Journal of Individual Differences

The Measurement Invariance of the Short Dark Triad: Implications for High- and Low-Stakes Contexts

Mihaela Grigoraș, Andreea Butucescu, Amalia Miulescu, Cristian Opariuc-Dan, and Dragoș Iliescu

Online First Publication, May 13, 2020. http://dx.doi.org/10.1027/1614-0001/a000322

Original Article

The Measurement Invariance of the Short Dark Triad

Implications for High- and Low-Stakes Contexts

Mihaela Grigoraș[1], Andreea Butucescu[1], Amalia Miulescu[1], Cristian Opariuc-Dan[2], and Dragoș Iliescu[1]

[1]Department of Psychology, University of Bucharest, Romania
[2]Department of Administration Science, "Ovidius" University of Constanta, Romania

Abstract: Given the fact that most of the dark personality measures are developed based on data collected in low-stake settings, the present study addresses the appropriateness of their use in high-stake contexts. Specifically, we examined item- and scale-level differential functioning of the Short Dark Triad (SD3; Paulhus & Jones, 2011) measure across testing contexts. The Short Dark Triad was administered to applicant (N = 457) and non-applicant (N = 592) samples. Item- and scale-level invariances were tested using an Item Response Theory (IRT)-based approach and a Structural Equation Modeling (SEM) approach, respectively. Results show that more than half of the SD3 items were flagged for Differential Item Functioning (DIF), and Exploratory Structural Equation Modeling (ESEM) results supported configural, but not metric invariance. Implications for theory and practice are discussed.

Keywords: measurement invariance, narcissism, psychopathy; machiavellianism, dark triad, selection

This paper addresses the increasing interest in the investigation of Dark Triad personality measurements and provides a better understanding of the appropriateness of such assessments in high-stake testing situations (e.g., employment selection, promotion) versus low-stake contexts (e.g., research, personnel development). Specifically, this study examines the item- and scale-level differential functioning of the Short Dark Triad (SD3; Paulhus & Jones, 2011), a popular measure of Dark Triad (DT) traits, across these two types of contexts.

Whereas several approaches on the dark side characteristics exist (e.g., aberrant, dysfunctional, maladaptive personality, etc.) the three constructs of machiavellianism, narcissism, and psychopathy known as the Dark Triad (Paulhus & Williams, 2002) have dominated the research in industrial-organizational psychology (Schyns, 2015; Spain, Harms, & LeBreton, 2013). Over the past 17 years, the measurement of the dark side of personality in work contexts has become of significant interest, the Dark Triad being linked to a range of work outcomes, such as job performance (O'Boyle, Forsyth, Banks, & McDaniel, 2012), leadership (Kaiser, LeBreton, & Hogan, 2015; Krasikova, Green, & LeBreton, 2013), counterproductive behavior (Scherer, Baysinger, Zolynsky, & LeBreton, 2013; Wu & LeBreton, 2011), team processes (Baysinger, Scherer, & LeBreton, 2014), and so forth. It would therefore appear that dark personality tests can add value to the hiring process and could be used by employers in order to make personnel selection decisions (LeBreton, Shiverdecker, & Grimaldi, 2017).

However, as with all assessment methods, especially with personality inventories, several points need to be covered before their usage in real-life decisions. Specifically, two important validity threats related to measurement have been recently raised in conjunction with personality questionnaires: differences among factor structures and differences in item functioning, between high- and low-stake contexts (Stark, Chernyshenko, Chan, Lee, & Drasgow, 2001; Ziegler & Buehner, 2009).

Both these issues derive from the singular fact that most personality questionnaires are developed based on voluntary samples that are collected under conditions that may be significantly different from those faced by a job applicant (Schmit & Ryan, 1993). Current practice is such that if a measure has been found to show adequate psychometric properties in a low-stake population (e.g., volunteers), it is also used in high-stake contexts (e.g., job applicants), without considering the fact that the instrument may not measure the same psychological constructs in both the volunteer and the job applicant group, and as a result scores may have different meanings in the two populations.

https://doi.org/10.1027/1614-0001/a000322

Factor Structure Invariance of Personality Items in High- Versus Low-Stake Contexts

Several theoretical explanations were advanced in the literature for why personality measurements may not function similarly across high- and low-stake contexts. All these explanations gravitate around the fact that the structure and functioning of personality instruments appear to depend on the schemas of test-takers (Schmit & Ryan, 1993). These schemas are manifested in test-taking situations, where test-takers are guided by an image of themselves that they wish to convey to the tester (R. Hogan & Hogan, 1992). In a low-stake context (e.g., research) respondents are placed in a situation in which they must compare themselves to other people *in general*. However, when respondents are taking a personality test as part of a personnel selection process, the circumstances are likely to differ: the image they wish to convey is focused on competence, and the comparison group includes individuals whom the respondents consider to be competent workers (Schmit & Ryan, 1993).

Empirical studies have also indicated that the structure of normal personality questionnaires differs under the two different conditions of high- versus low-stake contexts. For example, Schmit and Ryan (1993) found different factor structures in an applicant sample versus a non-applicant sample – a sixth factor ("ideal employee") was identified in the candidate sample, containing items mostly from the agreeableness and conscientiousness dimensions.

Vecchione, Alessandri, and Barbaranelli (2012) tested the invariance of factor structures across applicant and non-applicant samples and found evidence of configural (i.e., the number of loadings on each latent factor is similar in the two groups) and metric invariance (i.e., each item has the same or equivalent loading on the latent construct across groups), but not of scalar invariance (i.e., equivalence of item-level intercepts). More recently, Ion and Iliescu (2017) also showed that the measurement of five-factor model (FFM) personality traits is non-equivalent across high and low-stake test-taking contexts, having a potential important impact on the validity of decisions. Facets of agreeableness were involved in four out of seven causes of misfit, suggesting the existence of a common factor not specified in the model (Ion & Iliescu, 2017).

The Dark Triad traits have shown consistent associations with the FFM. Dark personalities are characterized by their antagonistic, dishonest, and egocentric (i.e., showing low agreeableness) interpersonal styles (Miller et al., 2010). People high on machiavellianism are better at controlling their behavior in high stakes situations in order to pursue a goal, they "do not act on temptation like psychopaths" (Furnham, Richards, & Paulhus, 2013, p. 208), a strategic-calculating orientation being considered one of the defining characteristics of machiavellianism (Jones & Paulhus, 2014). Although people high on machiavellianism are self-disciplined, status and achievement oriented, and engage in deliberate action, meta-analytic studies (Muris, Merckelbach, Otgaar, & Meijer, 2017; O'Boyle et al., 2015) report a negative association between machiavellianism and conscientiousness, possibly because these individuals fail to adhere to moral rules and values (a defining feature of conscientiousness). Among Dark Triad traits, psychopathy is considered the most malevolent trait, as it is characterized not only by very low levels of agreeableness and conscientiousness (Eysenck & Eysenck, 1985) but also by high levels of impulsivity and thrill-seeking (Furnham et al., 2013). For narcissism, high negative correlations were observed in relation with the modesty and straightforwardness facets of agreeableness, and positive high associations were recorded with the assertiveness facet of extraversion (Miller et al., 2010) and the achievement-striving and competence facets of conscientiousness (Miller et al., 2010). The picture emerging from the abovementioned studies is that from all the normal personality traits measured by the FFM, the SD3 probably captures best and most consistently the two traits of interpersonal antagonism (low agreeableness) and slacking (low conscientiousness). The meta-analytical results reported by O'Boyle et al. (2015), as well as those of Muris et al. (2017) also showed that agreeableness is negatively correlated with all three components of SD3, and that conscientiousness is negatively associated with machiavellianism and psychopathy. Given the fact that the sources of lack of invariance in FFM studies connected almost exclusively with these two personality traits, that is, agreeableness (see Ion & Iliescu, 2017; Schmit & Ryan, 1993) and conscientiousness (Schmit & Ryan, 1993; Vecchione et al., 2012), we believe that lack of invariance in the SD3 when comparing low-stake and high-stake conditions is also likely.

The measurement invariance in the SD3 or in its components has been investigated before, but not in conditions of low- versus high-stake testing. For example, Chiorri, Garofalo, and Velotti (2019) investigated the invariance of the Dirty Dozen scale and found it to be invariant across sex. Other studies investigated the factorial invariance of machiavellianism across gender (Du & Tang, 2005), the invariance of various narcissism inventories in cross-cultural settings (Wetzel et al., 2020; Żemojtel-Piotrowska et al., 2019), the invariance of measures of clinical narcissism across adolescents and young adults (Somma, Pincus, Fontana, Cianfanelli, & Fossati, 2019), the measurement invariance across sex and race of self- and parent-report scales assessing psychopathic characteristics in adolescents (Horan, Brown, Jones, & Aber, 2015).

Differential Functioning of Personality Items in High- Versus Low-Stake Contexts

Item-level (in)equivalence among various testing situations has also been explored in personality questionnaires. For example, O'Brien and LaHuis (2011) showed that 50 out of 90 items of the 16 Personality Questionnaire Select were flagged as displaying Differential Item Functioning (DIF) across different testing contexts. In a study on the effects of high-stake testing (and associated faking) on the functioning of items and scales of the 16 Personality Factors Questionnaire, Stark et al. (2001) also found the presence of differential item/test functioning across testing situations. These findings suggest that high-stake testing adversely affects the item functioning and thus the construct validity of personality scales, with an impact on selection and other high-stake decisions (Ion & Iliescu, 2017).

Although we have not been able to find any research that investigates the measurement invariance of Dark Triad measures across testing contexts (high- vs. low-stake), there are studies that show the differential functioning of Dark Triad items across gender and age groups (e.g., Kubarych et al., 2013; Webster & Jonason, 2013). Although this is not the objective of the current study, the item-level (in) equivalence across gender or age groups of some Dark Triad measures suggests that Dark Triad items may be non-equivalent across different groups or contexts and that differences in test results may not show actual differences in trait levels.

The Current Study

We purport that Dark Triad instruments, like any other personality measures, are likely to show some level of non-equivalence between low- and high-stake testing contexts. There are several measures of the Dark Triad constructs in the literature. We chose the SD3 (with its 28-items form) out of the measures that dominate research on the DT at work as it is probably the most feasible for organizational screening. It assesses all three DT constructs simultaneously with reasonable reliability – both internal consistency, with coefficients ranging from .70 to .80, and stability, with test-retest coefficients ranging from .77 to .84 (Paulhus & Jones, 2015) – and with a broader predictive power (Egan, Hughes, & Palmer, 2015; Jones & Paulhus, 2014) than another popular measure, the Dirty Dozen (Jonason & Webster, 2010; Maples, Lamkin, & Miller, 2014).

We were unable to identify any empirical study on the measurement invariance of SD3 across testing contexts. We therefore contribute to the literature by examining item- and scale-level differential functioning of SD3 in low- and high-stake contexts, using a mixed-method approach to measurement invariance (i.e., both Classical Test Theory and Item Response Theory). There is evidence that DT traits are associated with faking and impression management, especially in high-stake contexts. For example, Roulin and Bourdage (2017) show that all three DT traits predicted the use of deceptive impression management tactics in selection. Therefore, our investigation is guided by the following hypothesis:

> *Hypothesis 1 (H1)*: The Dark Triad will demonstrate significant differential functioning at both (H1a) item – and (H1b) scale-level, across high-and low-stake contexts.

Method

Participants

High-Stake Samples

A total of 457 job applicants form Sample 1; these participants stem from two different high-stake sub-samples. Sample 1a was composed of candidates for admission to training institutions of the Romanian Ministry of Internal Affairs (i.e., future officers in police, gendarmerie, and fire-fighting forces; N = 212); 76.4% of them were male, with an age range between 17 and 28 years (M = 20.45, SD = 2.25). Sample 1b consisted of current law enforcement employees (i.e., gendarmes; N = 245), participating in a selection process for a higher professional body; 71% of them were male, aged between 18 and 49 years (M = 30.97, SD = 7.30).

Low-Stake Samples

Two non-applicant samples were employed in our study and were merged into one single low-stake sample totaling N = 592 (Sample 2). Sample 2a is a general population sample (N = 373) where 49.6% of participants were male, aged between 19 and 75 years (M = 45.88, SD = 15.42). Sample 2b (N = 219) consisted of white-collar employees of a telecommunications company; participants were 50.7% male, with ages ranging from 21 to 53 years (M = 38.96, SD = 9.51).

Figures E1 and E2 of the Electronic Supplementary Material, ESM 1, display the age and gender characteristics for each sample.

Procedure

The high-stake sample was obtained in the context of two formal selection procedures organized by law enforcement units. The SD3 measure was completed along other personality and ability measures in a screen-out stage of the selection procedure. The non-applicant Sample 2a

completed the SD3 along with several other measures as part of the translation and adaptation process of a personality inventory, whereas Sample 2b completed the SD3 as part of a research project on predictors of counterproductive work behavior. Both these latter situations represent low-stake contexts, as the completion of measures was anonymous, and results were used only for research purposes.

Measures

Dark Triad

The Dark Triad was assessed with the Short Dark Triad (SD3; Paulhus & Jones, 2011), a short-form measure of machiavellianism (α = .78), subclinical psychopathy (α = .88), and subclinical narcissism (α = .73). Items were rated on a 5-point Likert scale (1 = *Strongly Disagree*; 5 = *Strongly Agree*). We note that this is an early version of the SD3 (Jones & Paulhus, 2014). The minor differences between the 2011 and 2014 versions reside in 3 small changes in wording and 1 exclusion, in items of the machiavellianism and psychopathy scales. In the 2011 version machiavellianism had 10 items as opposed to 9 items in the 2014 version; item 2 was changed from "Generally speaking, people won't work hard unless they have to" to "I like to use clever manipulation to get my way," item 9 was removed and item 10 was modified from "Most people deserve respect" to "Most people can be manipulated." In the psychopathy scale, item 8 was changed from: "I like to pick on losers" to "I enjoy having sex with people I hardly know."

Analytical Strategy

Invariance at the item and scale-level is usually tested via two well-established methods (Raju, Laffitte, & Byrne, 2002; Vandenberg & Lance, 2000). One method is based on structural equation modeling (SEM) and the other on item response theory (IRT). Each approach has certain advantages and they provide information that is complementary. SEM analyses provide unique information regarding the relationship between latent factors, whereas IRT analyses provide unique information on the equivalence of scale items. Therefore, both approaches should ideally be considered when investigating invariance.

Differential Item Functioning

Item invariance across contexts was tested using the lordif package (Choi, Gibbons, & Crane, 2011) of the R software (R Core Team, 2017). We used the graded response model (GRM) for polytomous items described by Samejima (1969) to estimate trait levels. In GRM, the item response depends on (a) a discrimination (slope) parameter indicating the strength of association between item and latent construct, and (b) location (severity) parameters – "item difficulty" in aptitude tests – that refers to where the item is situated in the construct score continuum. For readers familiar with IRT, the two-parameter-logistic (2PL) model is a special case of GRM for dichotomous items. IRT parameters were obtained using the R mirt package (Chalmers, 2012). To detect the presence of DIF, we formed hierarchical models for each item including additional explanatory variables. In Model 1, item responses were predicted only by the trait level, Model 2 added group membership, and Model 3 also added the interaction between these to predict item responses. We tested for uniform DIF (i.e., a constant effect across trait levels) by comparing Models 1 and 2, non-uniform DIF (i.e., a variable effect by trait level) by comparing Models 2 and 3, and an overall DIF effect by comparing Models 1 and 3. We used a statistical significance p value of .01 and relied on two effect size measures to detect meaningful DIF: first, McFadden's pseudo R^2 change > .02 (Cohen, 1988), or second, 10% change in β1 (Choi et al., 2011). We have used 10% change in β1 as a criterion to determine the presence of uniform DIF in an item, as it evaluates the magnitude of the relationship between item responses and overall ability level while controlling for demographics (Crane, van Belle, & Larson, 2004). The low-stake sample served as the reference group for all analyses.

Factorial Invariance

We examined scale-level measurement invariance using Exploratory Structural Equations Modeling (ESEM; Asparouhov & Muthén, 2009). We chose ESEM over traditional Confirmatory Factor Analysis (CFA) because it allows for more flexibility in model estimation addressing the strict assumptions of CFA (i.e., no cross-loadings of items or facets) that often resulted in poor fit and inflated factor correlations in investigations of personality data in general (e.g., Marsh et al., 2010) and specifically in motivated testing contexts (Lee, Mahoney, & Lee, 2017). At the same time, ESEM uses all typical CFA parameters (Asparouhov & Muthén, 2009). We conducted all ESEM analyses using Mplus7 (Muthén & Muthén, 1998–2012). To account for the non-normality of the data, we used a robust maximum likelihood (MLR) estimation method, which adjusts standard errors of parameter estimates, followed by an oblique Geomin rotation (the default in Mplus) with an ε value of .50.

We first fitted ESEM models for the high-stake and low-stake groups separately, prior to examining measurement equivalence. Then, we tested several levels of group invariance, conducting multiple-group ESEM to examine the configural, metric, and scalar invariance of the SD3 scales. Configural or weak invariance holds when the same

number of factors exists in the data from each group and items have the same pattern of zero and nonzero loadings in both groups. Next, metric or strong invariance implies that factor loadings are equal across groups. Strict, or scalar invariance implies that item intercepts are also equal across groups. We evaluated model fit based on an overall assessment of the following indices (Hu & Bentler, 1999): Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Squared Residuals (SRMR), as the chi-square (χ^2)-statistic is overly sensitive to sample size (Byrne, 1998). The interpretation of fit indices in ESEM relies on the same cut-off values as in CFA, with an acceptable model fit indicated by values higher than 0.90 for CFI and TLI and lower than 0.05 for RMSEA and SRMR (e.g., Asparouhov & Muthén, 2009). Model invariance was evaluated based on decreases in CFI, where changes in CFI (ΔCFI) of less than .01 would indicate invariance (Chen, 2007).

Comparison of factor intercorrelations for the ESEM solution was performed using an online calculator (Soper, 2018) that transforms correlation coefficient values into *z* scores.

Results

Descriptive statistics and Cronbach's alpha coefficients of SD3 scales are shown in Table E1 in ESM 1. Higher levels for Cronbach's α were obtained in the low-stake group (ranging from .80 to .83) than in the high-stake group (α ranging from .60 to .68). Also, higher means and standard deviations were found for the three dark traits in the low-stake group, especially for psychopathy.

Differential Item Functioning

Although the SD3 uses a 5-point Likert response scale, response categories were collapsed to less than 5 points for a total of 16 items due to low endorsement (due to having less than five observations per category). As a result, 12 items had four response categories and 4 items had three. Based on the pseudo R2 criterion, 14 out of 28 (50%) items of SD3 were flagged as displaying context-related DIF (see Table E2 in ESM 1): seven machiavellianism items (item 1, item 2, item 4, item 6, item 8, item 9, and item 10), five narcissism items (item 2, item 5, item 6, item 8 and item 9), two psychopathy items (item 2 and item 3). Using the DIF criterion of 10% beta change, two psychopathy items were flagged with DIF: item 1 and item 6. These results supported Hypothesis 1a, showing significant DIF of SD3 scales across high- and low-stake contexts.

Individual-level impact is examined by comparing DIF adjusted and unadjusted trait (theta) estimates for each SD3 scale (see Figures 1, 2, and 3). For all the three figures, the graph in the left (panel) shows a box plot of the difference scores (before and after accounting for DIF). The interquartile range (the shaded box) represents the middle 50% of the difference scores, whereas the median of the differences is the bolded line (around zero for all SD3 scales). The scatter plot on the right shows the difference scores plotted against initial scores (ignoring DIF) separately for each testing contexts. The solid line placed at 0.0 indicates no difference and the mean difference is indicated by the dotted line (which is close to zero for all SD3 scales). The high-stake group generally reports lower levels of the trait (theta) than the low-stake group. For the high-stake group, accounting for DIF in most cases led to slightly higher scores in the case of machiavellianism and psychopathy and to slightly lower scores for narcissism. Although these results seem to suggest that the impact of DIF on scores was minimal, as Choi et al. (2011) recommended to take into account both visual and model-based information, we also considered the IRT parameters of the SD3 items (see Tables E3 and E4 in ESM 1). For the high-stake sample, machiavellianism and psychopathy offer most information for traits levels above zero, while narcissism offers information for trait levels between −3 and 4 (Figure E3, ESM 1). We also draw attention to the fact that for psychopathy 8 out of 9 items have b1 values greater than zero (see Table E3, ESM 1), showing that the trait must be above zero for people to endorse these items. For the low-stake sample all three SD3 scales had a wider range of trait measurement, machiavellianism, and narcissism between 4 and 3, and psychopathy between −3 and 4 (Figure E4, ESM 1).

Factorial Invariance

The results of the ESEM analyses are presented in Table 1. Fit indexes were marginally acceptable, but improvable both in the low-stake (CFI = .84, RMSEA = .066) and in the high-stake (CFI = .87, RMSEA = .044) groups, when tested separately. As zero-order constraints on cross-loadings were not imposed in ESEM analyses, we specified the following correlated errors: items 4 and 2 for narcissism, items 8 and 9 for psychopathy, items 4 and 5 for psychopathy, items 7 and 10 for machiavellianism in the case of the low-stake group and item 3 for machiavellianism with item 5 for narcissism, as well as items 3 and 4 for machiavellianism for the high-stake group. The revised models exhibited an improved and marginally acceptable fit to the data for each group (CFI = .88, RMSEA = .058 in the low-stake group and CFI = .90, RMSEA = .039 in the high-stake group). These revised models (Figures E5

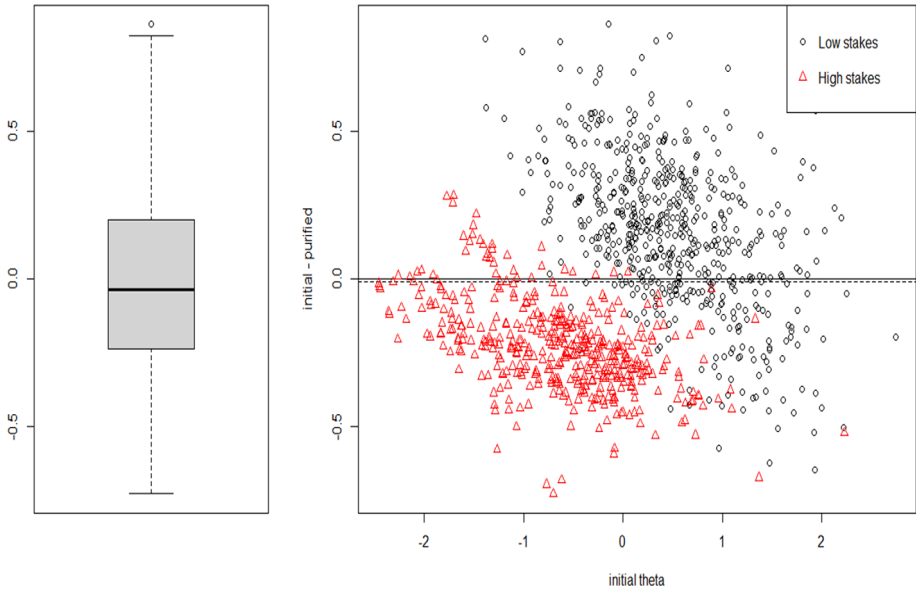

Figure 1. DIF impact at individual level for machiavellianism.

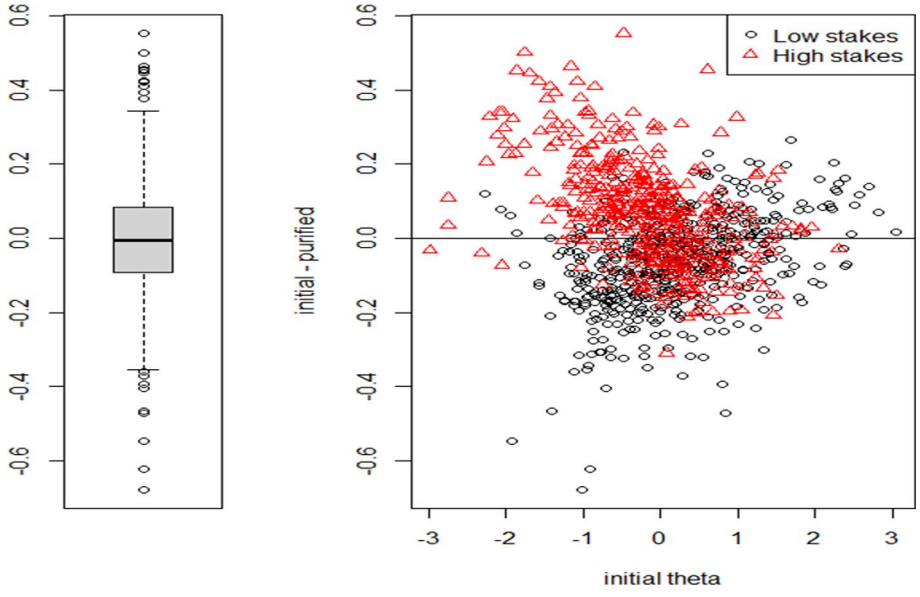

Figure 2. DIF impact at individual level for narcissism.

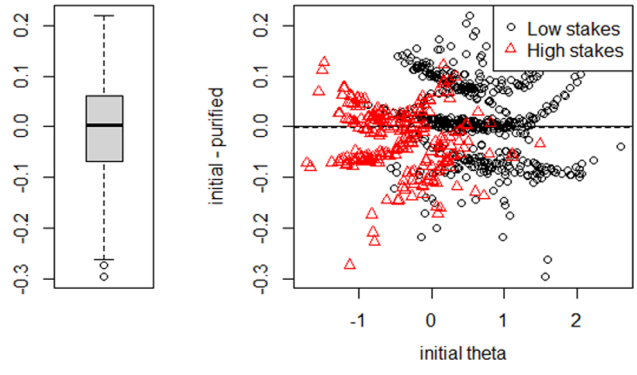

Figure 3. DIF impact at individual level for psychopathy.

and E6, ESM 1) were further considered baseline models for the low-stake and high-stake groups in the invariance analyses.

Tests of configural invariance yielded marginally acceptable values of goodness-of-fit (CFI = .88, RMSEA = .050). Constraining factor loadings to be equal across groups led to a change in CFI (ΔCFI = −.04) that exceeded the cut-off value of .01 suggested by Chen (2007). Thus, metric invariance was not supported. Given that metric invariance was not supported, as recommended by Putnick and Bornstein (2016), the test of scalar invariance was not performed. We concluded that the results of the scale-level analyses offered support for Hypothesis 1b.

Table 1. Fit statistics for the exploratory structural equation models

Model	χ^2	*df*	CFI	TLI	RMSEA	SRMR
High-stake (*N* = 457)	563.43	297	.87	.83	.044	.042
High-stake (revised model)	499.84	295	.90	.87	.039	.040
Low-stake (*N* = 592)	1,059.73	297	.84	.79	.066	.043
Low-stake (revised model)	866.57	293	.88	.84	.058	.039
Configural invariance	1,366.92	588	.88	.85	.050	.040
Weak	1,711.10	663	.84	.82	.055	.065

Note. CFI = Comparative Fit Index; TLI = Tucker–Lewis Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Squared Residuals.

Factor Loadings

The ESEM solutions for the high- and low-stake groups are presented in Table 2. The Geomin matrix showed a more straightforward solution for the low-stake sample. In the low-stake sample, with one exception (item 9 from machiavellianism), all items loaded appropriately (> .30) and had their highest loadings on the hypothesized factor.

Factor Intercorrelations

Table 3 presents the intercorrelations among the three factors across the low- and the high-stake samples. All factor correlations were significant and moderate in both groups, ranging from .17 to .45. Comparison of factor intercorrelations across testing contexts yielded a significant difference only for the narcissism and psychopathy association, which was larger in the low-stake sample ($z = 4.42$, $p < .01$), whereas the differences for narcissism-machiavellianism and psychopathy-machiavellianism were not significant ($z = -1.13$, $p = 0.25$ and $z = -1.36$, $p = 0.16$).

Both the IRT and ESEM analyses highlighted 4 problematic items. Two items had close to zero discrimination parameters and extreme beta scores. These items are: item 1 from the machiavellianism scale ("It's not wise to tell your secrets") ($a = 0.03$) and item 6 from the narcissism scale ("I feel embarrassed if someone compliments me") ($a = 0.04$). The other two items with very low discrimination parameters ($a = 0.37$ and $a = 0.23$) are: item 9 from narcissism ("I insist on getting the respect that I deserve") and item 2 from psychopathy ("I avoid dangerous situations"). These items also showed poor factor loadings on their respective factors (0.15, 0.17, 0.15, and 0.09, respectively).

Discussion

Summary of Findings

The present study examined the measurement invariance of the Short Dark Triad inventory in high- and low-stake contexts, using both item- and scale-level analyses. Our results show that the Short Dark Triad does not function equivalently across testing contexts, at neither the item nor the scale-level.

Item-level analyses showed that a sizeable number of items – more than 50% – are affected by DIF. This indicates that test takers with the same trait levels respond differently to more than half of the SD3 items, as a function of the testing situation. Of the three SD3 scales, machiavellianism is the most affected by DIF, with 7 out of the 10 items exhibiting DIF.

Consistent with previous research, in particular with research on situational as opposed to trait impression management (Henry & Raju, 2006), results show that DIF is not simply a matter of elevated scores in favor of one group. In the present study, whereas scales were uniformly biased in favor of the high-stake group, items were not uniformly biased in favor of one sample across all three scales, but manifested unpredictable relationships depending on the item, DIF was uniform, non-uniform, or combined, favoring either the high-stake or the low-stake sample.

As a high number of items (16 out of 28) had response categories that collapsed due to low endorsement, a legitimate question would be if these items have any relationship with items flagged for DIF. It is less likely that range restriction impacts our results. For example, the psychopathy scale, which has almost all items with collapsed response categories, has the lowest number of items flagged with DIF (see Table E2 in ESM 1).

Scale-level ESEM analyses also showed that the SD3 does not function equivalently across testing contexts: factor structures were not invariant in high- and low-stake samples. This finding is also in line with previous research regarding the effect of the testing context on the factor structure of non-cognitive measures (e.g., Ion & Iliescu, 2017; Ziegler & Buehner, 2009). Consistent with previous research (Ion & Iliescu, 2017; Lee et al., 2017; Schmit & Ryan, 1993) the high-stake sample exhibited a more complex pattern of factor loadings, presenting more items with primary loadings on non-parent scales and/or cross-loadings as compared to the low-stake sample.

Results show that there are no significant differences in factor intercorrelations across testing contexts, except for

Table 2. Geomin factor loadings (exploratory structural equation modeling)

	Machiavellianism		Narcissism		Psychopathy	
Items	High-stake	Low-stake	High-stake	Low-stake	High-stake	Low-stake
Machiavellianism						
1	.15	**.34**	.04	−.02	−.09	.05
2	**.43**	**.43**	−.06	.24	−.11	.03
3	**.46**	**.59**	.12	.13	−.04	−.05
4	**.37**	**.66**	−.03	.01	−.12	−.01
5	**.55**	**.79**	.03	−.15	.13	.03
6	**.49**	**.35**	.01	−.04	.25	.35
7	**.64**	**.54**	−.19	.11	.00	−.02
8	**.42**	**.29**	.01	.26	.00	.00
9	**.36**	.21	.09	.08	.27	.28
10	.19	**.42**	−.13	.07	.27	.14
Narcissism						
1	−.08	.10	**.64**	**.60**	.00	−.18
2	−.28	−.07	**.46**	**.43**	−.04	.22
3	.23	.06	**.40**	**.34**	.02	.19
4	.10	.04	**.60**	**.61**	.08	.07
5	.06	.01	**.31**	**.53**	−.07	.02
6	−.15	−.15	.17	**.65**	−.21	.07
7	.00	−.02	**.49**	**.75**	.10	−.11
8	.05	.13	.22	**.46**	.19	.08
9	.10	.12	.15	**.55**	−.04	−.03
Psychopathy						
1	.13	.21	−.04	.05	**.44**	**.47**
2	−.02	.14	.04	.11	.09	**.48**
3	−.03	.02	.02	−.11	**.74**	**.58**
4	−.01	−.20	−.10	.00	**.72**	**.79**
5	.19	.03	.00	.06	**.55**	**.42**
6	.17	.16	.01	−.03	**.43**	**.53**
7	.01	−.04	.07	.10	.14	**.63**
8	−.10	.04	.01	.00	**.71**	**.63**
9	−.08	−.01	−.02	.14	**.60**	**.48**

Note. Factor loadings higher than .30 are in boldface.

Table 3. Factor intercorrelations

Factor	1	2	3
Low-stake			
1. Machiavellianism	–	.42***	.45***
2. Narcissism		–	.42**
3. Psychopathy			–
High-stake			
1. Machiavellianism	–	.36**	.38***
2. Narcissism		–	.17**
3. Psychopathy			–

Note. $^{**}p < .01$; $^{***}p < .001$.

the narcissism-psychopathy correlation, which is larger in the high-stake sample. This finding is not consistent with previous results of increased correlations between traits caused by high-stake testing (e.g., Pauls & Crost, 2005). A possible explanation that has already been discussed in the literature (e.g., Ziegler & Buehner, 2009) is that, because high-stake testing (and associated faking) affects those scales that are considered relevant for the job more than those that are irrelevant, in heterogeneous samples, that is, samples containing applicants for different jobs (targeting different scales for impression management), as in our study, this general effect is then present in several or all scales, blurring the differences between them. Unfortunately, as no validity scale was employed in our study, we cannot robustly test this hypothesis.

Model fit for both the low- and the high-stake sample was improved by including several correlated errors into the model. These correlated errors belonged mainly to psychopathy (for the low-stake sample) and machiavellianism

(for the high-stake sample) and suggest a common underlying factor that has not been included into the model (possibly faking) and an insufficient differentiation of the items involved.

Theoretical and Practical Implications

The present study has important implications for personality research and for test development and use.

From a practical point of view we should consider the impact of the shown level of inequivalence on the practical decisions based on the SD3 – as noted, the SD3 has largely the same factor structure (configural invariance) across testing contexts but does not display metric invariance (similar loadings across groups). Given the large group differences detected for some items and scales, DIF may have an important consequence in practice. Unfortunately, this effect cannot be directly assessed as the study design did not include collection of external criteria. We believe that in light of our results it is especially important to use for decisions only normative data developed in high-stake contexts, as well as to establish cut-off scores based on high-stake populations; use of either normative or cut-off scores based on low-stake populations may result in large differences in pass-rates and may therefore decrease the utility of selection systems using the SD3 (Bott, O'Connell, Ramakrishnan, & Doverspike, 2007). We also pinpoint two important implications for the test development process. Firstly, using applicant (i.e., high-stakes) samples (and not only voluntary/low-stakes samples) during the test development process may be of higher importance than suggested by current practice. Secondly, the reported item-level non-invariance (DIF) suggests that the development of computer-adaptive measures in this domain may be more difficult. Also, regarding the four problematic items highlighted both by the IRT and ESEM analyses, we suggest that special attention should be devoted to these items in future research conducted on the SD3 scale and, if the results replicate, we recommend removal of these items for high-stake uses.

From a theoretical point of view, we feel compelled to speculate on the reasons of lack of measurement invariance for dark personality traits. The acknowledged existence of a certain level of item and factor non-equivalence between low- and high-stake contexts in personality questionnaires has been traditionally ascribed to the fact that high-stake testing creates more incentive for impression management. However, in the case of dark triad measures, such motivation may not only be induced by the context, but by the very nature of the assessed trait. Firstly, some of the items of the dark triad measures refer to highly undesirable behaviors (Dilchert, Ones, & Krueger, 2014; Watson, Stasik, Ro, & Clark, 2013), inducing an extra incentive to cheat. Secondly, the defining characteristics of a "dark triad personality" would include the willingness or even preference to manipulate, deceive, and lie, making these participants more skilled and more willing to engage in faking (Spain et al., 2013). We can further speculate that, of the three DT traits, machiavellianism is most affected by DIF due to its key characteristics of manipulativeness and strategic-calculating orientation (Jones & Paulhus, 2014). People high on machiavellianism are especially likely to engage in faking, as a selection context offers them a concrete incentive (i.e., a job offer) and the core of machiavellianism is to deceive in order to reach one's goals. As recent research shows, machiavellianism should be conceptualized as a less extreme form of psychopathy (Persson, 2019; Vize, Lynam, Collison, & Miller, 2018) and this could lead to higher endorsement rates for machiavellianism (Persson, Kajonius, & Garcia, 2017).

Strengths and Limitations

Several limitations of this study should be mentioned. Firstly, our study used a between-subject design in investigating invariance across testing contexts. Using within-subject data would rule out individual difference biases, other than the testing context, which could impact our conclusions. Secondly, the samples used are occupationally diverse and meta-analytic results show that the traits affected by the testing context are mainly those considered job relevant (Birkeland, Manson, Kisamore, Brannick, & Smith, 2006). Moreover, our high-stake sample is a law-enforcement sample, therefore the results cannot be generalized to other populations. The non-negligible higher proportion of men in the high-stake sample represents another limit of our study. While this is understandable in light of the sample being comprised of law-enforcement applicants, it also constitutes a potential weakness, in light of previous findings in the literature. For example, Webster and Jonason (2013) showed that men tend to have slightly lower thresholds than women in the endorsement of DT items, especially in psychopathy (about 0.5 *SD*). It is thus possible that the endorsement ratings reported by our study are slightly lower than a more balanced sample would have yielded. At the same time the second and more gender-balanced sample shows comparative results; Webster and Jonason (2013) also state that by and large there is "substantial overlap between men and women" (p. 305) in their DIF analysis – while bias in this respect cannot be completely ruled out, we believe it to be improbable. Also, the high-stake sample has a much younger mean age and age range than the low-stake sample. Further research should use within-subject data or at least employ gender and age balanced samples and focus on occupationally pure samples to rule out potentially individual differences

biases. Another limitation of our study is that no validity scale was included – we therefore cannot link faking to our results.

In spite of these limitations, this study has also a number of strengths: it is based on robust applicant and non-applicant samples and has a comprehensive approach, employing two complementary methods in the investigation of invariance: one based on structural equation modeling and the other on item response theory, providing both item- and scale-level information.

Electronic Supplementary Material

The electronic supplementary material is available with the online version of the article at https://doi.org/10.1027/1614-0001/a000322

ESM 1. The file contains figures and tables for descriptive statistics for the study variables for the high- and low-stake samples (Table E1), differential item functioning (DIF) results (Table E2), IRT parameters of the Short Dark Triad (SD3) items across low- and high-stake samples (Table E3), IRT slope-intercept parameters of the Short Dark Triad (SD3) items across low- and high-stake samples (Table E4), the gender and age distribution across samples (Figures E1, E2), the scale information curves for SD3 traits for the high- and low-stake sample (Figures E3, E4), and the revised exploratory structural equation model of SD3 for the high- and low-stake group (Figures E5, E6).

References

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling, 16*, 397–438. https://doi.org/10.1080/10705510903008204

Baysinger, M., Scherer, K. T., & LeBreton, J. M. (2014). Exploring the disruptive effects of psychopathy and aggression on group processes and group performance. *Journal of Applied Psychology, 99*, 48–65. https://doi.org/10.1037/a0034317

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*, 317–335. https://doi.org/10.1111/j.1468-2389.2006.00354.x

Bott, J. P., O'Connell, M. S., Ramakrishnan, M., & Doverspike, D. (2007). Practical limitations in making decisions regarding the distribution of applicant personality test scores based on incumbent data. *Journal of Business and Psychology, 22*, 123–134. https://doi.org/10.1007/s10869-007-9053-x

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts,applications, and programming*. Mahwah, NJ: Erlbaum.

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*, 1–29. https://doi.org/10.18637/jss.v048.i06

Chen, F. F. (2007). Sensitivity of goodness of fit indices to lack of measurement invariance. *Structural Equation Modeling, 14*, 464–504. https://doi.org/10.1080/10705510701301834

Chiorri, C., Garofalo, C., & Velotti, P. (2019). Does the dark triad manifest similarly in men and women? Measurement invariance of the Dirty Dozen across sex. *Current Psychology, 38*, 659–675. https://doi.org/10.1007/s12144-017-9641-5

Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software, 39*, 1–30. https://doi.org/10.18637/jss.v039.i08

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Crane, P. K., van Belle, G., & Larson, E. B. (2004). Test bias in a cognitive test: Differential item functioning in the CASI. *Statistics in Medicine, 23*, 241–256. https://doi.org/10.1002/sim.1713

Dilchert, S., Ones, D. S., & Krueger, R. F. (2014). Maladaptive personality constructs, measures, and work behaviors. *Industrial and Organizational Psychology, 7*, 98–110. https://doi.org/10.1111/iops.12115

Du, L., & Tang, T. L.-P. (2005). Measurement invariance across gender and major: The love of money among university students in People's Republic of China. *Journal of Business Ethics, 59*, 281–293. https://doi.org/10.1007/s10551-004-6395-4

Egan, V., Hughes, N., & Palmer, E. J. (2015). Moral disengagement, the dark triad, and unethical consumer attitudes. *Personality & Individual Differences, 76*, 123–128. https://doi.org/10.1016/j.paid.2014.11.054

Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences*. New York, NY: Plenum Press.

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

Henry, M. S., & Raju, N. S. (2006). The effects of trailed and situational impression management on a personality test: An empirical analysis. *Psychology Science, 48*, 247–267.

Hogan, R., & Hogan, J. (1992). *Hogan Personality Inventory Manual*. Tulsa, OK: Hogan Assessment Systems.

Horan, J. M., Brown, J. L., Jones, S. M., & Aber, J. L. (2015). Assessing invariance across sex and race/ethnicity in measures of youth psychopathic characteristics. *Psychological Assessment, 27*, 657–668. https://doi.org/10.1037/pas0000043

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional versus new alternatives. *Structural Equation Modeling, 6*, 1–55. https://doi.org/10.1080/10705519909540118

Ion, A., & Iliescu, D. (2017). The measurement equivalence of personality measures across high- and low-stake test taking settings. *Personality and Individual Differences, 110*, 1–6. https://doi.org/10.1016/j.paid.2017.01.008

Jones, D. N., & Paulhus, D. L. (2014). Introducing the short dark triad (SD3): A brief measure of dark personality traits. *Assessment, 21*, 28–41. https://doi.org/10.1177/1073191113514105

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

Kaiser, R. B., LeBreton, J. M., & Hogan, J. (2015). The dark side of personality and extreme leader behavior. *Applied Psychology: International Review, 64*, 55–92. https://doi.org/10.1111/apps. 12024

Krasikova, D., Green, S., & LeBreton, J. M. (2013). Destructive leadership: A theoretical review, integration, and future research agenda. *Journal of Management, 39*, 1308–13038. https://doi.org/10.1177/0149206312471388

Kubarych, T. S., Aggen, S. H., Kendler, K. S., Torgersen, S., Reichborn-Kjennerud, T., & Neale, M. C. (2013). Measurement non-invariance of DSM-IV narcissistic personality disorder criteria across age and sex in a population-based sample of Norwegian twins. *International Journal of Methods Psychiatric Research, 19*, 156–166. https://doi.org/10.1002/mpr.318

LeBreton, J. M., Shiverdecker, L. K., & Grimaldi, E. M. (2017). The dark triad and workplace behavior. *Annual Review of Organizational Psychology and Organizational Behavior, 5*, 387–414. https://doi.org/10.1146/annurev-orgpsych-032117-104451

Lee, P., Mahoney, K. T., & Lee, S. (2017). An application of the exploratory structural equation modeling framework to the study of personality faking. *Personality and Individual Differences, 119*, 220–226. https://doi.org/10.1016/j.paid.2017.07.029

Maples, J. L., Lamkin, J., & Miller, J. D. (2014). A test of two brief measures of the dark triad: The dirty dozen and short dark triad. *Psychological Assessment, 26*, 326–331. https://doi.org/10.1037/a0035084

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment, 22*, 471–491. https://doi.org/10.1037/a0019227

Miller, J. D., Dir, A., Gentile, B., Wilson, L., Pryor, L. R., & Campbell, W. K. (2010). Searching for a vulnerable dark triad: Comparing factor 2 psychopathy, vulnerable narcissism, and borderline personality disorder. *Journal of Personality, 78*, 1529–1564. https://doi.org/10.1111/j.1467-6494.2010.00660.x

Muris, P., Merckelbach, H., Otgaar, H., & Meijer, E. (2017). The malevolent side of human nature: A meta-analysis and critical review of the literature on the dark triad (narcissism, machiavellianism, and psychopathy). *Perspectives on Psychological Science, 12*, 183–204. https://doi.org/10.1177/1745691616666070

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide*. Los Angeles, CA: Author.

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

O'Boyle, E. H., Forsyth, D. R., Banks, G. C., Story, P. A., & White, C. D. (2015). A meta-analytic test of redundancy and relative importance of the dark triad and five-factor model of personality: Dark triad and the five-factor model. *Journal of Personality, 83*, 644–664. https://doi.org/10.1111/jopy.12126

O'Brien, E., & LaHuis, D. M. (2011). Do applicants and incumbents respond to personality items similarly? A comparison of dominance and ideal point response models. *International Journal of Selection and Assessment, 19*, 109–118. https://doi.org/10.1111/j.1468-2389.2011.00539.x

Paulhus, D. L., & Jones, D. N. (2015). Measures of dark personalities. In G. J. Boyle, D. H. Saklofske, & G. Matthews (Eds.), *Measures of personality and social psychological constructs* (pp. 562–594). San Diego, CA: Academic Press. https://doi.org/10.1016/B978-0-12-386915-9.00020-6

Paulhus, D. L., & Jones, D. N. (2011, January). *Introducing a short measure of the dark triad*. Poster presented at the meeting of the Society for Personality and Social Psychology, San Antonio, TX.

Paulhus, D. L., & Williams, K. M. (2002). *The dark side of normal personality: Self-report andbehavioral correlates* Unpublished manuscript, University of British Columbia.

Pauls, C. A., & Crost, N. W. (2005). Effects of different instructional sets on the construct validity of the NEO-PI-R. *Personality and Individual Differences, 39*, 297–308. https://doi.org/10.1016/j.paid.2005.01.003

Persson, B. N. (2019). Searching for Machiavelli but finding psychopathy and narcissism. *Personality Disorders: Theory, Research, and Treatment, 10*, 235–245. https://doi.org/10.1037/per0000323

Persson, B. N., Kajonius, P. J., & Garcia, D. (2017). Testing construct independence in the short dark triad using item response theory. *Personality and Individual Differences, 117*, 74–80. https://doi.org/10.1016/j.paid.2017.05.025

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *The Journal of Applied Psychology, 87*, 517–529. https://doi.org/10.1037/0021-9010.87.3.517

Roulin, N., & Bourdage, J. S. (2017). Once an impression manager, always an impression manager? Antecedents of honest and deceptive impression management use and variability across multiple job interviews. *Frontiers in Psychology, 8*, 29. https://doi.org/10.3389/fpsyg.2017.00029

Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores*. Richmond, VA: Psychometric Society. (Psychometric Monograph No. 17). Retrieved from http://www.psychometrika.org/journal/online/MN17.pdf

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

Schmit, M. J., & Ryan, A. M. (1993). The big five in personnel selection: Factor structure in applicant and nonapplicant populations. *Journal of Applied Psychology, 78*, 966–974. https://doi.org/10.1037/0021-9010.78.6.966

Scherer, K. T., Baysinger, M. J., Zolynsky, D., & LeBreton, J. M. (2013). Predicting counterproductive work behaviors with subclinical psychopathy: Beyond the five factor model of personality. *Personality & Individual Differences, 55*, 300–305. https://doi.org/10.1016/j.paid.2013.03.007

Somma, A., Pincus, A. L., Fontana, A., Cianfanelli, B., & Fossati, A. (2019). Measurement invariance of three versions of the pathological narcissism inventory across gender-matched Italian adolescent high school and young adult university students. *Journal of Psychopathology and Behavioral Assessment*. https://doi.org/10.1007/s10862-019-09758-7

Soper, D. S. (2018). *Significance of the difference between two correlations calculator* [Software]. Retrieved from https://www.danielsoper.com/statcalc/calculator.aspx?id=104

Spain, S. M., Harms, P., & LeBreton, J. M. (2013). The dark side of personality at work. *Journal of Organizational Behavior, 35*, 141–160. https://doi.org/10.1002/job.1894

Stark, S., Chernyshenko, O. S., Chan, K. Y., Lee, W. C., & Drasgow, F. (2001). Effects of the testing situation on item responding: Cause for concern. *The Journal of Applied Psychology, 86*, 943–953. https://doi.org/10.1037/0021-9010.86.5.943

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4–70. https://doi.org/10.1177/109442810031002

Vecchione, M., Alessandri, G., & Barbaranelli, C. (2012). The five factor model in personnel selection: Measurement equivalence between applicant and non-applicant groups. *Personality and Individual Differences, 52*, 503–508. https://doi.org/10.1016/j.paid.2011.11.014

Vize, C. E., Lynam, D. R., Collison, K. L., & Miller, J. D. (2018). Differences among dark triad components: A meta-analytic investigation. *Personality Disorders: Theory, Research, and Treatment, 9*, 101–111. https://doi.org/10.1037/per0000222

Watson, D., Stasik, S. M., Ro, E., & Clark, L. A. (2013). Integrating normal and pathological personality: Relating the DSM-5 trait-dimensional model to general traits of personality. *Assessment, 20*, 312–326. https://doi.org/10.1177/1073191113485810

Webster, G. D., & Jonason, P. K. (2013). Putting the "irt" in "dirty": Item response theory analyses of the dark triad dirty dozen – An efficient measure of narcissism, psychopathy, and Machiavellianism. *Personality and Individual Differences, 54*, 302–306. https://doi.org/10.1016/j.paid.2012.08.027

Wetzel, E., Lang, F. J., Back, M., Vecchione, M., Rogoza, R., & Roberts, B. (2020). *Measurement invariance of three narcissism questionnaires across the US, the UK, and Germany*. [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/rf6ep

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

Żemojtel-Piotrowska, M., Piotrowski, J., Rogoza, R., Baran, T., Hitokoto, H., & Maltby, J. (2019). Cross-cultural invariance of NPI-13: Entitlement as culturally specific, leadership and grandiosity as culturally universal. *International Journal of Psychology, 54*, 439–447. https://doi.org/10.1002/ijop.12487

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

History
Received November 20, 2018
Revision received January 30, 2020
Accepted March 6, 2020
Published online May 13, 2020

Authorship
The first 2 authors have contributed equally to this work.

ORCID
Dragoș Iliescu
https://orcid.org/0000-0002-5958-3920

Dragoș Iliescu
Department of Psychology
University of Bucharest
Panduri Street, no. 90
Bucharest
Romania
dragos.iliescu@fpse.unibuc.ro