Article

Assessment
2018, Vol. 25(4) 513–526

DOI: 10.1177/1073191116641181
journals.sagepub.com/home/asm
SAGE

Forced-Choice Assessment of Work-Related Maladaptive Personality Traits: Preliminary Evidence From an Application of Thurstonian Item Response Modeling

Nigel Guenole[1,2], Anna A. Brown[3], and Andrew J. Cooper[1]

Abstract

This article describes an investigation of whether Thurstonian item response modeling is a viable method for assessment of maladaptive traits. Forced-choice responses from 420 working adults to a broad-range personality inventory assessing six maladaptive traits were considered. The Thurstonian item response model's fit to the forced-choice data was adequate, while the fit of a counterpart item response model to responses to the same items but arranged in a single-stimulus design was poor. Monotrait heteromethod correlations indicated corresponding traits in the two formats overlapped substantially, although they did not measure equivalent constructs. A better goodness of fit and higher factor loadings for the Thurstonian item response model, coupled with a clearer conceptual alignment to the theoretical trait definitions, suggested that the single-stimulus item responses were influenced by biases that the independent clusters measurement model did not account for. Researchers may wish to consider forced-choice designs and appropriate item response modeling techniques such as Thurstonian item response modeling for personality questionnaire applications in industrial psychology, especially when assessing maladaptive traits. We recommend further investigation of this approach in actual selection situations and with different assessment instruments.

Keywords

maladaptive personality, forced choice, Thurstonian IRT

Extensive use of the Minnesota Multiphasic Personality Inventory (MMPI; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) for police officer selection (e.g., Bartol, 1991; Costello, Schneider, & Schoenfeld, 1996; Hiatt & Hargrave, 1988; Kauder & Thomas, 2003) shows that industrial psychologists have long been interested in assessing maladaptive personality traits in occupational settings. However, legal challenges to the appropriateness of the MMPI, along with the emergence of the five-factor model of personality, led to a reduction in research focus on maladaptive traits in the workplace. Selection practitioners have recently expressed renewed interest in the measurement of subclinical maladaptive personality (Guenole, 2014; Schyns, 2015; Spain, Harms, & LeBreton, 2014). This interest is heightened following recurring examples of senior business executives behaving in ethically inappropriate ways (e.g., Hargie, Stapleton, & Tourish, 2010), and has been spurred on by adaptations of clinical personality instruments to make them more suited to workplace applications.

It is unlikely that selection based on psychometric testing alone will serve as a remedy for executives' poor behavior at work, a problem that has been noted to be "irregular, complex and extends well beyond the boundary of the firm" (Hess & Broughton, 2014, p. 542). However, recent research has linked maladaptive personality traits to outcomes including contextual performance (Wille, De Fruyt, & De Clercq, 2013), career development (De Fruyt, Wille, & Furnham, 2013), and leader effectiveness and emergence (Harms, Spain, & Hannah, 2011; Hogan & Hogan, 2001; Kaiser, LeBreton, & Hogan, 2015; Padilla, Hogan, & Kaiser, 2007). A number of recent meta-analyses of both broad maladaptive models, as well as particular constellations of maladaptive traits, further underscore the idea that maladaptive personality assessment has a role to play in mitigating the risk of bad behavior at work (Gaddis & Foster, 2015; Grijalva, Harms, Newman, Gaddis, & Fraley,

[1]Goldsmiths, University of London, London, UK
[2]IBM Smarter Workforce, London, UK
[3]University of Kent, Canterbury, Kent, UK

Corresponding Author:
Nigel Guenole, Institute of Management Studies, Goldsmiths, University of London, New Cross, London SE14 6NW, UK.
Email: n.guenole@gold.ac.uk

2015; O'Boyle, Forsyth, Banks, & McDaniel, 2012). For these reasons, awareness is growing that focused assessment of the maladaptive poles of personality could have positive benefits.

Self-Report Assessments of Maladaptive Personality Traits

The most common approach to assessing maladaptive traits is to use single-stimulus self-reports. In addition to being cost effective and easy to obtain, Paulhus and Vazire (2007, p. 227) noted that selves have " . . . the opportunity to observe a wide range of behaviors covering a wide swath of time. These include behaviors that are typically performed in private—for example, masturbation, academic cheating, napping, and singing in the shower." They also argued that self-reports are directly interpretable, provided by the person with the most interest in the self, and for constructs like self-efficacy, it may be the only valid report.

No source, self-report or other, has been shown to be more valid for all traits in all situations. For instance, whether a self- or informant report is most accurate in the context of maladaptive personality assessment depends on whether the disorder is related to internalizing or externalizing (Carlson, Vazire, & Oltmanns, 2008). Vazire and Mehl (2008) have shown that self- and informant ratings sometimes can have incremental predictive power over one another. These studies support the self–other knowledge asymmetry model (Vazire, 2010), which builds on the realistic accuracy model (Funder, 1995) to identify conditions under which self-perceptions and other perceptions are likely to be most accurate.

Despite their benefits, self-reports present interpretational complexity. Selves may not be able to accurately report on their traits (Klonsky, Oltmanns, & Turkheimer, 2002). In fact, it is well established that this itself is sometimes diagnostic (Meehl, 1945). Self-reports are also open to systematic and nonsystematic distortions. Systematic response styles such as preferences for scale midpoints (Hamamura, Heine, & Paulhus, 2008; Lee, Jones, Mineyama, & Zhang, 2002), acquiescence (Messick, 1967), or extreme responding (Jin & Wang, 2014) and nonsystematic response distortion, such as careless responding (Meade & Craig, 2012), do not require item comprehension. Socially desirable responding (SDR; Paulhus, 1984, 2002) and malingering (Rogers, 2008) are forms of systematic bias on self-reports that do require item comprehension.

Controlling Response Bias

Response bias indicators, such as the Marlowe–Crowne social desirability index (Crowne & Marlowe, 1960), try to capture evidence of distortion. Major personalities inventories contain such scales, including the California Psychological Inventory (Gough & Bradley, 1996), the NEO Personality Inventory–Revised (McCrae & Costa, 2010), and the MMPI (Butcher et al., 1989). McGrath, Mitchell, Kim, and Hough (2010) examined evidence of whether response bias indicators suppressed scores or moderated the validity of substantive scales, finding little support for either hypothesis. However, their review did not entirely rule out that response bias might be a problem. Instead, response bias indicators might be a poor approach for controlling response bias, as they might measure substantive personality traits, or the effects might be too subtle for response bias indicators to detect.

Notwithstanding the true extent of response bias, model-based, rather than observed-score, methods to control the effects of self-report dissimulation are receiving considerable attention. Among the methods advocated are applying multiple process item response models to separate content-related latent factors from latent factors that capture response styles (Böckenholt, 2012; Plieninger & Meiser, 2014); using mixture modeling to separate latent classes of "fakers" and "honest respondents" (e.g., Zickar, Gibby, & Robie, 2004); comparing observed response patterns with expected response patterns at the item level using optimal appropriateness techniques (Zickar & Drasgow, 1996), and applying multinomial processing trees (Khorramdel & von Davier, 2014).

Forced-Choice Designs. An alternative approach to controlling response biases is preventing biases from occurring in the first place. Forced-choice questionnaires present candidates with several stimuli at a time, asking to rank or partially rank items. Significant research has been conducted on the California Q-Sort where clinicians rank-order 100 statements according to how descriptive each statement is (e.g., Colvin, Block, & Funder, 1995; Kolar, Funder, & Colvin, 1996). There is little factor-analytic support for the Q-Sort's structure (John & Naumann, 2010), however, and research has reported item-ordering effects (Serfass & Sherman, 2013). Today, it is common that items are organized in blocks wherein items assess different constructs. Most commonly, pairs, triads, or quads are used. An example of a triad block measuring traits of antagonism, detachment, and disinhibition might ask a candidate which of the three statements is most like them and which is least like them:

A. Put rude people in their place
B. Prefer being alone
C. Act quickly and manage the consequences later

Challenges Presented by Forced-Choice Designs. The forced-choice approach is not without its own challenges. While increasing resistance to generalized faking where individuals try to favorably respond against all traits assessed, it is still possible that a candidate can consistently fake a particular trait, or even set of traits. Test developers try to

counteract this by placing items of similar desirability together in blocks. Another challenge is that classical scoring of forced-choice questionnaires results in the same total score for every individual—so-called ipsative data (Clemans, 1966; Meade, 2004). Problems of ipsative data include biased reliability estimates, biased estimates of construct and criterion-related validity due to the constraint that covariances of all traits necessarily sum to zero, and biased individual scores (Brown & Maydeu-Olivares, 2013).

Solving the Challenge of Ipsative Data. Several groups of researchers have proposed models for accurately modeling the forced-choice response process (Andrich, 1989, 1995; Brown & Maydeu-Olivares, 2011; McCloy, Heggestad, & Reeve, 2005; Stark, Chernyshenko, & Drasgow, 2005; Stark, Chernyshenko, Drasgow, & White, 2012; Zinnes & Griggs, 1974). For an overview of these different methods, the reader is referred to Brown (2014). We focus on the Thurstonian item response modeling approach (Brown & Maydeu-Olivares, 2011, 2012), because it is the only method at present that allows estimating multidimensional forced-choice data using commonly available structural equation modeling software such as M*plus* (Muthén & Muthén, 1998-2010). So far, Thurstonian item response modeling has been applied to normal-range personality questionnaires and other noncognitive constructs (e.g., Anguiano-Carrasco, MacCann, Geiger, Seybert, & Roberts, 2014; Brown & Maydeu-Olivares, 2011, 2013; Joubert, Inceoglu, Bartram, Dowdeswell, & Lin, 2015). But maladaptive personality is a natural application for Thurstonian approach because response distortion is likely to be exacerbated beyond levels seen with normal-range questionnaires.

Models of Maladaptive Personality

Maladaptive Personality Models in Industrial Psychology. Guenole (2014) classified the existing maladaptive models used in occupational settings into four broad categories. First, specific traits linked with health-related outcomes have been examined (e.g., Judge & LePine, 2007). Second, models that use dimensional representations of *Diagnostic and Statistical Manual of Mental Disorders–Fourth edition* (*DSM-IV*) have been extensively studied (Hogan & Hogan, 1997, 2001; Moscoso & Salgado, 2004). Third, coherent clusters of personality traits that often manifest together, such as the "dark triad" of narcissism, Machiavellianism, and psychoticism have been studied (Paulhus & Williams, 2002; Wu & LeBreton, 2011). Along with extensions to include sadism, this model is sometimes represented as a "dark quad" (Furnham, Richards, & Paulhus, 2013). Fourth, compounds of normal-range personality that predict dysfunction have been studied (Wille et al., 2013).

These maladaptive frameworks have advanced our understanding of personality at work in important ways. Nevertheless, drawbacks are associated with each. Studies of specific health-related personality traits would benefit from a taxonomic structure that is better grounded, both theoretically and empirically. For example, larger criterion-related validities have been observed when traits were studied using the Big Five framework at the domain trait level instead of frameworks that contain blends of the Big Five traits (Salgado, 2003). While the dark triad (and quad) represents a subset of important troublesome personality traits, there is more to maladaptive personality than these particular traits. Finally, the normal personality items that are reconstituted to form maladaptive compounds have shown coherent relations with maladaptive personality, but they do not contain sufficiently extreme item content to be considered a comprehensive model of maladaptive personality traits. For these reasons, Guenole (2015) recommended assessing medium-range variations of the *DSM-5* maladaptive trait model in occupational settings.

The DSM-5 *Maladaptive Trait Model.* The *DSM-5* personality trait model was developed as part of a new approach to diagnosing personality disorders to address concerns with the diagnostic procedures in *DSM-IV* and earlier editions. Broadly speaking, the earlier typology recommended a categorical interpretation of disorder, while mounting evidence suggests personality disorders are dimensional (e.g., Krueger & Eaton, 2010; Widiger & Simonsen, 2005). The indicators of the disorder types in earlier editions did not adequately differentiate between personality disorders and other mental health problems (Krueger, Markon, Patrick, Benning, & Kramer, 2007). While ultimately not adopted, the new approach was presented in the emerging model section of the *DSM-5* and is receiving considerable research attention (Krueger, 2013; Krueger & Markon, 2013; Thomas et al., 2012). This model posits that personality disorder can be diagnosed based on an individual's standing on a maladaptive trait model and severity of problems with interpersonal functioning (i.e., empathy and intimacy) and goal regulation (i.e., identity and self-direction; American Psychiatric Association [APA], 2013). Definitions of the traits from Skodol et al. (2011) are provided in Table 1. The *DSM-5* does not rule out bipolarity of trait pathology. Rather, the existing evidence is more supportive of a unipolar perspective on trait pathology.

All individuals can be profiled on the types of maladaptive traits included in the model, whether or not they have a personality disorder (O'Connor, 2002; Saulsman & Page, 2004; Trull & Durrett, 2005). Equally important, the breadth of the trait model means the individual maladaptive traits (e.g., neuroticism) and maladaptive trait clusters (e.g., dark triad) that have been studied in occupational psychology are subsumed under the trait model.

DSM-5 *Personality Assessments and the Present Study.* For a comprehensive review of clinical assessments that measure

Table 1. *DSM-5* Domain Trait Definitions From Skodol et al. (2011).

DSM-5 domain trait	Definition	Big Five adaptive variant
Negative Affect	Experiences a wide range of negative emotions (anxiety, depression, guilt/shame, worry, etc.), and the behavioral and interpersonal manifestations of those experiences	Emotional Stability
Detachment	Withdrawal from other people, ranging from intimate relationships to the world at large; restricted affective experience and expression; limited hedonic capacity	Extraversion
Antagonism	Exhibits diverse manifestations of antipathy toward others, and a correspondingly exaggerated sense of self-importance	Agreeableness
Disinhibition	Diverse manifestations of being present- (vs. future- or past-) oriented so that behavior is driven by current internal and external stimuli rather than by past learning and consideration of future consequences	Conscientiousness
Compulsivity[a]	The tendency to think and act according to a narrowly defined and unchanging ideal, and the expectation that this ideal should be adhered to by everyone	Conscientiousness
Psychoticism	Exhibits a range of odd or unusual behaviors and cognitions, including both process (e.g., perception) and content (e.g., beliefs)	Openness

Note. DSM-5 = Diagnostic and Statistical Manual of Mental Disorders–Fifth edition.
[a]In the final version of the maladaptive trait model, Compulsivity content became a reverse-scored facet of the Disinhibition domain trait, indicating compulsivity is the opposite of disinhibition.

traits relevant to the *DSM-5* model, readers are referred to Krueger, Derringer, Markon, Watson, and Skodol (2012). The American Psychiatric Association's official measure, the Personality Assessment for *DSM-5* (PID-5), contains five domain factors that can be decomposed into 37 facets comprising some 220 items. However, the items in this questionnaire are too extreme for use among normal working populations in selection settings. For example, perceptual dysregulation from the psychoticism factor is measured with the item "Sometimes I get this weird feeling that parts of my body feel like they're dead or not really me." Candidates might question the relevance of such items for personnel selection, and these items would provide information at levels of the latent trait where very few working individuals would fall. The PID-5 and most of the measures mentioned previously use self-report rating scales rather than a forced-choice format, which are especially needed in occupational settings to ameliorate dissimulation.

We sought a measure of the *DSM-5* traits where behaviors are sampled that are more extreme than expected in normal-range personality traits, but not so extreme as to be statistically uninformative and face-invalid in occupational contexts. We also sought a relatively brief inventory of domain traits to minimize testing time while still providing accurate markers of the maladaptive model, for example, a questionnaire similar to that of the International Personality Item Pool Big Five markers (Goldberg, 1992) but for the *DSM-5* domain traits. Developing a *DSM-5* domain trait measure that fulfilled the above criteria was the motivation behind the design of the Goldsmiths-50 (G-50) questionnaire (Guenole, 2015).

Development of this brief inventory involved the following key steps. First, four items were written to assess each of target definitions for facets originally proposed for the *DSM-5* trait model by Skodol et al. (2011), save for facets of self-harm (e.g., cutting and burning) and intimacy avoidance (e.g., avoidance of sexual intimacy) which were deemed inappropriate for an occupational questionnaire. An item review process involving subject matter expert ratings followed by discussion to consensus shortened the item pool to three items per facet. Responses to this item pool were collected from 696 English-speaking working adults in the United Kingdom using a convenience sample. Separate categorical item factor analyses using a diagonally weighted least squares estimator were used to identify the best performing items for each factor. The strongest item from each facet for a given domain was selected, and the process was repeated until an omega reliability coefficient (McDonald, 1999) of .70 was achieved.

Resulting item responses were subjected to categorical item factor analyses from a single-factor model through to a seven-factor model. The six-factor model was adopted based on a substantial improvement in fit moving from five to six factors but a negligible improvement for the seven-factor model. The minimum average factor loading for any domain trait in this model was .46. The maximum average cross-loading was just .11. The development process is described in detail in Guenole (2015). The technical documentation for this measure reports uncorrected correlations between the G-50 and the PID-5 domain traits as falling between $r = .65$ and $r = .73$—strong correlations considering that the PID-5 domain traits are measured with between 38 and 53 items per domain trait, whereas the G-50 uses just 8 to 10 items per domain trait.

In the present study, we develop a forced-choice version of the G-50 for comparison with the existing single-stimulus version. The study is exploratory in nature, but based on the robustness of the forced-choice format to

response biases reported in the literature (e.g., Bartram, 2007), we expect to observe improvement in goodness of fit and perhaps in the item discriminations over the single-stimulus method.

Method

Participants

An Internet-based data collection strategy with snowball sampling of researcher personal networks and convenience sampling via advertising on LinkedIn were used. The personal networks sample was almost solely from the United Kingdom, whereas the LinkedIn sample included other English-speaking countries including the United States, Australia, South Africa, and New Zealand. We excluded participants from non-English-speaking countries due to uncertainty of their fluency in English. The personal network data collection resulted in a sample of 204 participants with female overrepresentation (57% female), while the LinkedIn collection yielded a sample of 216 participants with male overrepresentation (58% male). Overall, the sample was balanced for gender. Due to the demographics section being at the end of the questionnaire, the majority of people did not complete other demographic questions. These two samples were combined for the purposes of this study, yielding the overall sample size ($N = 420$). We expect this sample is representative of people who might actually apply for jobs and complete tests such as that being studied in the present article, because our personal networks mainly included working professionals in industries outside academia and LinkedIn is a professional networking group.

Convenience sampling of working adults might still raise concerns about the appropriateness of our sample for the current investigation. Perhaps respondents were experienced test takers able to deduce the aims of the research. In a worst-case scenario, perhaps maladaptive traits led individuals to answer in a style they thought would deliberately derail the comparisons between the single-stimulus and forced-choice formats or "helpful" participants may have responded in a way they thought would support the hypotheses. While this is impossible to rule out, there are reasons to think that this is unlikely. First, the research was positioned as a study into the measurement of maladaptive personality, with the ethical approval of the primary author's institution, in which participants would be asked potentially sensitive questions without emphasizing that scores between single-stimulus and forced-choice formats would be compared. Because the single-stimulus questionnaire was presented first by the time that any helpful (or disruptive) participants discerned that a comparison between response formats was involved, they would have been responding to the forced-choice section designed to be resistant to response distortion.

Measures

The G-50 (Guenole, 2015) described earlier was used as the basis for the development of a forced-choice questionnaire assessing *DSM-5* maladaptive traits among working adults. To this end, the G-50 items were supplemented with 10 additional items from item pools created and tested during the original instrument development, described in Guenole (2015). This provided 10-item measures of each of the six maladaptive domain traits. Compulsivity and impulsiveness are measured separately as opposed to jointly, thus following the initial rationally derived domain trait structure of the *DSM-5* maladaptive trait model. The 60 items were arranged into forced-choice blocks to create an instrument we refer to as the G-60-FC. The G-60-FC consists of 20 blocks of three items (triads) measuring different traits, wherein every triad is approximately balanced for desirability based on the response data described in Guenole (2015). This forced-choice design is completely balanced, with each trait measured by 10 items (appearing in 10 blocks) compared with every other trait exactly four times. For instance, the first triad consisted of the three items with the highest mean rating in their respective scales. The actual blocks and items are presented in Table 2.

The item set used in the G-60-FC was also used to create the single-stimulus counterpart of the maladaptive inventory (G-60-SS), and therefore the content of the G-60-SS exactly parallels the content of the G-60-FC. In the single-stimulus version, a 5-point Likert-type scale ranging from *strongly disagree* to *strongly agree* was used to gather responses. In the present sample, the Cronbach's alphas for the six scales of the G-60-SS were .69 for Antagonism, .78 for Compulsivity, .77 for Detachment, .70 for Disinhibition, .82 for Negative Affect, and .79 for Psychoticism.

Measurement Models

To analyze single-stimulus responses to the G-60-SS, a confirmatory item factor analysis model was fitted to the polychoric correlations of ordered categorical item responses. This approach was selected to evaluate the whole model in its entirety, rather than trait by trait. Due to the large number of traits assessed in the G-60-SS, the full information maximum likelihood estimation of categorical responses would be computationally infeasible. The hypothesized structural model was that of independent clusters (McDonald, 1999), with 10 items indicating each of the maladaptive traits, and no cross-loadings. Figure 1 depicts this measurement model. The model was estimated using the software M*plus* version 7.2 (Muthén & Muthén, 1998-2010). The diagonally weighted least squares estimator with robust errors denoted weighted least squares means and variance adjusted in M*plus* was used here and in all subsequent estimations.

To analyze forced-choice responses, the rankings within blocks were dummy-coded as described in Brown and

Table 2. Items in the Goldsmiths 60-Item Questionnaires (G-60-SS [Single Stimulus] and G-60-FC [Forced-Choice]), With Standardized Factor Loadings in the Corresponding Measurement Models.

Block	Trait	Item	Loading	
			SS	FC
1	CO	1. Avoid danger	.43	.11(n/s)
	DE	2. Have few close friends	.34	.27
	NE	3. Worry about things that might happen	.31	.43
2	CO	4. Believe there is usually a right way to do things	.58	.62
	NE	5. Imagine what others think about me	.35	.65
	PS	6. Sense things in strange ways	.49	.73
3	CO	7. Never give up	.51	.39
	DI	8. Like to do crazy things	.26	**.76**
	PS	9. Have unusual thought processes	.43	.73
4	AN	10. Put rude people in their place	.29	.74
	DE	11. Prefer being alone	.43	.74
	DI	12. Act quickly and manage the consequences later	.41	.66
5	AN	13. Behave boldly	.19	.66
	DI	14. Find concentrating for long periods dull	.56	.57
	PS	15. Have bizarre experiences	.58	.76
6	AN	16. Am better than my colleagues	.43	.51
	CO	17. Do everything properly	.69	.61
	PS	18. Have unconventional beliefs	.51	**.86**
7	DE	19. Lack energy	.49	.77
	DI	20. Throw caution to the wind	.27	.45
	NE	21. Wonder if friends changed their minds about me	.60	.54
8	AN	22. Tell people when they're annoying me	.36	.72
	DE	23. Find daily tasks demoralizing	.49	.74
	NE	24. Am a victim of my emotions	.54	.76
9	CO	25. Am a detail person	.68	.73
	DE	26. Keep interaction to a minimum	.67	.77
	PS	27. Seem eccentric to other people	.56	.76
10	AN	28. Don't back down	.34	.69
	DI	29. Don't overthink important decisions	.22	.34
	NE	30. Often feel down	.73	**.99**
11	DE	31. Keep my distance from colleagues	.51	.52
	NE	32. Have a poor opinion of myself	.49	.85
	PS	33. Wonder why I do the things I do	.65	.62
12	AN	34. Fight fire with fire	.59	**.86**
	CO	35. Believe most things should be done a certain way	.40	.75
	DI	36. Have a short attention span	.67	.72
13	CO	37. Can't stand errors	.57	.76
	DE	38. Have a cold personality	.61	.37
	DI	39. Can't concentrate for long	**.75**	.40
14	AN	40. Am mean to people who deserve it	**.72**	.63
	NE	41. Feel a lot of guilt	.54	.60
	PS	42. Have weird perceptions	**.68**	.74
15	AN	43. Never tell the whole story	.46	.16(n/s)
	CO	44. Am a perfectionist	.54	**.84**
	NE	45. Have a negative outlook	**.83**	.85
16	DE	46. Avoid others whenever possible	**.78**	.72
	DI	47. Don't follow through	.53	.67
	PS	48. Have peculiar mannerisms	.56	.62
17	CO	49. Insist on everything being perfect	**.81**	.73
	DI	50. Get into trouble	.54	.75

(continued)

Table 2. (continued)

Block	Trait	Item	Loading SS	Loading FC
	NE	51. Feel miserable	**.83**	.92
18	AN	52. Can get whatever I want through flattery	.41	.35
	NE	53. Rarely feel happy	.75	.96
	PS	54. Forget how I got places	.53	.61
19	AN	55. See how far I can push people	.66	.57
	CO	56. See failure as a result of lack of effort	.31	.23
	DE	57. Have no feelings to express	.57	**.80**
20	DI	58. Break agreements	.67	.24
	DE	59. Find that people are against me	.72	.77
	PS	60. Feel like someone else is controlling what I do	**.69**	.10(n/s)

Note. AN = Antagonism; CO = Compulsivity; DE = Detachment; DI = Disinhibition; NE = Negative Affect; PS = Psychoticism. Marker items are indicated in bold. Marker items are indicated in bold; n/s = not significant at the 0.05 level.

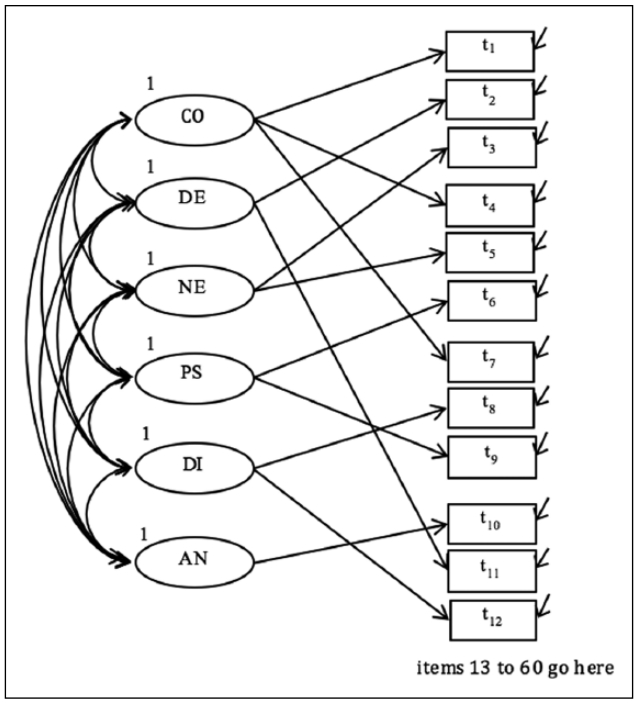

Figure 1. A fragment of the single-stimulus measurement model for Goldsmiths 60-Item Questionnaire–Single Stimulus. *Note.* CO = Compulsiveness; DE = Detachment; NE = Negative Affect; PS = Psychoticism; DI = Disinhibition; AN = Antagonism. Observed ordinal item responses are denoted $t_1 - t_{12}$.

Maydeu-Olivares (2012). Thus, three pairwise comparisons {A, B}, {A, C}, and {B, C} were created from each block of three items A, B, and C, with each pair coded as 1 when the first item was preferred to the second, and 0 otherwise. For instance, if in a triad (A, B, C), a respondent assigned the ranks A = 2, B = 1, C = 3, the corresponding dummy coding would be {A, B} = 0; {A, C} = 1; and {B, C} = 1. Thus, any rank ordering of three items was equivalently coded as a set of three binary outcome variables, each representing a contrast between two items measuring two traits.

A Thurstonian item response model was then fitted to tetrachoric correlations of the dummy-coded binary outcomes. The theoretical basis for this modeling is Thurstone's (1927) law of comparative judgment, which has a long tradition in analysis of forced-choice data (e.g., Maydeu-Olivares & Bockenholt, 2005). According to Thurstone, the observed outcome of each pairwise comparison (i.e., Item A vs. Item B) is determined by the unobserved psychological values (or "utilities") of the two items. Thus, the item with a larger utility for a respondent gets chosen. The utilities of items are unobserved, only the outcome of their implicit comparison (i.e., whether the utility of A was greater that the utility of B, or otherwise) is observed. Hence, the utilities of items are modeled as latent variables, which causally determine the observed outcomes; specifically, the sign of the difference of two utilities, $t_i - t_k$, determines whether the outcome of comparison between items i and k is 1 (when $t_i - t_k > 0$) or 0 (when $t_i - t_k < 0$). The latent utilities, in turn, are assumed to be caused by broader personality traits—latent variables of the higher order. A fragment of this second-order factor model, which is mathematically equivalent to its Thurstonian item response theory counterpart (Brown & Maydeu-Olivares, 2011), but much easier to relate to for applied researchers, presented in Figure 2. Syntax for the Thurstonian model in the present article was generated with the Excel macro supplied in Brown and Maydeu-Olivares (2012). Full technical detail on setting up and testing these models is also provided within that article.

The same factorial structure of the item utilities was hypothesized for the single-stimulus and forced-choice responses, with the only difference that in the single-stimulus format, the observed responses were direct indicators of utilities, while in the forced-choice format, the responses were indicators of differences between pairs of item utilities.

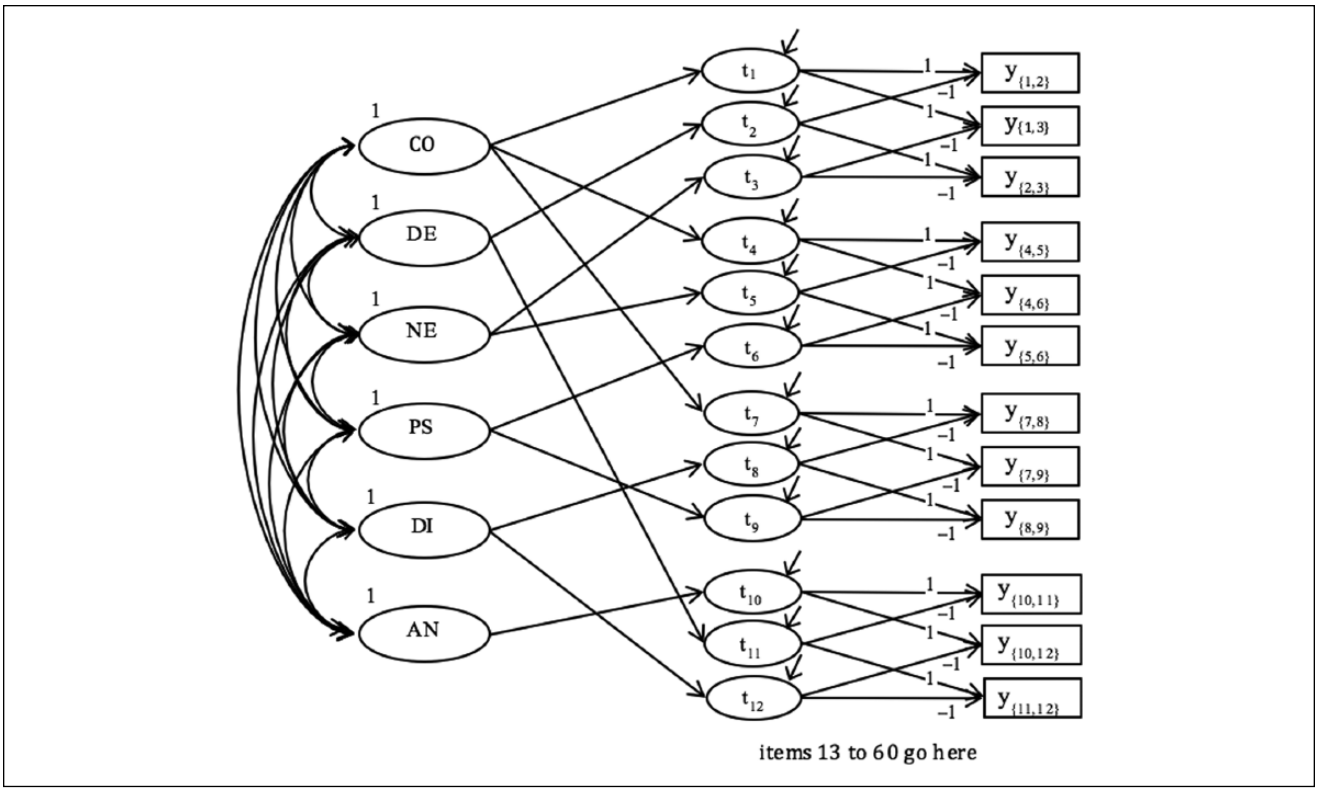

Figure 2. A fragment of the forced-choice measurement model for Goldsmiths 60-Item Questionnaire–Fixed Choice.
Note. Latent item utilizes are denoted $t_1 - t_2$; observed dummy contrasts (outcomes of pairwise comparisons) are denoted according to the items contrasted, for example *y*{1, 2} represents the contrast between Items 1 and 2.

Model Fit

Goodness of fit was assessed for both single-stimulus and forced-choice models. We considered the recommended goodness-of-fit indices (Hu & Bentler, 1998): the chi-square test (χ^2), the root mean square error of approximation (RMSEA), the comparative fit index (CFI), and the standardized mean root square residual (SRMR). A significant chi-square indicates rejection of the hypothesis that the model perfectly fits the data. For the other indices, it is customary to compare the obtained values with established cutoffs. For RMSEA, the cutoff .05 has been suggested for close fit, and .08 for adequate fit. For CFI, the cutoff .95 is considered for good fit and .90 for adequate fit. For SRMR, a value of .08 or less is considered indicative of an acceptable model. We draw the reader's attention to the fact that the above-mentioned indices, as well as the cutoffs have been established for structural equation modeling with continuous data. In the context of categorical data analysis with limited information estimators, however, the obtained fit indices evaluate how well the models reproduce the observed tetrachoric/polychoric correlations.

While we can evaluate the fit of the two models separately, comparing them with established cutoffs, we cannot compare our single-stimulus and forced-choice models formally since they are not based on the same observed variables and therefore are not nested. Hence, we rely on the fit indices as some measures of suitability of each model to the respective data.

Latent Factor Intercorrelations

When the best fitting model for each response format was identified, the single-stimulus and forced-choice responses were analyzed together by fitting the two measurement models in a single M*plus* run. This final step has the advantage of modeling the correlations between the error-free latent *DSM-5* domain traits, rather than the unreliable estimated factor scores.

Results

Model Estimation

The fit for the hypothesized single-stimulus measurement model was as follows: $\chi^2 = 4857.54$, degrees of freedom (*df*) = 1,695, RMSEA = .068 (90% CI [.066, .070], CFI = .70, SRMR = .107). Only the RMSEA indicated adequate fit to the data, while the other indices indicated poor fit. However, modifications were not justified, since the

cross-loadings were small, and highest loadings were always on intended factors. Table 2 reports the standardized factor loadings of the single-stimulus items on their respective traits.

In the first test of the forced-choice model, a case of empirical nonidentification was observed. This is not uncommon in forced-choice designs with all positively keyed items and positively correlated traits (see Brown & Maydeu-Olivares, 2012). In this case, estimated factor loadings within blocks may be very similar, yielding a degenerate factor space (Brown, 2014). To remedy this situation, additional constraints are required. Since the underidentification related to the relationships between the factors AN (Antagonism), DI (Disinhibition), and NE (Negative Affect), we fixed the correlations between these three factors to their estimated values in the single-stimulus model. The forced-choice model then successfully converged, and the fit for the G-60-FC was as follows: $\chi^2 = 2369.88$, $df = 1{,}638$,[1] RMSEA = .033 (90% CI [.030, .036]), CFI = .85, SRMR = .098. The RMSEA indicated good fit to the data, while the other indices indicated close to adequate fit. Judging by all the fit indices, the Thurstonian item response theory model of the forced-choice responses fitted to the a priori six-domain structure better than the counterpart model of the single-stimulus responses.

Item Parameters

Table 2 reports the standardized factor loadings for the single-stimulus items on their respective traits, as well as for the forced-choice latent utilities. These estimates are comparable across models since we standardize on the utilities underlying both the polytomous item responses in the single-stimulus format, and the binary pairwise outcomes. It can be seen that marker items (items with the highest standardized factor loading) for each trait differed between the two response formats. Often, the marker item in one design showed rather weak loading in the other, indicating that the traits as conceptualized in the two response formats had somewhat different meanings. The average standardized factor loading for the single-stimulus design was .53, and for the forced-choice design, it was .63.

Relationships Between Measured Constructs Within and Between Formats

Table 3 reports the monotrait heteromethod correlations between single-stimulus and forced-choice *DSM-5* personality traits (on the diagonal). These correlations were all positive, significant, and moderately strong. These latent correlations not attenuated for the measurement error, however, ranged between .62 for compulsivity and .86 for disinhibition, which falls well below the levels expected if the constructs measured by the two formats were equivalent. The average monotrait heteromethod correlation was .76, which was still much higher than the average single-stimulus monomethod heterotrait correlation of .41, and the average forced-choice monomethod heterotrait correlation of .42.

Table 3. Latent Factor Intercorrelations for Single-Stimulus and Forced-Choice Designs.

	AN	CO	DE	DI	NE	PS
AN	.64**	.26**	.63**	.55**	.39**	.50**
CO	.48**	.62**	-0.03	-.12*	.19**	0.09
DE	.43**	0.11	.85**	.61**	.73**	.55**
DI	.55(f)	-0.12	.45**	.86**	.50**	.66**
NE	.39(f)	.19*	.76**	.50(f)	.84**	.66**
PS	.57**	0.09	.52**	.78**	.59**	.76**

Note. The hetero-trait mono-method single-stimulus correlations are above the diagonal; the forced-choice correlations are below the diagonal. The mono-trait hetero-method correlations are on the diagonal; ** - correlation is significant at the 0.01 level; * - correlation is significant at the 0.05 level; (f) indicates that the correlation was fixed for identification in the forced-choice design. AN = Antagonism; CO = Compulsivity; DE = Detachment; DI = Disinhibition; NE = Negative Affect; PS = Psychoticism.

Examining the Performance of Fit Statistics Across Model Types

Readers may wonder if a possible explanation for the differences in fit might simply be that the fit indices perform differently across the two types of models. To examine this issue, we conducted a brief simulation study. We generated 60 item utilities with the same means, factor loadings, and factor correlations as was established in the single-stimulus model (treating data as continuous) and creating 100 samples of the same size as the empirical sample ($N = 420$). We then obtained Likert-type responses by categorizing the continuous 60 utilities so that the proportions in each category equaled to those in the empirical single-stimulus sample. To obtain forced-choice responses, we computed the differences of the continuous utilities for all the item pairs, and dichotomized them in accordance with the Thurstonian law of comparative judgment. Now we had two data set formats based on the same utilities, but yielding completely different observed variables to compare fit across the single-stimulus and forced-choice models.

Theoretically, both models should fit the simulated data near perfectly, with rejection rates at the nominal rate .05. The differences we observed in fit between the two model types from analyses of these data sets was trivial, with both models yielding insignificant chi-squares with near-nominal rejection rates. The RMSEA and CFI were different only at the third decimal place, and SRMR was actually better for the single-stimulus data than for the forced-choice data. Much more comprehensive simulations would be required to completely resolve the issue of how the fit

indices perform across different response formats, and these simulations fall outside the scope of this article. However, here we note that this small simulation calls into question interpretations of the findings in the current study as being due to spurious fit differences. Fit results from these supplementary analyses are available from the corresponding author.

Discussion

Industrial psychologists with an interest in personality assessment are paying attention to developments in clinical psychology. However, assessment of maladaptive traits in occupation settings brings the added complication that examinees are motivated to fake their responses. For this reason, industrial psychologists are also closely following developments in the field of psychological measurement for strategies for handling response biases. This article investigated an important confluence of all three fields: The participants were working adults studied in industrial psychology, the personality trait model was one that emerged from clinical psychology, and the method for reducing faking involved Thurstonian modeling of forced-choice responses from psychological measurement. The goal was to examine the quality of forced-choice measurement of maladaptive traits and, in particular, to compare the appropriateness of single-stimulus and forced-choice measurement designs combined, respectively, with confirmatory factor analysis and Thurstonian item response modeling.

Goodness of Fit

Goodness of fit of the hypothesized factorial structure was better for the forced-choice data than for the single-stimulus data. The closer fit of the Thurstonian model to the forced-choice data than for the corresponding independent clusters model to the single-stimulus data highlights the utility of comparative item formats in the context of maladaptive traits. Other research using a within-subject design, but measuring normal-range personality, similarly discovered that the forced-choice data fitted better (Brown & Maydeu-Olivares, 2011). However, the difference in fit between the single stimulus and forced choice was not as large as in this study, suggesting that the Thurstonian approach may have particular relevance for the assessment of maladaptive content. Although in the low-stakes assessments carried out in the present study, participants had little reason to dissimulate, other than preserving their self-image, the more extreme item wording than usually seen for normal-range personality questionnaires is likely to have evoked at last some SDR. Given that forced-choice modeling approaches have been demonstrated to reduce uniform response biases, the observation of better fit for the forced-choice models is consistent with the interpretation that the forced-choice model reduced SDR.

Construct Interpretation

We observed substantial differences in item discrimination parameters as estimated in the single-stimulus versus forced-choice models. This potentially alters the meaning of latent constructs. In some cases, such as negative affect, the loadings for the marker item in the single-stimulus design were similar to the loading for the marker item in the forced-choice design and vice versa, indicating quite similar trait meanings across the methods. In other cases, however, the loadings across the designs were quite different. For example, the marker item for antagonism under forced choice, "Fight fire with fire," appears quite consistent with the central elements of the broad definition of antagonism in Table 1, which emphasizes antipathy to others and an exaggerated sense of self-importance. The antagonism marker under the single-stimulus design, on the other hand, was "Am mean to people who deserve it." This is more representative of a hostility subcomponent that, while important to the antagonism construct, does not appear in the core definition.

Similarly, the marker item for disinhibition under the forced-choice design was "Like to do crazy things." This is representative of the broad elements of the disinhibition definition in Table 1, which relates to behavior driven by current desires rather than past learning or future considerations. On the other hand, the disinhibition marker for the single-stimulus design "Can't concentrate for long" appears to represent more a lack of attention. In sum, the loading patterns for the same items under single-stimulus and forced-choice designs and the monotrait heteromethod correlations show that different trait meanings emerge based on the response format.

Supporting this interpretation, and while similar latent traits were observed to underlie the forced-choice and the single-stimulus responses, the cross-method correlations for matching traits is far below the levels expected when equivalent conceptual constructs are measured (i.e., $r = 1$). However, it is still important to note that there is considerable similarity in the constructs, because the average monotrait heteromethod correlation (i.e., the average correlation between corresponding traits estimated from single-stimulus and forced-choice designs) was greater than the average of the heterotrait monomethod correlations.

Theoretical Contribution

The higher loadings under the single-stimulus design manifested in different variance–covariance matrices among the latent traits between the single-stimulus and forced-choice formats. In turn, different latent variances and covariance

for latent traits can be expected to lead to different relationships with external variables (Guenole & Brown, 2014). The observed relationships between maladaptive traits and external variables that are closest to a veridical representation of the nomology of maladaptive personality will be provided by whichever response format and measurement model combination yields the closest fit to the data. The current study suggests this representation is that provided by the forced-choice format and Thurstonian item response modeling.

Practical Contribution

From a practical perspective, the closer fit and higher loadings of the indicators in the forced-choice design indicate that when asked in the forced-choice format, the indicators are better markers of their theoretical traits. The scores for individuals resulting from Thurstonian item response modeling are therefore more suited as operational implementations of the theoretical *DSM-5* model than the single-stimulus format. For more accurate profiling of individuals for applications in organizations, we recommend test developers consider this approach to eliminate uniform response biases and more accurately model the response process. The current research suggests the forced-choice method, when modeled appropriately, is a viable assessment method of *DSM-5* content for occupational testing. It remains to be seen whether the reduction in biases is followed by increased predictive validity.

Limitations and Future Directions

The present study used a convenience sampling strategy to attract working individuals. Convenience samples of working adults can yield representative samples (Gosling, Vazire, Srivastava, & John, 2004) that are arguably appropriate for studying covariation among maladaptive item content (Wright et al., 2012). However, it is impossible to rule out that disruptive participants discerned the goals of the research and responded in a way that they believed would distort the results. It is therefore important that future research examines how the current research findings generalize to other samples such as actual job applicants. It will also be important to try to understand the implications of the differences between the model fit and person scores derived from Thurstonian and Likert-type scaling. This could occur, for instance, by examining the extent to which single-stimulus and Thurstonian item scores correlate with response bias scales in inventories such as the MMPI and California Psychological Inventory and stand-alone scales designed to measure response distortion such as the Marlowe–Crowne social desirability index (Crowne & Marlowe, 1960).

We also emphasize the importance of examining the respective predictive validities of the single-stimulus and Thurstonian trait scores against diverse criteria in the context of maladaptive trait assessment because the differences between the two approaches so far observed in this article pertain only to overall model fit and loading patterns. Finally, along with the psychometric challenges that maladaptive personality assessment in occupational contexts come social and legal challenges that are yet to be fully resolved. Future research should continue to investigate the legal and social implications of maladaptive personality assessment at work.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Note

1. When the number of items per forced-choice block is three or larger, the *df* must be corrected by subtracting the number of redundancies in the model (Brown & Maydeu-Olivares, 2012). There is one redundancy per block consisting of three items, making 20 redundancies in total for the G-60-FC.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Andrich, D. (1989). A probabilistic IRT model for unfolding preference data. *Applied Psychological Measurement*, *13*, 193-296. doi:10.1177/014662168901300211

Andrich, D. (1995). Hyperbolic cosine latent trait models for unfolding direct responses and pairwise preferences. *Applied Psychological Measurement*, *20*, 269-290. doi:10.1177/014662169501900306

Anguiano-Carrasco, C., MacCann, C., Geiger, M., Seybert, J. M., & Roberts, R. D. (2014). Development of a forced-choice measure of typical-performance emotional intelligence. *Journal of Psychoeducational Assessment*, *33*, 83-97. doi:10.1177/0734282914550387

Bartol, C. R. (1991). Predictive validation of the MMPI for small-town police officers who fail. *Professional Psychology*, *22*, 127-132.

Bartram, D. (2007). Increasing validity with forced-choice criterion measurement formats. *International Journal of Selection and Assessment*, *15*, 263-272. doi:10.1111/j.1468-2389.2007.00386.x

Block, J. (1961). *The Q-sort method in personality assessment and psychiatric research*. Springfield, IL: Thomas.

Böckenholt, U. (2012). Modeling multiple response processes in judgment and choice. *Psychological Methods*, *17*, 665-678. doi:10.1037/a0028111

Brown, A. (2014). Item response models for forced-choice questionnaires: A common framework. *Psychometrika*, *81*, 135-160. doi:10.1007/s11336-014-9434-9

Brown, A., & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement*, *71*, 460-502. doi:10.1177/0013164410375112

Brown, A., & Maydeu-Olivares, A. (2012). Fitting a Thurstonian IRT model to forced-choice data using M*plus*. *Behavior Research Methods*, *44*, 1135-1147. doi:10.3758/s13428-012-0217-x

Brown, A., & Maydeu-Olivares, A. (2013). How IRT can solve problems of ipsative data in forced-choice questionnaires. *Psychological Methods*, *18*, 36-52. doi:10.1037/a0030641

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *The Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration and scoring*. Minneapolis: University of Minnesota Press.

Carlson, E. N., Vazire, S., & Oltmanns, T. F. (2008). Self-other knowledge asymmetries in personality pathology. *Journal of Personality*, *81*, 155-170.

Clemans, W. V. (1966). *An analytical and empirical examination of some properties of ipsative measures* (Psychometric Monograph No. 14). Retrieved from https://www.psychometricsociety.org/sites/default/files/pdf/MN14.pdf

Colvin, C. R., Block, J., & Funder, D. C. (1995). Overly positive self-evaluations and personality: Negative implications for mental health. *Journal of Personality and Social Psychology*, *68*, 1152-1162.

Costello, R. M., Schneider, S. L., & Schoenfeld, L. S. (1996). Validation of a preemployment MMP index correlated with disciplinary suspension days of police officers. *Psychology, Crime & Law*, *2*, 299-306.

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

De Fruyt, F., Wille, B., & Furnham, A. (2013). Assessing aberrant personality in managerial coaching: Measurement issues and prevalence rates across employment sectors. *European Journal of Personality*, *27*, 555-564. doi:10.1002/per.1911

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, *102*, 652-670.

Furnham, A., Richards, S. C., & Paulhus, D. L. (2013). The dark triad of personality: A 10 year review. *Social & Personality Psychology Compass*, *7*, 199-216. doi:10.1111/spc3.12018

Gaddis, B. H., & Foster, J. L. (2015). Meta-analysis of dark side personality characteristics and critical work behaviors among leaders across the globe: Findings and implications for leadership development and executive coaching. *Applied Psychology*, *64*, 25-54. doi:10.1111/apps.12017

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, *4*, 26-42. doi:10.1037/1040-3590.4.1.26

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist*, *59*, 93-104. doi:10.1037/0003-066x.59.2.93

Gough, H. G., & Bradley, P. (1996). *CPI Manual.* Ed. 3. Palo Alto, CA: Consulting Psychologists.

Grijalva, E., Harms, P. D., Newman, D. A., Gaddis, B., & Fraley, R. C. (2015). Narcissism and leadership: A meta-analytic review of linear and nonlinear relationships. *Personnel Psychology*, *68*, 1-47. doi:10.1111/peps.12072

Guenole, N. (2014). Maladaptive personality at work: Exploring the darkness. *Industrial and Organizational Psychology*, *7*, 85-97. doi:10.1111/iops.12114

Guenole, N. (2015). The hierarchical structure of work related maladaptive personality traits. *European Journal of Psychological Assessment*, *31*, 83-90. doi:10.1027/1015-5759/a000209

Guenole, N., & Brown, A. (2014). The consequences of ignoring measurement invariance for path coefficients in structural equation models. *Frontiers in Psychology, 5*. doi:10.3389/fpsyg.2014.00980

Hamamura, T., Heine, S. J., & Paulhus, D. L. (2008). Cultural differences in response styles: The role of dialectical thinking. *Personality and Individual Differences*, *44*, 932-942.

Hargie, O., Stapleton, K., & Tourish, D. (2010). Interpretations of CEO public apologies for the banking crisis: Attributions of blame and avoidance of responsibility. *Organization*, *17*, 721-742. doi:10.1177/1350508410367840

Harms, P. D., Spain, S. M., & Hannah, S. T. (2011). Leader development and the dark side of personality. *Leadership Quarterly*, *22*, 495-509. doi:10.1016/j.leaqua.2011.04.007

Hess, M. F., & Broughton, E. (2014). Fostering an ethical organization from the bottom up and the outside in. *Business Horizons*, *4*, 541-549. doi:10.1016/j.bushor.2014.02.004

Hiatt, D., & Hargrave, G. E. (1988). MMPI profiles of problem peace officers. *Journal of Personality Assessment*, *52*, 722-731.

Hogan, R., & Hogan, J. (1997). *Hogan Development Survey Manual*. Tulsa, OK: Hogan Assessment Systems.

Hogan, R., & Hogan, J. (2001). Assessing leadership: A view from the dark side. *International Journal of Selection and Assessment*, *9*, 40-51.

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*, 424-453. doi:10.1037/1082-989X.3.4.424

Jin, K. Y., & Wang, W. C. (2014). Generalized IRT models for extreme response style. *Educational and Psychological Measurement*, *74*, 116-138.

John, O. P., & Naumann, L. P. (2010). Surviving two critiques by Block? The resilient Big Five have emerged as the paradigm for personality trait psychology. *Psychological Inquiry*, *21*, 44-49.

Joubert, T., Inceoglu, I., Bartram, D., Dowdeswell, K., & Lin, Y. (2015). A comparison of the psychometric properties of the forced choice and Likert scale versions of a personality instrument. *International Journal of Selection and Assessment*, *23*, 92-97. doi:10.1111/ijsa.12098

Judge, T. A., & LePine, J. A. (2007). The bright and dark sides of personality: Implications for personnel selection in individual and team contexts. In *Research companion to the dysfunctional workplace: Management challenges and symptoms* (pp. 332-355). Cheltenham, England: Edward Elgar.

Kaiser, R. B., LeBreton, J. M., & Hogan, J. (2015). The dark side of personality and extreme leader behavior. *Applied Psychology*, *64*, 55-92.

Kauder, B. S., & Thomas, J. C. (2003). Relationship between MMPI-2 and Inwald Personality Inventory (IPI) scores and ratings of police officer probationary performance. *Applied HRM Research*, *8*, 81-84.

Khorramdel, L., & von Davier, M. (2014). Measuring response styles across the Big Five: A multiscale extension of an approach using multinomial processing trees. *Multivariate Behavioral Research*, *49*, 161-177. doi:10.1080/00273171.2013.866536

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.

Kolar, D. W., Funder, D. C., & Colvin, C. R. (1996). Comparing the accuracy of personality judgments by the self and knowledgeable others. *Journal of Personality*, *64*, 311-338.

Krueger, R. F. (2013). Personality disorders are the vanguard of the post-*DSM-5.0* era. *Personality Disorders*, *4*, 355-362. doi:10.1037/per0000028

Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychological Medicine*, *42*, 1879-1890. doi:10.1017/S0033291711002674

Krueger, R. F., & Eaton, N. (2010). Personality traits and the classification of mental disorders: Toward a more complete integration in DSM 5 and an empirical model of psychopathology. *Personality Disorders*, *1*, 97-118. doi:10.1037/a0018990

Krueger, R. F., & Markon, K. E. (2013). The role of the DSM-5 personality trait model in moving toward a quantitative and empirically based approach to classifying personality and psychopathology. *Annual Review of Clinical Psychology*, *10*, 477-501. doi:10.1146/annurev-clinpsy-032813-153732

Krueger, R. F., Markon, K. E., Patrick, C. J., Benning, S. D., & Kramer, M. (2007). Linking antisocial behavior, substance use, and personality: An integrative quantitative model of the adult externalizing spectrum. *Journal of Abnormal Psychology*, *116*, 645-666.

Lee, J. W., Jones, P. S., Mineyama, Y., & Zhang, X. E. (2002). Cultural differences in responses to a Likert scale. *Research in Nursing & Health*, *25*, 295-306.

McCloy, R., Heggestad, E., & Reeve, C. (2005). A silk purse from the sow's ear: Retrieving normative information from multidimensional forced-choice items. *Organizational Research Methods*, *8*, 222-248. doi:10.1177/1094428105275374

McCrae, R. R., & Costa, P. T., Jr. (2010). *NEO Inventories: Professional manual*. Lutz, FL: Psychological Assessment Resources.

Maydeu-Olivares, A. (2001). Limited information estimation and testing of Thurstonian models for paired comparison data under multiple judgment sampling. *Psychometrika*, *66*, 209-228.

Maydeu-Olivares, A., & Bockenholt, U. (2005). Structural equation modeling of paired comparisons and ranking data. *Psychological Methods*, *10*, 285-304.

Maydeu-Olivares, A., & Joe, H. (2014). Assessing approximate fit in categorical data analysis. *Multivariate Behavioral Research*, *49*, 305-328. doi:10.1080/00273171.2014.911075

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.

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

Meade, A. W. (2004). Psychometric problems and issues involved with creating and using ipsative measures for selection. *Journal of Occupational and Organizational Psychology*, *77*, 531-552. doi:10.1348/0963179042596504

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*, 437-455.

Meehl, P. E. (1945). The dynamics of "structured" personality tests. *Journal of Clinical Psychology*, *1*, 296-303.

Messick, S. J. (1967). The psychology of acquiescence: An interpretation of research evidence. In I. A. Berg (Ed.), *Response set in personality assessment* (pp. 115-145). Chicago, IL: Aldine.

Moscoso, S., & Salgado, J. F. (2004). "Dark side" personality styles as predictors of task, contextual, and job performance. *International Journal of Selection and Assessment*, *12*, 356-362. doi:10.1111/j.0965-075X.2004.00290.x

Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Author.

O'Boyle, E. H., Jr., Forsyth, D. R., Banks, G. C., & McDaniel, M. A. (2012). A meta-analysis of the dark triad and work behavior: A social exchange perspective. *Journal of Applied Psychology*, *97*, 557-579. doi:10.1037/a0025679

O'Connor, B. (2002). A quantitative review of the comprehensiveness of the five-factor model in relation to popular personality inventories. *Assessment*, *9*, 188-203. doi:10.1177/1073191102092010

Padilla, A., Hogan, R., & Kaiser, R. B. (2007). The toxic triangle: Destructive leaders, susceptible followers, and conducive environments. *Leadership Quarterly*, *18*, 176-194. doi:10.1016/j.leaqua.2007.03.001

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, *46*, 598-609.

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. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 224-239). New York, NY: Guilford.

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

Plieninger, H., & Meiser, T. (2014). Validity of multi-process IRT models for separating content and response styles. *Educational and Psychological Measurement*, *74*, 875-899. doi:10.1177/0013164413514998

Rogers, R. (Ed.). (2008). *Clinical assessment of malingering and deception*. New York, NY: Guilford.

Salgado, J. F. (2003). Predicting job performance using FFM and non-FFM personality measures. *Journal of Occupational and Organizational Psychology*, *76*, 323-346.

Saulsman, L. M., & Page, A. C. (2004). The five-factor model and personality disorder empirical literature: A meta-analytic review. *Clinical Psychology Review*, *23*, 1055-1085. doi:10.1016/j.cpr.2008.07.002

Schyns, B. (2015). Dark personality in the workplace: Introduction to the special issue. *Applied Psychology*, *64*, 1-14. doi:10.1111/apps.12041

Serfass, D. G., & Sherman, R. A. (2013). A methodological note on ordered Q-Sort ratings. *Journal of Research in Personality*, *47*, 853-858.

Skodol, A. E., Bender, D. S., Morey, L. C., Clark, L. A., Oldham, J. M., Alarcon, R. D., . . . Siever, L. J. (2011). Personality disorder types proposed for DSM-5. *Journal of Personality Disorders*, *25*, 136-169. doi:10.1521/pedi.2011.25.2.136

Spain, S. M., Harms, P., & LeBreton, J. M. (2014). The dark side of personality at work. *Journal of Organizational Behavior*, *35*, 41-60. doi:10.1002/job.1894

Stark, S., Chernyshenko, O., & Drasgow, F. (2005). An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: The multi-unidimensional pairwise-preference model. *Applied Psychological Measurement*, *29*, 184-203. doi:10.1177/0146621604273988

Stark, S., Chernyshenko, O. S., Drasgow, F., & White, L. A. (2012). Adaptive testing with multidimensional pairwise preference items: Improving the efficiency of personality and other noncognitive assessments. *Organizational Research Methods*, *3*, 663-487. doi:10.1177/1094428112444611

Thomas, K. M., Yalch, M. M., Krueger, R. F., Wright, A. G., Markon, K. E., & Hopwood, C. J. (2012). The convergent structure of DSM-5 personality trait facets and five-factor model trait domains. *Assessment*, *3*, 308-311. doi:10.1177/1073191112457589

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 79*, 281-299.

Trull, T. J., & Durrett, C. (2005). Categorical and dimensional models of personality disorders. *Annual Review of Clinical Psychology*, *1*, 355-380. doi:10.1146/annurev.clinpsy.1.102803.144009

Vazire, S. (2010). Who knows what about a person? The self-other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, *98*, 281-300.

Vazire, S., & Mehl, M. R. (2008). Knowing me, knowing you: The accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *Journal of Personality and Social Psychology*, *95*, 1202-1216.

Widiger, T. A., & Simonsen, E. (2005). Alternative dimensional models of personality disorder: Finding a common ground. *Journal of Personality Disorders*, *19*, 110-130. doi:10.1521/pedi.19.2.110.62628

Wille, B., De Fruyt, F., & De Clercq, B. (2013). Expanding and reconceptualizing aberrant personality at work: Validity of five-factor model aberrant personality tendencies to predict career outcomes. *Personnel Psychology*, *66*, 173-223. doi:10.1111/peps.12016

Wright, A. G., Thomas, K. M., Hopwood, C. J., Markon, K. E., Pincus, A. L., & Krueger, R. F. (2012). The hierarchical structure of *DSM-5* pathological personality traits. *Journal of Abnormal Psychology*, *121*, 951-957. doi:10.1037/a0027669

Wu, J., & LeBreton, J. M. (2011). Reconsidering the dispositional basis of counterproductive work behavior: The role of aberrant personality. *Personnel Psychology*, *64*, 593-626. doi:10.1111/j.1744-6570.2011.01220.x

Zickar, M. J., & Drasgow, F. (1996). Detecting faking on a personality instrument using appropriateness measurement. *Applied Psychological Measurement*, *20*, 71-87. doi:10.1177/014662169602000107

Zickar, M. J., Gibby, R. E., & Robie, C. (2004). Uncovering faking samples in applicant, incumbent, and experimental data sets: An application of mixed-model item response theory. *Organizational Research Methods*, *7*, 168-190. doi:10.1177/1094428104263674

Zinnes, J. L., & Griggs, R. A. (1974). Probabilistic, multidimensional unfolding analysis. *Psychometrika*, *39*, 327-350.