effects are more likely attributable to the manipulation of the study rather than the type of relationship between target and informant. The specific instructions for each condition are given in the Supporting Information materials (https://osf.io/eb49g/?view_only=28d4eceb865049558c8b1818596cb684). The "look good" instructions framed the personality ratings as part of a job selection process for a job you/your friend really wanted whereas the "look bad" a job that you/your friend really did not want to take.

Following the instruction screen, participants were asked "what did the instructions ask you to do?" and participants had to select a response from nine options (e.g., "Rate myself honestly", "Complete some intelligence tests"). If they did not answer correctly, the instructions were displayed a second time. There were four manipulation checks, all of which were data check items, in place to ensure attention to the task and instructions (see Supporting Information materials). All protocols were approved by the Human Research Ethics Committee of the first author's institution.

2.5 | Analysis

Hypotheses 1 to 3 were tested using 3 × 2 ANOVAs.[2] We used contrast coding as follows, for the six conditions informant/fake-bad, informant/fake-good, informant/honest, self/fake-bad, self/fake-good, self/honest: (a) self-report versus informant-report (0.33, 0.33, 0.33, −0.33, −0.33 −0.33); (b) fake-bad versus honest (0.5, 0 −0.5, .5, −0.5); (c) fake-good versus honest (0, 0.5, −0.5, 0.5, −0.5); (d) the interaction of "look bad" with measurement type (i.e., contrast a * contrast b); and (e) the interaction of faking good with measurement type (i.e., contrast a * contrast c). A separate ANOVA was conducted for each of the Dark Triad domains and facets. We evaluated the effect size with respect to η_P^2, with values of 0.01, 0.06, and 0.14 considered as "small," "medium," and "large," respectively (Cohen, 1988). Posthoc analyses were conducted to test whether the association between the Dark Triad trait constellations and self-interested, or other-interested variables differed between honest and faking conditions. This analysis represents a departure from the preregistration.

3 | RESULTS

3.1 | Reliability and descriptive statistics

Table 1 presents the reliability, descriptive statistics for the Dark Triad domain and facet scores. Mean differences across conditions are also shown. Internal consistency of the Dark Triad was good across honest, look good, and look bad conditions, except for LSRP total psychopathy in three conditions: look-bad/self-report (α = .57), honest/informant-report (α = .67), and look-bad/informant report (α = .64). These reliability estimates are consistent with prior research on the LSRP (Schoenleber et al., 2015). Effect sizes (Cohen's *d*, see Table 1) ranged from small to very large for "look good" on self-ratings, very small to extremely large for "look bad" on self-ratings, small to very large for "look good" on informant-ratings, and large to extremely large for "look bad" on informant-ratings. The correlations among condition variables—honest responding, fake good, and fake bad—are detailed in Tables 2, 3, and 4, respectively. For all tables, self-report data are positioned below the diagonal, while informant-report data are located above it.

TABLE 1 Reliability and descriptive statistics for each Dark Triad domain and facet.

	Honest (N = 281) (self n = 143, informant n = 138)			Look good (N = 281) (self n = 147, informant n = 134)					Look bad (N = 272) (self n = 132, informant n = 140)				
Trait constellations	M	SD	α	M	SD	α	t	d	M	SD	α	t	d
Self-report													
PNI Total narcissism	2.98	0.61	.89	2.46	0.59	.89	−7.35***	−0.86	3.64	0.60	.80	9.24***	1.11
PNI Grandiose narcissism	3.26	0.62	.77	3.20	0.72	.81	−0.84	−0.10	3.15	0.92	.80	−1.16	−0.14
PNI Vulnerable narcissism	2.76	0.72	.86	1.91	0.68	.90	−10.37***	−1.22	4.02	0.66	.78	15.09***	1.82
DD Narcissism	2.65	0.89	.71	2.19	0.84	.77	−4.46***	−0.52	4.03	1.19	.79	10.79***	1.32
FFMI Machiavellianism	2.96	0.39	.85	3.44	0.25	.92	12.00***	1.42	2.37	0.45	.90	−4.64***	−0.56
DD Machiavellianism	2.21	0.85	.65	1.71	0.76	.78	−5.26***	−0.62	4.18	1.17	.86	15.79***	1.93
LRSP Egocentricity	1.84	0.64	.89	1.67	0.70	.92	−2.25	−0.27	3.59	0.58	.87	23.86***	2.88
LRSP Callous	1.66	0.56	.55	1.47	0.66	.83	−2.56	−0.30	3.58	0.70	.76	24.94***	3.05
LRSP Antisocial	2.32	0.54	.52	1.59	0.72	.89	−9.84***	−1.16	3.64	0.58	.75	19.33***	2.35
DD Psychopathy	1.87	0.71	.82	1.33	0.67	.75	−7.21***	−0.85	4.50	0.80	.72	28.85***	3.50
Informant-report													
PNI Total narcissism	2.84	0.49	.83	2.38	0.53	.87	−7.55***	−0.92	3.68	0.54	.82	13.67***	1.64
PNI Grandiose narcissism	3.03	0.59	.75	2.93	0.70	.81	−1.42	−0.17	3.37	0.76	.77	3.99***	0.48
PNI Vulnerable narcissism	2.70	0.60	.80	1.96	0.61	.88	−10.05***	−1.22	3.94	0.64	.82	16.74***	2.01
DD Narcissism	2.30	0.94	.81	1.88	0.80	.82	−3.97***	−0.48	3.83	1.07	.75	17.00***	1.51
FFMI Machiavellianism	3.10	0.41	.85	3.31	0.24	.89	5.17***	0.62	2.90	0.32	.93	−4.58***	−0.55
DD Machiavellianism	1.99	0.85	.70	1.47	0.56	.69	−5.91***	−0.71	3.95	1.06	.79	12.59***	2.04
LRSP Egocentricity	1.94	0.61	.87	1.43	0.47	.86	−7.78***	−0.94	3.46	0.60	.88	20.97***	2.52
LRSP Callous	1.83	0.62	.72	1.36	0.47	.65	−7.09***	−0.86	3.40	0.70	.74	19.58***	2.36
LRSP Antisocial	2.08	0.56	.59	1.44	0.51	.74	−9.99***	−1.35	3.60	0.56	.82	22.38***	2.69
DD Psychopathy	1.73	0.71	.81	1.22	0.40	.52	−7.36***	−1.21	4.38	0.37	.77	30.85***	3.70

Note: Reliability and descriptive statistics for each Dark Triad domain and facet under answer honestly, look good, and look bad instruction conditions for self-report and informant-report (Cohen's d compares faking conditions to answer honestly). Cohen's d compares the standardized mean difference for the honest condition to "look good" and "look bad" conditions. The t statistic was derived from an independent samples t test. To account for multiple comparisons, α was set at .001 (α/40 comparisons).

Abbreviations: DD, Dirty Dozen (5-point scale); FFMI, Five Factor Machiavellianism Inventory (5-point scale); LRSP, Levenson's Self-Report Psychopathy Scale (4-point scale); PNI, Pathological Narcissism Inventory (6-point scale).

***$p < .001$.

3.2 | Hypothesis testing

Table 5 reports the ANOVAs (testing Hypotheses 1–3).

3.2.1 | Hypothesis 1a: Dark Triad scores will be lower for "look good" instructions

In comparison to the "answer honestly" condition, participants in the "look good" condition showed significantly lower scores on eight of the 10 Dark Triad variables we measured. The two exceptions were PNI grandiose narcissism (where there were no significant differences) and FFMI Machiavellianism (where "faking good" instructions produced significantly higher scores, in contrast to hypotheses—note that Dirty Dozen Machiavellianism showed the opposite effect, as hypothesized). The effect sizes were: (a) very large (partial eta squared [η_P^2] > 0.20) for PNI vulnerable narcissism, Dirty Dozen total narcissism, and LSRP secondary psychopathy; (b) large (η_P^2 = 0.14) for PNI total narcissism, FFMI total Machiavellianism (where the effect was in the opposite direction to hypotheses), and LSRP total psychopathy; (c) moderate to large (η_P^2 = 0.10) for Dirty Dozen psychopathy; and (d) moderate (η_P^2 = 0.06) for Dirty Dozen

TABLE 2 Correlations between honest conditions self-report variables (lower matrix) and informant-report variables (upper matrix).

	PNI N	PNI G	PNI V	FFMI	LSRP E	LSRP C	LSRP A	DD N	DD M	DD P
PNI P		.761***	.876***	.217*	.519***	.242**	.468***	.520***	.514***	.354***
PNI G	.836***		.353***	.399***	.447***	.190*	.226**	.549***	.542***	.232**
PNI V	.937***	.592***		.015	.416***	.208*	.506***	.342***	.338***	.338***
FFMI M	.105	.332***	-.058		.467***	.216*	−.095	.313***	.253**	.253**
LSRP E	.411***	.375***	.365***	.376***		.602***	.334***	.328***	.531***	.567***
LSRP C	.055	.085	.027	.309***	.544***		.295***	.245**	.497***	.556***
LSRP A	.469***	.373***	.452***	−.075	.356***	.172*		.261**	.352***	.485***
DD N	.495***	.483***	.419***	.256**	.379***	.177*	.321***		.519***	.272**
DD M	.291***	.445***	.144	.541***	.403***	.312***	.217**	.405***		.504***
DD P	.172*	.125	.172*	.204*	.432***	.371***	.260**	.355***	.404***	

Abbreviations: A, antisocial; C, callous; DD, Dirty Dozen; E, egocentricity; FFMI, Five Factor Machiavellianism Inventory; LSRP, Levenson's Self-report Psychopathy scale; M, Machiavellianism; N, narcissism; P, sychopathy; PNI, Pathological Narcissism Inventory; V, vulnerable narcissism.
*$p < .05$; **$p < .01$; ***$p < .001$.

TABLE 3 Correlations between fake bad conditions self-report variables (lower matrix) and informant-report variables (upper matrix).

	PNI N	PNI G	PNI V	FFMI	LSRP E	LSRP C	LSRP A	DD N	DD M	DD P
PNI N		.757***	.811***	.187*	.555***	.378***	.396***	.458***	.375***	.375***
PNI G	.787***		.232**	.555***	.399***	.257**	.08	.493***	.371***	.111
PNI V	.767***	.208*		−.218**	.470***	.332***	.518***	.241**	.226**	.459***
FFMI M	.133	.414***	−.220*		.219**	.13	−.117	.304***	.176*	-.069
LSRP E	.345***	.272**	.265**	.293***		.704***	.528***	.417***	.507***	.549***
LSRP C	.082	.072	.055	.135	.617***		.498***	.292***	.441***	.536***
LSRP A	.054	−.16	.251**	−.263**	.408***	.502***		.240**	.268**	.620***
DD N	.407***	.461***	.166	.387***	.388***	.215*	.067		.508***	.377***
DD M	.283**	.307***	.129	.376***	.465***	.364***	.141	.610***		.507***
DD P	.216*	.173*	.163	.176*	.410***	.436***	.314***	.525***	.580***	

Abbreviations: A, antisocial; C, callous; DD, Dirty Dozen; E, egocentricity; FFMI, Five Factor Machiavellianism Inventory; LSRP, Levenson's Self-report Psychopathy scale; M, Machiavellianism; N, narcissism; P, sychopathy; PNI, Pathological Narcissism Inventory; V, vulnerable narcissism.
*$p < .05$; **$p < .01$; ***$p < .001$.

Machiavellianism and LSRP primary psychopathy. We therefore have mixed support for Hypothesis 1a.

3.2.2 | Hypothesis 1b: Dark Triad scores will be higher for "look bad" instructions

In comparison to the "answer honestly" condition, participants in the "look bad" condition showed significantly higher scores on eight of the 10 Dark Triad variables we measured. As for faking good, the two exceptions were PNI grandiose narcissism (where there were no significant differences), and FFMI Machiavellianism (where "look bad" instructions produced significantly lower scores in contradistinction to hypotheses—note that Dirty Dozen Machiavellianism showed the opposite effect, as hypothesized). The effect sizes were: (a) very large ($\eta_P^2 > 0.20$) for LRSP total, primary, and secondary psychopathy, Dirty Dozen psychopathy, Dirty Dozen Machiavellianism, PNI total narcissism, and PNI vulnerable narcissism; and (b) moderate ($\eta_P^2 = 0.06$) for FFMI Machiavellianism (where the effect was in the opposite direction to hypotheses), and c) small ($\eta_P^2 = 0.03$) for Dirty Dozen total narcissism. We therefore have mixed support for Hypotheses 1b.

3.2.3 | Hypothesis 2: Individuals will give better scores to themselves than others

Contrast 3 (Table 5) tests whether mean differences between self- and informant-ratings differ significantly from zero. For narcissism, there were no significant differences between self- and informant-

TABLE 4 Correlations between fake good conditions self-report variables (lower matrix) and informant-report variables (upper matrix).

	PNI N	PNI G	PNI V	FFMI	LSRP E	LSRP C	LSRP A	DD N	DD M	DD P
PNI N		.784***	.844***	.157	.575***	.159	.320***	.548***	.415***	.293***
PNI G	.805***		.328***	.368***	.382***	.014	.032	.480***	.400***	.201*
PNI V	.886***	.438***		−.079	.543***	.229**	.459***	.418***	.285***	.272**
FFMI M	.202*	.384***	.005		.170*	.089	−.243**	.297***	.306***	.118
LSRP E	.456***	.326***	.437***	.246**		.458***	.440***	.449***	.480***	.401***
LSRP C	.298***	.223**	.276***	.184*	.714***		.233**	.182*	.155	.219*
LSRP A	.347***	.096	.451***	−.192*	.673***	.539***		.165	.188*	.305***
DD N	.522***	.476***	.419***	.353***	.391***	.244**	.225**		.573***	.413***
DD M	.501***	.473***	.389***	.190*	.429***	.372***	.256**	.590***		.503***
DD P	.321***	.098	.409***	−.009	.360***	.308***	.321***	.431***	.468***	

Abbreviations: A, antisocial; C, callous; DD, Dirty Dozen; E, egocentricity; FFMI, Five Factor Machiavellianism Inventory; LSRP, Levenson's Self-report Psychopathy scale; M, Machiavellianism; N, narcissism; P, sychopathy; PNI, Pathological Narcissism Inventory; V, vulnerable narcissism.
*$p < .05$; **$p < .01$; ***$p < .001$.

ratings. For FFMI Machiavellianism, the effect was in the hypothesized direction such that self-ratings were significantly lower than informant-ratings (with a small effect size). For Dirty Dozen Machiavellianism, Dirty Dozen psychopathy, and LSRP psychopathy, the total score and both facets the effects were in the opposite direction to hypotheses such that self-ratings were significantly higher than informant-ratings (with a small effect size). In summary, there was very little support for Hypotheses 2.

3.2.4 | Hypothesis 3: People will fake more for others than for themselves

Differences in distorting to "look good" for informant versus self-ratings

The interaction effects for "look good" instructions with rater-type were significant for FFMI Machiavellianism, and LSRP total psychopathy, and LSRP primary psychopathy, (but not for PNI total narcissism, PNI vulnerable narcissism, PNI grandiose narcissism nor any of the three Dirty Dozen scales). For FFMI Machiavellianism, instructions to "look good" (compared to honest instructions) resulted in a very large score increase to self-ratings and a moderate to large score increase for informant ratings. There was a small, but significant difference between self- and informant-ratings such that there is greater response distortion on self-ratings than informant-ratings (i.e., both the direction of response distortion and the difference between self-ratings and informant-ratings were in opposite directions to hypotheses). For LSRP total psychopathy, there was a moderate difference between "look good" and "answer honestly" for self-ratings, and a very large difference for informant-ratings. This difference between self- and informant-ratings was small but significant, indicating greater response distortion for informant-ratings than self-ratings. For LSRP primary psychopathy there was a small difference between "look good" and "answer honestly" for self-ratings, and a large difference for informant-ratings. This difference between self- and informant-ratings was small, but significant, indicating greater response distortion for informant-ratings than self-ratings. Hypothesis 3a was thus supported only for LSRP total and primary psychopathy, and not for the other Dark Triad constructs. That is, people engage in greater faking good of primary and total psychopathy for others as compared to themselves.

Differences in distorting to "look bad" for informant versus self-ratings

The interaction effects of "look bad" instructions with self- versus informant-ratings were significant only for PNI grandiose narcissism and LSRP primary psychopathy. For grandiose narcissism, the difference between "look bad" and "answer honestly" was very small for self-reports but small to moderate for informant-reports. The difference between self- and informant-ratings was significant with a small effect size, indicating that people were distorting responses more for others than for themselves. For primary psychopathy, there was an extremely large difference between "look bad" and "answer honestly" for both self- and informant-ratings. The difference between self- and informant-ratings was significant with a small effect size, indicating that people were distorting responses more for themselves than for others. Although results were mixed, overall there is no support for the idea that people "distort to look bad" more for others than for themselves.

4 | DISCUSSION

The aim of the current study was to examine the extent to which people can distort their responses on self and informant-ratings of the Dark Triad under simulated "high stakes" conditions of obtaining or avoiding employment. The purpose of employing the job simulation condition was to ensure an element of choice over the

TABLE 5 Results of the 3 × 2 between-subjects ANOVA testing instruction condition by target (self-report vs. informant-report) and interactions.

Narcissism	PNI total narcissism			PNI grandiose narcissism			PNI vulnerable narcissism			DD narcissism		
	Contrast	*F*	η_P^2	Contrast	*F*	η_P^2	Contrast	*F*	η_P^2	Contrast	*F*	η_P^2
ψ1: honest vs. look good	−13.81	107.93***	0.12	−1.07	2.12	0.00	−12.74	208.02***	0.20	−1.75	28.91***	0.03
ψ2: honest vs. look bad	21.30	252.90***	0.23	1.30	3.11	0.00	20.00	504.60***	0.38	5.80	312.68***	0.27
ψ3: self vs. informant	−1.63	2.24	0.00	−1.14	3.62	0.00	−0.49	0.45	0.00	−1.16	18.79***	0.02
ψ4: interaction 1 (ψ1 × ψ3)	0.23	0.27	0.00	−0.09	0.15	0.00	0.32	1.20	0.00	0.02	0.04	0.00
ψ5: interaction 2 (ψ2 × ψ3)	0.83	3.49	0.00	0.87	12.64***	0.02	−0.04	0.02	0.00	0.09	0.74	0.00

Machiavellianism	FFMI Total Machiavellianism			DD Machiavellianism		
	Contrast	*F*	η_P^2	Contrast	*F*	η_P^2
ψ1: honest vs. look good	17.48	129.58***	0.14	−2.03	44.93***	0.05
ψ2: honest vs. look bad	−11.35	53.79***	0.06	7.84	663.00***	0.45
ψ3: self vs. informant	3.09	6.00**	0.01	−0.921	13.79***	0.02
ψ4: interaction 1 (ψ1 × ψ3)	−2.22	18.82***	0.02	−0.01	0.01	0.00
ψ5: interaction 2 (ψ2 × ψ3)	0.31	0.37	0.00	0.00	0.00	0.00

Psychopathy	LSRP egocentricity			LSRP callous			LSRP antisocial			DD psychopathy		
	Contrast	*F*	η_P^2	Contrast	*F*	η_P^2	Contrast	*F*	η_P^2	Contrast	*F*	η_P^2
ψ1: honest vs. look good	−3.44	45.42***	.05	−1.31	38.66***	0.05	−3.50	196.72***	0.19	−2.11	88.74***	0.10
ψ2: honest vs. look bad	16.40	1008.14***	.55	6.94	1066.71***	0.56	7.04	798.32***	0.50	10.57	2189.54***	0.73
ψ3: self vs. informant	−0.93	4.89*	.01	−0.15	0.70	0.00	−0.71	12.05***	0.01	−0.50	7.35**	0.01
ψ4: interaction 1 (ψ1 × ψ3)	−0.56	10.76***	.01	−0.19	7.40**	0.01	0.07	0.80	0.00	0.02	0.10	0.00
ψ5: interaction 2 (ψ2 × ψ3)	−0.39	5.20*	.01	−0.24	11.49***	0.01	0.16	3.91*	0.01	0.01	0.01	0.00

Note: The planned contrasts for conditions informant/fake-bad, informant/fake-good, informant/honest, self/fake-bad, self/fake-good, self/honest were coded (0, 0.5, −0.5, 0, 0.5, −0.5), (0.5, 0, −0.5, 0.5, 0, −0.5), (0.333, 0.333, 0.333, −0.333, −0.333, −0.333), (0, 0.1665, −0.1665, 0, −0.1665, 0.1665), and (0.1665, 0, −0.1665, −0.1665, 0, 0.1665) to test contrasts 1–5.

Abbreviations: η_P^2, partial eta squared; Ψ, ANOVA contrast; ANOVA, analysis of variance; DD, Dirty Dozen; FFMI, Five Factor Machiavellianism Inventory; LSRP, Levenson's Self-report Psychopathy scale; PNI, Pathological Narcissism Inventory.

*p < .05; **p < .01; ***p < .001.

extent to which people would fake under simulated high-stakes conditions (as opposed to asking the participants explicitly to "fake"). The results of this study demonstrated three things. First, people can distort their responses on self-report measures of the Dark Triad when instructed to do so. Second, the magnitude of response distortion on the Dark Triad is substantial in almost all cases (all but grandiose narcissism) and is particularly pronounced for "look bad" on the psychopathy domain and facets. Third, informants (who were instructed to think of a friend and report on their friend's characteristics) can distort their responses at least as much as people rating themselves. In the case of "look good" on psychopathy, informants may even be distorting more than people who self-rate.

Although people clearly distorted their responses on both self- and informant-reports, the direction of response distortion was not uniform across the measures of the Dark Triad. Narcissism and psychopathy scores decreased under "look good" and increased under "look bad" conditions (in line with hypotheses). However, results were more complicated for Machiavellianism. When measured using the Dirty Dozen, the direction of results were as expected, such that scores decreased for faking-good and increased for faking-bad. However, when measured using the FFMI (Collison et al., 2018), the direction for Machiavellianism scores was in the opposite direction. That is, scores increased under "look good" instructions and decreased under "look bad" instructions. This was consistent across both self- and informant-reports, but is in stark contrast prior research on Machiavellianism, where scores increased for "look bad" and decreased for "look good" (Skinner et al., 1976; Skinner, 1982; Young, 2018).

The unexpected pattern of results for Machiavellianism may be due to the instrument used (The FFMI), which is based on 13 facets of the Five Factor Model of personality (Collison et al., 2018). Big five facets are known to include evaluative items that enhance socially desirable responses (Bäckström et al., 2009; John & Robins, 1993)—to increase under "look good" instructions and decrease under "look bad" (Viswesvaran & Ones, 1999). Furthermore, the FFMI was developed from an inventory developed to capture normal variance in personality, and although the FFMI may capture elements of Machiavellianism, it is unlikely to capture variance at the upper end of the scale (Suzuki et al., 2015). The FFMI shows large positive correlations with several facets of Conscientiousness, with the largest for Self-Discipline, Competence, and Achievement-Striving (r = .44–0.48, Kückelhaus et al., 2021). Of the six conscientiousness facets, these three (i.e., those with the strongest associations with the FFMI) also show the largest score increases under instructions to look good (with effect sizes of 1.16–1.58; Ziegler et al., 2010). As such, it is not surprising that the FFMI shows a pattern of response distortion that is inconsistent with other Dark Triad scales, but consistent with response distortion on Big Five measures. Where Dark Triad researchers are usually concerned about test-takers suppressing their Dark Triad scores, these findings, while preliminary, suggest that scores on the FFMI may be inflated rather than suppressed when there is motivation to distort responses.

4.1 | Are informant-reports the solution to faking?

Concern relating to the susceptibility of self-report scales to response distortion is not a new observation. Since the early 20th century, it has been suggested that personality scales may be measuring social desirability rather than the personality construct itself (Allport, 1928; Edwards, 1957; Meehl & Hathaway, 1946). Consequently, there has been considerable focus on mitigating these biases in personality research, particularly within the context of the Big Five/FFM models of personality (MacCann, 2013). The present study sought to extend prior research and examine these phenomena within the context of the Dark Triad of personality. As the use of self-report rating scales have been criticized for potential response distortion, the use of informant-reports became a popular way to confirm the accuracy of self-reported personality traits (Kim et al., 2019; Vazire, 2006) despite early suggestions that informant-reports may be susceptible to halo and friendship effects (Allport, 1928).

A recent meta-analysis examining self-other agreement concluded the high convergence between self- and informant-reports was evidence for limited response distortion on self-report scales (Kim et al., 2019), indicating that informants may be less prone to social desirability biases. Kim et al. explored additional factors that may impact the accuracy of self and informant-reports, such as the extent to which the evaluativeness of a trait could impact the mean differences found between self- and informant-reports. Specifically, they hypothesized that more evaluative traits would show larger mean differences compared to less evaluative traits. Although this hypothesis was not supported, there is a need for future research to understand the role of item valence and responses to evaluative traits. Additionally, Kim et al. discuss the role of self-informant closeness in informant-report bias, proposing that friendship may introduce biases similar to self-report response distortion. These suggestions support prior research which shows supervisors, for instance, can distort their responses to rate their employee's conscientiousness and extraversion more favorably (König et al., 2017).

Building on these ideas, the nuanced role of social desirability across both self- and informant-reports requires a more refined discussion. Our results suggest that honest informant-reports are generally more favorable than honest self-reports (except for PNI vulnerable narcissism and LSRP egocentricity). When considering the Dark Triad dimensions, it is possible that "low visibility" trait characteristics, such as lacking remorse, may introduce uncertainty in the informant's responses. When informants are uncertain about the internal workings of their target, they may be more inclined to rate the target favorably thereby introducing a form of social desirability divergent from the more traditional understanding of impression management, for example.

Therefore, the question "Are informant-reports the solution to faking?" is less fundamental than exploring how different types of social desirability biases, such as self-deceptive enhancement and impression management, manifest differently across self- and informant-reports. As far as informant-reports are concerned, more

favorable responses are not necessarily driven solely by friendship effects, but could also be influenced by the low visibility of Dark Triad dimensions, as well as the evaluative content included in the scales measuring these dimensions.

4.2 | Future directions

Although these results show that people can distort their responses on self- and informant-reported subclinical Dark Triad measures, results do not indicate the extent to which this occurs in practice, nor the motivations behind why someone would distort their responses for someone else. Further research should examine the underlying motivations and antecedents of informant response distortion. For example, are informant response distortions driven by how much the informant likes the target (Leising et al., 2010), whether there is a sense of reciprocity (e.g., when the informant needs to look good, they will expect the target to reciprocate), or driven by other prosocial or nonprosocial reasons. In addition, there may be more nefarious reasons someone may "look good" for someone else. For example, if an employer hopes to "get rid" of an employee, the employer may be inclined to provide a glowing review of that staff member to ensure they get another job. Importantly, the motivations underlying distorting one's responses on an informant-report may differ substantially than on a self-report. For example, distorting to look bad on self-reported psychopathy may be less likely for a job interview where a candidate can simply decline the job, but an informant distorting to make someone else "look bad" when reporting their suitability for a job is only likely in a real-world scenario where this benefits the informant, or the informant simply does not like the target. There are myriad potential motivations underlying why someone may be inclined to distort informant-reports, each requiring further examination.

Several aspects of the study design warrant discussion. Although we were interested in comparing the individual differences between people, an optimal design for future research could include dyadic or multi-informant self and informant ratings. Dyadic or multi-informant ratings may be used to measure not only the extent to which people can distort their responses on Dark Triad measures, but also measure sources of agreement and disagreement. Additionally, using a dyadic or multi-informant design will provide the opportunity to examine whether the psychometric properties of the Dark Triad measures are affected as a result of distorting to look good, or look bad. Although the results of this study offer insights into the comparison of self- and informant-ratings relating to Dark Triad traits, the interpretation of these results should consider assumptions inherent to the research design. Informants were required to think of a close friend of the same age and sex as the informant and to provide ratings of their chosen friend (either honest, or fake good/bad). One assumption inherent to this design is that the average latent Dark Triad trait reflected in the self-report rating groups is the same as/similar to that in the targets that were rated by informants. However, even though individuals with higher Dark Triad scores are socially perceived similarly to any other member of a social group (Rogoza et al., 2021), there may be observing genuine differences in the traits between the informant and their target. This could potentially lead to an over- or underestimation of the effects of self- versus informant-report ratings on the Dark Triad traits.

An instructed faking paradigm is a useful, and necessary starting point to determine the extent to which people can distort their responses on assessments. Substantial additional research is needed in this area to go beyond the artificial nature of lab-based instructed faking paradigms to determine the extent to which people do fake on these measures, the extent to which response distortion impacts their predictive validity, and the practical consequences of informant-report response distortion.

4.3 | Conclusion

This study provides the first comprehensive assessment of instructed faking investigating the extent to which faking can occur on both self- and informant-reports of the Dark Triad. The current data confirm that people can substantially distort their responses on self-report measures of the Dark Triad. One of the more significant findings to emerge from this study is that informants can also distort their responses on behalf of their target. Before this study, there was limited evidence that people can distort their responses on informant-reports. Additionally, these results suggest that the Dark Triad measures may contain a substantial amount of evaluative content. These results demonstrate the importance of closely examining the extent to which response distortion is relevant for self- and informant-reports. If people self-report themselves favorably and informants also rate their targets favorably, then the Dark Triad assessment utility is diminished. Although self-report scales continue to undergo intense scrutiny regarding their accuracy in the face of response distortion there has not been the same scrutiny for informant-reports. This study has established that informants can distort their responses and lays the foundation for future research to continue to explore informant response distortion. In particular, future research could address whether informants *do* distort their responses in a high-stakes real-world context, and what the implications are for score interpretation if they do. For confidence in informant-reports to continue, a similar level of scrutiny must be applied to the psychometric properties of informant-reports of personality.

ACKNOWLEDGMENTS

Open access publishing facilitated by The University of Sydney, as part of the Wiley - The University of Sydney agreement via the Council of Australian University Librarians.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in OSF at https://osf.io/eb49g/?view_only=28d4eceb865049558c8b1818596cb684.

ORCID

Sarah A. Walker http://orcid.org/0000-0002-6767-8604

Carolyn MacCann http://orcid.org/0000-0001-7789-6368

ENDNOTES

[1] The preregistrations states "ratings target" here; however, we have altered this to "rater" (i.e., self or informant) for clarity.

[2] We preregistered dummy-coding (rather than contrast coding) for comparison of conditions (honest vs. fake-good vs. fake-bad). Contrast coding is therefore a minor departure from the preregistration (but we feel that it is clearer). We did not conduct the dummy-coded analysis.

REFERENCES

Allport, G. W. (1928). A test for ascendance-submission. *The Journal of Abnormal and Social Psychology*, *23*(2), 118–136. https://doi.org/10.1037/h0074218

Anderson, J. L., Sellbom, M., Wygant, D. B., & Edens, J. F. (2013). Examining the necessity for and utility of the psychopathic personality inventory–revised (PPI–R) validity scales. *Law and Human Behavior*, *37*, 312–320. https://doi.org/10.1037/lhb0000018

Arthur, Jr, W., Glaze, R. M., Villado, A. J., & Taylor, J. E. (2010). The magnitude and extent of cheating and response distortion effects on unproctored Internet-based tests of cognitive ability and personality. *International Journal of Selection and Assessment*, *18*(1), 1–16. https://doi.org/10.1111/j.1468-2389.2010.00476.x

Bäckström, M., Björklund, F., & Larsson, M. R. (2009). Five-factor inventories have a major general factor related to social desirability which can be reduced by framing items neutrally. *Journal of Research in Personality*, *43*(3), 335–344. https://doi.org/10.1016/j.jrp.2008.12.013

Bartone, P. T. (2010). Preventing prisoner abuse: leadership lessons of Abu Ghraib. *Ethics & Behavior*, *20*(2), 161–173. https://doi.org/10.1080/10508421003595984

Beckmann, N., Birney, D. P., Beckmann, J. F., Wood, R. E., Sojo, V., & Bowman, D. (2020). Inter-individual differences in intra-individual variability in personality within and across contexts. *Journal of Research in Personality*, *85*, 103909. https://doi.org/10.1016/j.jrp.2019.103909

Berg, J. M., Lilienfeld, S. O., & Waldman, I. D. (2013). Bargaining with the devil: Using economic decision-making tasks to examine the heterogeneity of psychopathic traits. *Journal of Research in Personality*, *47*(5), 472–482. https://doi.org/10.1016/j.jrp.2013.04.003

Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, *14*(4), 317–335. https://doi.org/10.1111/j.1468-2389.2006.00354.x

Brinkley, C. A., Diamond, P. M., Magaletta, P. R., & Heigel, C. P. (2008). Cross-validation of Levenson's Psychopathy Scale in a sample of federal female inmates. *Assessment*, *15*, 464–482. https://doi.org/10.1177/1073191108319043

Christian, E., & Sellbom, M. (2016). Development and validation of an expanded version of the three-factor Levenson Self-Report Psychopathy Scale. *Journal of Personality Assessment*, *98*(2), 155–168. https://doi.org/10.1080/00223891.2015.1068176

Christie, R., & Geis, F. (1970). Scale construction. *Studies in Machiavellianism*, *34*(4), 10–34.

Cleckley, H. M. (1951). The mask of sanity. *Postgraduate Medicine*, *9*(3), 193–197.

Clifton, A., Turkheimer, E., & Oltmanns, T. F. (2005). Self-and peer perspectives on pathological personality traits and interpersonal problems. *Psychological Assessment*, *17*(2), 123–131. https://doi.org/10.1037/1040-3590.17.2.123

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Collison, K. L., Vize, C. E., Miller, J. D., & Lynam, D. R. (2018). Development and preliminary validation of a five factor model measure of Machiavellianism. *Psychological Assessment*, *30*(10), 1401–1407. https://doi.org/10.1037/pas0000637

Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, *136*(6), 1092–1122. https://doi.org/10.1037/a0021212

Dickinson, K. A., & Pincus, A. L. (2003). Interpersonal analysis of grandiose and vulnerable narcissism. *Journal of Personality Disorders*, *17*(3), 188–207. https://doi.org/10.1521/pedi.17.3.188.22146

Donaldson, S. I., & Grant-Vallone, E. J. (2002). Understanding self-report bias in organizational behavior research. *Journal of Business and Psychology*, *17*(2), 245–260. https://doi.org/10.1023/A:1019637632584

Dunlop, P. D., Lee, K., Ashton, M. C., Butcher, S. B., & Dykstra, A. (2015). Please accept my sincere and humble apologies: The HEXACO model of personality and the proclivity to apologize. *Personality and Individual Differences*, *79*, 140–145. https://doi.org/10.1016/j.paid.2015.02.004

Edwards, A. L. (1957). The social desirability variable in personality assessment and research. *Academic Medicine*, *33*(8), 610–611.

Furnham, A. (1990). Faking personality questionnaires: Fabricating different profiles for different purposes. *Current Psychology*, *9*(1), 46–55. https://doi.org/10.1007/BF02686767

Furnham, A., Richards, S. C., & Paulhus, D. L. (2013). The Dark Triad of personality: A 10 year review. *Social and Personality Psychology Compass*, *7*(3), 199–216. https://doi.org/10.1111/spc3.12018

Garofalo, C., Noteborn, M. G. C., Sellbom, M., & Bogaerts, S. (2019). Factor structure and construct validity of the Levenson Self-Report Psychopathy Scale (LSRP): A replication and extension in Dutch nonclinical participants. *Journal of Personality Assessment*, *101*(5), 481–492. https://doi.org/10.1080/00223891.2018.1519830

Grös, D. F., Antony, M. M., Simms, L. J., & McCabe, R. E. (2007). Psychometric properties of the state-trait inventory for cognitive and somatic anxiety (STICSA): Comparison to the state-trait anxiety inventory (STAI). *Psychological Assessment*, *19*(4), 369–381. https://doi.org/10.1037/1040-3590.19.4.369

Hare, R. D. (2003). *The psychopathy checklist–Revised*. Multi-Health Systems.

Hogan, R., & Foster, J. (2016). Rethinking personality. *International Journal of Personality Psychology*, *2*, 37–43.

Holden, R. R., & Passey, J. (2010). Socially desirable responding in personality assessment: Not necessarily faking and not necessarily substance. *Personality and Individual Differences*, *49*(5), 446–450. https://doi.org/10.1016/j.paid.2010.04.015

Hollander, E. P. (1956). The friendship factor in peer nominations. *Personnel Psychology*, *9*, 435–447. https://doi.org/10.1111/j.1744-6570.1956.tb01078.x

John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The big five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality*, *61*(4), 521–551. https://doi.org/10.1111/j.1467-6494.1993.tb00781.x

Jonason, P. K., & Webster, G. D. (2010). The dirty dozen: A concise measure of the dark triad. *Psychological Assessment*, *22*(2), 420–432. https://doi.org/10.1037/a0019265

Karpman, B. (1941). On the need of separating psychopathy into two distinct clinical types: the symptomatic and the idiopathic. *Journal of Criminal Psychopathology*, *3*, 112–137.

Kim, H., Di Domenico, S. I., & Connelly, B. S. (2019). Self–other agreement in personality reports: A meta-analytic comparison of self-and informant-report means. *Psychological Science*, *30*(1), 129–138. https://doi.org/10.1177/0956797618810000

Klonsky, E. D., Jane, J. S., Turkheimer, E., & Oltmanns, T. F. (2002). Gender role and personality disorders. *Journal of Personality Disorders*, *16*(5), 464–476. https://doi.org/10.1521/pedi.16.5.464.22121

Klonsky, E. D., Oltmanns, T. F., & Turkheimer, E. (2002). Informant reports of personality disorder: Relation to self-reports and future research directions. *Clinical Psychology: Science and Practice*, *9*, 300–311. https://doi.org/10.1093/clipsy.9.3.300

König, C. J., Steiner Thommen, L. A., Wittwer, A. M., & Kleinmann, M. (2017). Are observer ratings of applicants' personality also faked? Yes, but less than self-reports. *International Journal of Selection and Assessment*, *25*(2), 183–192. https://doi.org/10.1111/ijsa.12171

Kückelhaus, B. P., Blickle, G., Kranefeld, I., Körnig, T., & Genau, H. A. (2021). Five factor Machiavellianism: Validation of a new measure. *Journal of Personality Assessment*, *103*(4), 509–522. https://doi.org/10.1080/00223891.2020.1784182

Lee, K., & Ashton, M. C. (2014). The Dark Triad, the Big Five, and the HEXACO model. *Personality and Individual Differences*, *67*, 2–5. https://doi.org/10.1016/j.paid.2014.01.048

Leising, D., Erbs, J., & Fritz, U. (2010). The letter of recommendation effect in informant ratings of personality. *Journal of Personality and Social Psychology*, *98*(4), 668–682. https://doi.org/10.1037/a0018771

Levenson, M. R., Kiehl, K. A., & Fitzpatrick, C. M. (1995). Assessing psychopathic attributes in a noninstitutionalized population. *Journal of Personality and Social Psychology*, *68*(1), 151–158. https://doi.org/10.1037/0022-3514.68.1.151

MacCann, C. (2013). Instructed faking of the HEXACO reduces facet reliability and involves more Gc than Gf. *Personality and Individual Differences*, *55*(7), 828–833. https://doi.org/10.1016/j.paid.2013.07.007

Meehl, P. E., & Hathaway, S. R. (1946). The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *Journal of Applied Psychology*, *30*(5), 525–564. https://doi.org/10.1037/h0053634

Miller, J. D., & Campbell, W. K. (2008). Comparing clinical and social-personality conceptualizations of narcissism. *Journal of Personality*, *76*(3), 449–476. https://doi.org/10.1111/j.1467-6494.2008.00492.x

Miller, J. D., Hoffman, B. J., Gaughan, E. T., Gentile, B., Maples, J., & Keith Campbell, W. (2011). Grandiose and vulnerable narcissism: A nomological network analysis. *Journal of Personality*, *79*(5), 1013–1042. https://doi.org/10.1111/j.1467-6494.2010.00711.x

Miller, J. D., & Lynam, D. R. (2012). An examination of the Psychopathic Personality Inventory's nomological network: A meta-analytic review. *Personality Disorders: Theory, Research, and Treatment*, *3*(3), 305–326. https://doi.org/10.1037/a0024567

Mills, J. F., & Kroner, D. G. (2005). An investigation into the relationship between socially desirable responding and offender self-report. *Psychological services*, *2*(1), 70–80. https://doi.org/10.1037/1541-1559.2.1.70

Mount, M. K., Barrick, M. R., & Strauss, J. P. (1994). Validity of observer ratings of the big five personality factors. *Journal of Applied Psychology*, *79*(2), 272–280. https://doi.org/10.1037/0021-9010.79.2.272

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). Lawrence Erlbaum Associates Publishers.

Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (Vol. 1, pp. 224–239). The Guilford Press.

Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, *36*(6), 556–563. https://doi.org/10.1016/S0092-6566(02)00505-6

Quilty, L. C., Cosentino, N., & Bagby, R. M. (2018). Response bias and the personality inventory for DSM–5: Contrasting self- and informant-report. *Personality Disorders: Theory, Research, and Treatment*, *9*(4), 346–353. https://doi.org/10.1037/per0000246

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), 1–14. https://doi.org/10.1037/a0026482

Rodman, A. M., Powers, K. E., & Somerville, L. H. (2017). Development of self-protective biases in response to social evaluative feedback. *Proceedings of the National Academy of Sciences*, *114*(50), 13158–13163. https://doi.org/10.1073/pnas.1712398114

Rogers, R., & Bender, S. D. (2018). *Clinical assessment of malingering and deception* (Fourth Edition). Guilford Publications.

Rogers, R., Sewell, K. W., Martin, M. A., & Vitacco, M. J. (2003). Detection of feigned mental disorders: A meta-analysis of the MMPI-2 and malingering. *Assessment*, *10*(2), 160–177. https://doi.org/10.1177/1073191103010002007

Rogoza, R., Danieluk, B., Kowalski, C. M., Kwiatkowska, K., & Kwiatkowska, M. M. (2021). Making and maintaining relationships through the prism of the dark triad traits: A longitudinal social network study. *Journal of Personality*, *89*(2), 338–356.

Schoenleber, M., Roche, M. J., Wetzel, E., Pincus, A. L., & Roberts, B. W. (2015). Development of a brief version of the pathological narcissism inventory. *Psychological Assessment*, *27*(4), 1520–1526. https://doi.org/10.1037/pas0000158

Van Scotter, J. R., & Roglio, K. D. D. (2020). CEO bright and dark personality: Effects on ethical misconduct. *Journal of Business Ethics*, *164*(3), 451–475. https://doi.org/10.1007/s10551-0018-4061-5

Sellbom, M. (2011). Elaborating on the construct validity of the Levenson Self-Report Psychopathy Scale in incarcerated and non-incarcerated samples. *Law and Human Behavior*, *35*, 440–451. https://doi.org/10.1007/s10979-010-9249-x

Semrad, M., & Scott-Parker, B. (2020). Police, personality and the ability to deceive. *International Journal of Police Science & Management*, *22*(1), 50–61. https://doi.org/10.1177/1461355719880568

Skinner, N. F. (1982). Personality correlates of Machiavellianism: III. A simulation procedure for identifying high Machs. *Social Behavior and Personality: An International Journal*, *10*(2), 197–199. https://doi.org/10.2224/sbp.1982.10.2.197

Skinner, N. F., Giokas, J. A., & Hornstein, H. A. (1976). Personality correlates of Machiavellianism: 1. Consensual validation. *Social Behavior and Personality: An International Journal*, *4*(2), 273–276. https://doi.org/10.2224/sbp.1976.4.2.273

Sleep, C. E., Sellbom, M., Campbell, W. K., & Miller, J. D. (2017). Narcissism and response validity: Do individuals with narcissistic features underreport psychopathology? *Psychological Assessment*, *29*(8), 1059–1064. https://doi.org/10.1037/pas0000413

Spain, S. M., Harms, P., & LeBreton, J. M. (2014). The dark side of personality at work. *Journal of Organizational Behavior*, *35*(S1), S41–S60. https://doi.org/10.1002/job.1894

Suzuki, Y., Tamesue, D., Asahi, K., & Ishikawa, Y. (2015). Grit and work engagement: A cross-sectional study. *PloS one*, *10*(9), e0137501. https://doi.org/10.1371/journal.pone.0137501

Vazire, S. (2006). Informant reports: A cheap, fast, and easy method for personality assessment. *Journal of Research in Personality*, *40*(5), 472–481. https://doi.org/10.1016/j.jrp.2005.03.003

Vazire, S., & Carlson, E. N. (2010). Self-knowledge of personality: Do people know themselves? *Social and Personality Psychology Compass*, *4*(8), 605–620. https://doi.org/10.1111/j.1751-9004.2010.00280.x

Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement*, *59*(2), 197–210. https://doi.org/10.1177/00131649921969802

Walker, S. A., Double, K. S., Birney, D. P., & MacCann, C. (2022). How much can people fake on the dark triad? A meta-analysis and systematic review of instructed faking. *Personality and Individual Differences*, *193*, 111622. https://doi.org/10.1016/j.paid.2022.111622

Walker, S. A., & MacCann, C. (2023). Faking good on self-reports versus informant-reports of emotional intelligence. *Assessment*. Advance online publication. 0731911231203960. https://doi.org/10.1177/107319112312039

Webber, T. A., Sullivan-Baca, E., Modiano, Y. A., Taiwo, Z., & Grabyan, J. M. (2022). Validity of informant report interpretations: Role of examinee performance and symptom invalidity. *Psychological Assessment*, *34*(2), 125–138. https://doi.org/10.1037/pas0001074

Wink, P. (1991). Two faces of narcissism. *Journal of Personality and Social Psychology*, *61*(4), 590–597. https://doi.org/10.1037/0022-3514.61.4.590

Young, A. L. (2018). *Faking resistance of a forced-choice measure of the dark triad* (Order No. 11017976). Available from ProQuest Central Student. (2195372959). Retrieved from https://search.proquest.com/docview/2195372959?accountid=6724

Zhang, H., Luo, Y., Zhao, Y., Zhang, R., & Wang, Z. (2017). Differential relations of grandiose narcissism and vulnerable narcissism to emotion dysregulation: Self-esteem matters. *Asian Journal of Social Psychology*, *20*(3–4), 232–237. https://doi.org/10.1111/ajsp.12191

Ziegler, M., Danay, E., Schölmerich, F., & Bühner, M. (2010). Predicting academic success with the Big 5 rated from different points of view: Self-rated, other rated and faked. *European Journal of Personality*, *24*, 341–355. https://doi.org/10.1002/per.753

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Walker, S. A., MacCann, C. (2024). Faking good and bad on self-reports versus informant-reports of Dark Triad personality. *International Journal of Selection and Assessment*, 1–14. https://doi.org/10.1111/ijsa.12465