A peer-reviewed version of this preprint was published in PeerJ on 1 March 2016.

View the peer-reviewed version (peerj.com/articles/1748), which is the preferred citable publication unless you specifically need to cite this preprint.

The (mis)measurement of the Dark Triad Dirty Dozen: Exploitation at the core of the scale

Petri Kajonius, Björn Persson, Patricia Rosenberg, Danilo Garcia

Background: The dark side of human character has been conceptualized in the Dark Triad Model: Machiavellianism, psychopathy, and narcissism. These three dark traits are often measured using single long instruments for each one of the traits. Nevertheless, there is a necessity of short and valid personality measures in psychological research. As an independent research group, we replicated the factor structure, convergent validity and item response for one of the most recent and widely used short measures to operationalize these malevolent traits, namely, Jonason’s Dark Triad Dirty Dozen. We aimed to expand the understanding of what the Dirty Dozen really captures because the mixed results on construct validity in previous research.

Method: We used the largest sample to date to respond to the Dirty Dozen (N = 3,698). We firstly investigated the Dirty Dozen’s factor structure using Confirmatory Factor Analysis. Secondly, using sub-sample (n = 500) and correlation analyses, we investigated the Dirty Dozen dark traits convergent validity to Machiavellianism measured by the Mach-IV, psychopathy measured by Eysenck’s Personality Questionnaire Revised, narcissisms using the Narcissism Personality Inventory, and both neuroticism and extraversion from the Eysenck’s questionnaire. Finally, besides these Classic Test Theory analyses, we analyzed the responses for each Dirty Dozen item using Item Response Theory (IRT).

Results: The results confirmed previous findings of a bi-factor model fit: one latent core dark trait, plus the three dark traits. An additional exploratory distribution analysis showed that all three Dirty Dozen traits had a striking bi-modal distribution, which might indicate unconcealed social undesirability with the items. The three Dirty Dozen traits did converge to, although not strongly, with the contiguous single Dark Triad scales (r between .41-.49). The probabilities of filling out steps on the Dirty Dozen narcissism-items were much higher than on the Dirty Dozen items for Machiavellianism and psychopathy. Overall, the Dirty Dozen instrument delivered the most predictive value with persons with average and high Dark Triad traits (Theta > -0.5). Moreover, the Dirty Dozen scale was better conceptualized as measured of a combined Machiavellianism-psychopathy factor, not narcissism, that can
be replaced with item 4: ‘I tend to exploit others towards my own end’.

**Conclusion:** The Dirty Dozen showed a consistent factor structure, a relatively convergent validity similar to that found in earlier studies. Narcissism measured using the Dirty Dozen, however, did not contribute with information to the core constitution of the Dirty Dozen construct. More importantly, the results imply a Single Item Dirty Dark Triad (SIDDT) measure of a manipulative and anti-social core as the content of the Dirty Dozen scale.
The (mis)measurement of the Dark Triad Dirty Dozen: Exploitation at the core of the scale

Petri J. Kajonius\textsuperscript{1,2,3}, Björn N. Persson\textsuperscript{3,4}, Patricia Rosenberg\textsuperscript{3}, Danilo Garcia\textsuperscript{1,3,5,6*}

\textsuperscript{1}Department of Psychology, University of Gothenburg, Sweden

\textsuperscript{2}Department of Social Psychology, University of Skövde, Sweden

\textsuperscript{3}Network for Empowerment and Well-Being, Gothenburg, Sweden

\textsuperscript{4}Department of Cognitive Neuroscience and Philosophy, University of Skövde, Sweden

\textsuperscript{5}Blekinge Center of Competence, Blekinge County Council, Karlskrona, Sweden

\textsuperscript{6}Centre for Ethics, Law, and Mental Health (CELAM), University of Gothenburg, Sweden
Abstract

**Background:** The dark side of human character has been conceptualized in the Dark Triad Model: Machiavellianism, psychopathy, and narcissism. These three dark traits are often measured using single long instruments for each one of the traits. Nevertheless, there is a necessity of short and valid personality measures in psychological research. As an independent research group, we replicated the factor structure, convergent validity and item response for one of the most recent and widely used short measures to operationalize these malevolent traits, namely, Jonason’s Dark Triad Dirty Dozen. We aimed to expand the understanding of what the Dirty Dozen really captures because the mixed results on construct validity in previous research.

**Method:** We used the largest sample to date to respond to the Dirty Dozen ($N = 3,698$). We firstly investigated the Dirty Dozen’s factor structure using Confirmatory Factor Analysis. Secondly, using sub-sample ($n = 500$) and correlation analyses, we investigated the Dirty Dozen dark traits convergent validity to Machiavellianism measured by the Mach-IV, psychopathy measured by Eysenck’s Personality Questionnaire Revised, narcissisms using the Narcissism Personality Inventory, and both neuroticism and extraversion from the Eysenck’s questionnaire. Finally, besides these Classic Test Theory analyses, we analyzed the responses for each Dirty Dozen item using Item Response Theory (IRT).

**Results:** The results confirmed previous findings of a bi-factor model fit: one latent core dark trait, plus the three dark traits. An additional exploratory distribution analysis showed that all three Dirty Dozen traits had a striking bi-modal distribution, which might indicate unconcealed social undesirability with the items. The three Dirty Dozen traits did converge to, although not strongly, with the contiguous single Dark Triad scales ($r$ between .41-.49). The probabilities of filling out steps on the Dirty Dozen narcissism-items were much higher than on the Dirty Dozen items for Machiavellianism and psychopathy. Overall, the Dirty Dozen instrument delivered the most predictive value with persons with average and high Dark Triad traits ($\Theta > -0.5$). Moreover, the Dirty Dozen scale was better conceptualized as measured of a combined Machiavellianism-psychopathy factor, not narcissism, that can be replaced with item 4: ‘I tend to exploit others towards my own end’.

**Conclusion:** The Dirty Dozen showed a consistent factor structure, a relatively convergent validity similar to that found in earlier studies. Narcissism measured using the Dirty Dozen, however, did not contribute with information to the core constitution of the Dirty Dozen construct. More importantly, the results imply a Single Item Dirty Dark Triad (SIDDT) measure of a manipulative and anti-social core as the content of the Dirty Dozen scale.

**Keywords:** Dark Triad; Dark Triad Dirty Dozen; Gender; Item Response Theory; Machiavellianism; Narcissism; Psychopathy; Single Item Dirty Dark Triad.
Over the last 25 years, the vast majority of personality research has focused on the Big Five traits: openness, conscientiousness, extraversion, agreeableness, and neuroticism. The Big Five Model of personality is a theory developed from both language taxonomy as well as statistical factor analysis (Costa & McCrae, 1992). However, the Big Five has painted an unnecessarily “light” conceptualization of human nature (Lee & Ashton, 2014). Critics have also argued against – what they believe is – the overreliance on factor analysis (i.e., one of the methods in Classical Test Theory) to uncover the latent structure of personality, without a well-grounded theoretical basis and substantial variation in methodology (e.g., Block, 1995; Gould, 1981). In the last decade, personality psychologists have turned their attention to the dark side of human character: Machiavellianism, psychopathy, and narcissism. Together, these traits are widely known as the Dark Triad model (Paulhus & Williams, 2002). The validation studies of Dark Triad measures have mostly been conducted using Classic Test Theory methods and in very few cases using Item Response Theory (IRT) methods. The Dark Triad embodies interpersonal, sub-clinical, and maladaptive personality traits in the general population (Paulhus & Williams, 2002), which are characterized by manipulativeness (i.e., Machiavellianism), impulsivity and antagonism (i.e., psychopathy), and the sense of entitlement (i.e., narcissism). The Dark Triad traits are associated with a value system of unconventional and antisocial morality (Furnham, Richards, & Paulhus, 2013; Kajonius, Persson, & Jonason, 2015). In essence, individuals with high levels on any of these dark traits appear to operate in selfish and
competitive ways with a common core: uncooperativeness (see Jones & Figueredo, 2013). Thus, whether the dark traits constitute a ternary model of unique traits or a unified uncooperative general factor with three closely related anti-social sub-traits is still an open question. In this context, validation studies using IRT methods might shed some light on what different measures of the Dark Triad actually measure.

As with most personality psychology research, the measurement of individuals’ tendencies on the dark traits is often conducted using self-report measures. Most of the time, this has been done using one instrument for each trait. These single instruments to measure the dark traits are often long and time demanding. For the trait of Machiavellianism, for example, researchers often use Christie and Geis’ Mach-IV (1970), which was originally based on statements from the Italian Niccolò Machiavelli’s books *The Prince* and *The Discourse* (see also Jones & Paulhus, 2009, who point out that the instrument also captures behaviors from the Chinese military general, strategist, and philosopher Sun Tzu’s book *The Art of War*; behaviors such as planning, building a reputation, and creating alliances). The trait of psychopathy is often measured with the Self-report Psychopathy Scale (Hare, 1985). This instrument was first used on prisoners and later on validated in non-criminal populations as well (Hare, 1985). Nevertheless, also the psychoticism scale in the hierarchical three-factor model proposed by Eysenck (e.g., Eysenck, Eysenck, & Barrett 1985) has been used as a measure of psychopathy or "Impulsive Unsocialized Sensation Seeking" (Zuckerman, Kuhlman, Thornquist, & Kiers, 1991. See also Zuckerman, 1989, 1991; Linton & Power, 2013; Garcia & Sikström, 2014). Finally, narcissism is often measured using the Narcissism Personality Inventory, which comprises 80 (long version) or 32 (short version) paired-items (Raskin & Hall, 1979). Nevertheless, shorter measures comprising all three traits in one single instrument have been created to facilitate data collection.
One such measure is the Dark Triad Dirty Dozen (Jonason & Webster, 2010). See Table 1 for the statements and the key word in each one of the statements in the Dirty Dozen scale. The Dirty Dozen comprises 12 items that in four studies were demonstrated to retain its core of disagreeableness when compared to 91 items from questionnaires that measured the dark traits separately (Jonason & Webster, 2010). This is a reduced item count by 87%. Subsequent studies with smaller samples have explored the thin line between efficiency and accuracy in this short scale. The findings suggest a bi-factor structural model with both a general latent Dark Triad construct and the three dark traits of Machiavellianism, psychopathy, and narcissism. In addition, the findings also show relatively good convergent validity with the Mach-IV ($r = .53$), Self-report Psychopathy Scale III ($r = .32$), and Narcissism Personality Inventory-40 ($r = .53$) (Jonason & Luévano, 2013). Further validations, using a sample of young undergraduates, were reported with the ubiquitous Big Five Inventory developed by Benet-Martínez and John (1998). The findings revealed an unstable core of conscientiousness for psychopathy and agreeableness for both Machiavellianism and psychopathy, with no clear, correlational relationships with extraversion for narcissism (Jonason, Kaufman, Webster, & Geher, 2013). Again, a bi-factor model (i.e., one general factor plus three specific factors) fit the data best. This suggests that each dark trait measured something unique (Jonason, Kaufman, Webster, & Geher, 2013), in addition to the common variance captured by the general factor. In addition to validations using Classical Test Theory, current research has moved to IRT for the validation of the Dirty Dozen scale.

Table 1 should be here

There is a large diversity of models that have been developed using IRT. For simplicity reasons, we refer to most of them using the global term IRT throughout the rest of the paper. IRT was first proposed in the field of psychometrics for the purpose of ability assessment. For
instance, all major educational tests are developed using this technique because it significantly improves measurement accuracy and reliability, and it provides significant reductions in assessment time and effort (for a review see An & Yung, 2014). In recent years, this technique has also been applied in health and clinical research (e.g., Hays, Morales, & Reise, 2000; Edelen & Reeve, 2007; Holman, Glas, & de Haan, 2003; Reise & Waller, 2009). Using IRT models researchers have found a slightly lower endorsement threshold of the dark traits for males compared to females. This has been interpreted as differences in social undesirability sensitivity, or true differences as proposed by mating-strategy theory (Webster & Jonason, 2013). The latest validation study among onsite UK undergraduates and online Crowdflower-workers\(^1\), however, found conflicting results using Mokken analysis, a non-parametric form of IRT (Carter, Campbell, Muncer, & Carter, 2015). While the expected three traits of Machiavellianism, psychopathy, and narcissism, emerged among female students’ scores; only two traits emerged among male students’ scores. These two traits were a combined Machiavellianism-psychopathy factor and a narcissism factor. In contrast, among the online workers, only one core construct of the Dirty Dozen appeared. These differences were not explained by invariance over sex and age. Hence, casting some uncertainty on the evasive constructs measured by the Dirty Dozen scale or suggesting some kind of mismeasurement of the triad by this specific scale.

The Present Study

The possibility to replicate findings is one of the parameters that distinguish science from non-science. In short, replication should be at ”the heart of science” (Schmidt, 2009). By use of conceptual replications we can potentially confirm which findings about human nature that can

\(^1\) "CrowdFlower is a data enrichment, data mining and crowdsourcing company based in the Mission District of San Francisco, California. The company’s software as a service platform allows users to access an online workforce of millions of people to clean, label and enrich data. CrowdFlower is typically used by data scientists at academic institutions, start-ups and large enterprises.” Retrieved from: https://en.wikipedia.org/wiki/CrowdFlower.
be generalized and thus increase predictive validity in our regular use of psychological measurements. As researchers we expect that replication studies are common and that the methodology is well developed, however, particularly in social sciences the contrary is true, demonstrating an overall replication rate of only 1.07% (Makel, Plucker, & Hegarty, 2012; see also Lucas & Donnellan, 2013, and the Registered Replication Reports initiative by the Association for Psychological Science, http://www.psychologicalscience.org/index.php/replication). A major problem in current validations is the small sample sizes and a general lack of power and precision. This “results in lower precision in parameter estimates and systematically inflated effect size estimates” (Lucas & Donnellan, 2013, p. 453). In addition, the current validation studies that have been published provide many statistically significant low-powered findings even within the same study, which “paradoxically provide less support for a phenomenon than papers that report some failures to reach statistical significance” (Lucas & Donnellan, 2013, p. 453; see also Francis, 2012; Schimmack, 2012). To the best of our knowledge, the present study provides the largest single sample used to this date ($N = 3,698$) to replicate some of the most common findings with regard to the Dirty Dozen scale. For instance, previous validation studies on the Dirty Dozen have had limited, or at least unclear, generalizability, often only including undergraduates or homogenous age cohorts. In addition, although we do believe in researchers’ capacity for objectivity, we see as an important venue that an independent research group that had no ties to the construction of the Dirty Dozen scale conducted the present replication study.

In sum, we present a replication of Confirmative Factor Analysis (CFA), Convergence Analyses, and IRT Analyses of the Dirty Dozen, which is one of the most recent popular
personality short scales used among personality psychologists to measure the dark traits (see a review by Furnham, Richards, Rangel, & Jones, 2014). We have also conducted new analyses. First, we investigated the original factor structure, which has shown varied results in previous studies, this time using a sizable, heterogeneous sample from all walks and ages of life. For instance, with the large sample at hand, there were sufficient respondents for an exploratory distribution analysis, which has not been reported before. Second, we further establish the validity of the traits measured using the Dirty Dozen by investigating convergence with known, contiguous single long scales of the dark traits: the Mach-IV, Eysenck’s Personality Questionnaire Revised, and the Narcissism Personality Inventory. Third, using IRT, we explored what the Dirty Dozen truly endeavors to measure. We contend that what makes the Dark Triad, measured by this specific scale, “dark” is not a uniform stable core but, instead, a challenging mix of malevolent and tradition-laden anti-social and uncooperative traits. If so, we might be able to clarify what the Dirty Dozen measures and uncover what scale-items might be responsible for the many interpretations of its core.

Method

Ethical statement

After consulting with the Network for Empowerment and Well-Being’s Review Board we arrived at the conclusion that the design of the present study (e.g., all participants’ data were anonymous and will not be used for commercial or other non-scientific purposes) required only informed consent from the participants.

Participants

The participant data was collected through Mechanical Turk (MTurk), which has demonstrated reliability and validity, providing a wider range of socio-economic backgrounds compared to
other samples (Casler, Bickel, & Hackett, 2013). This is particularly useful when it comes to research on values, such as the undesirability of the Dark Triad traits’ values (cf. Kajonius et al., 2015). All participants were informed that the survey was voluntary, anonymous, and that the participants could terminate the survey at any time. The MTurk workers received 50 cents (US-dollars) as compensation for participating and only residents of the US were allowed to accept participation. Two control questions were added to the survey, to control for automatic responses (e.g., “This is a control question, please answer “neither agree or disagree”). A total of 50 participants responded erroneously to one or both of the control questions, the final sample constituted 3,698 ($M_{age} = 33.5, SD = 11.8$). As expected, males ($N_{males} = 1,726$) scored higher on all Dark Triad traits than females ($N_{females} = 1,972$), as summarized in the descriptive Table 2. A subsample ($N = 500$) also answered to single long instruments of the Dark Triad, extraversion, and neuroticism.

**Measures**

The Dark Triad Dirty Dozen (Jonason & Webster, 2010) is a 12-item self-report questionnaire measurement of the three Dark Triad traits. Participants are asked to rate how much they agreed (1 = *Strongly disagree*; 7 = *Strongly agree*) with statements such as: “I tend to manipulate others to get my way” (Machiavellianism), “I tend to lack remorse” (psychopathy), and “I tend to want others to admire me” (narcissism). Items were averaged to create each dimension (Cronbach’s Alphas between .74 to .85; see Table 2 for Alphas for both males and females). We also constructed a composite score of the three dark traits by using the mean values from all of the items. For facilitating readability, all measures of the dark traits using the Dirty

---

2 We are well aware of the controversy concerning composite scores present in the literature (e.g., Glenn & Sellbom, 2015). While we’re inclined to agree with their arguments on a theoretical level, we’ve elected to use a composite score as it is a quick abbreviation of a general “dark personality”. Furthermore, we conducted an exploratory omega analysis (omega in R package *psych*, see also Revelle & Wilt, 2013), which yielded a $\omega_{hierarchical}$ coefficient of .72, which suggests that the Dirty Dozen is saturated by a general factor. Additionally, the correlation between the
Dozen are labeled as follows: DD Machiavellianism, DD psychopathy, and DD narcissism. High scores represent high degree in each of the dark traits or, in the case of the composite, a high degree of the Dark Triad Dirty Dozen core.

The Mach-IV (Christie & Geis, 1970) was used to also measure Machiavellianism. The Mach-IV consists of 20 items that reflect ways of thinking and opinions about people and different situations (e.g., “Never tell anyone the real reason you did something unless it is useful to do so”). Participants were requested to rate to what extent they agree with each statement on a 6-point Likert scale: 1 = *Strongly agree*, 6 = *Strongly disagree*. The Machiavellianism score was computed by summarizing the means across the 20 items, a high score representing high degree of Machiavellianism.

The short version of the Eysenck’s Personality Questionnaire Revised was used to measure extraversion (e.g., “Do you usually take the initiative in making new friends?”), neuroticism (e.g., “Do you ever feel ‘just miserable’ for no reason?”), and psychoticism (e.g., “Would you like other people to be afraid of you?”) (Eysenck, Eysenck, & Barrett, 1985). The Eysenck questionnaire consists of 12 items for each trait (forced binary answers: *Yes* or *No*). The score for each of the personality traits was computed as the sum of the 12 items, with *yes* responses coded as 1 and *no* responses coded as 0. Thus, a high score represents high degree in each of the three personality traits. As stated in the Introduction section, Eysenck’s psychoticism scale is better labeled as psychopathy (Zuckerman, 1989, 1991). Hence, for the rest of the paper we refer to the psychoticism scale as psychopathy.

---

composite score and an unrotated principal component was .994. These analyses are available from the corresponding author upon request.
The short version of the Narcissistic Personality Inventory was used to also measure narcissism (Ames, Rose & Anderson, 2006). The instrument consists of 16 pairs of items (one consistent and one inconsistent with narcissistic behavior in each pair) for what participants are instructed to choose, for each pair, one item that comes closest to describing their own feelings and beliefs about themselves. The narcissism score was computed as the sum of the 16 items, with narcissism-consistent responses (e.g., “I really like to be the center of attention”) coded as 1 and narcissism-inconsistent responses coded as 0 (e.g., “It makes me uncomfortable to be the center of attention”). Thus, a high score represents high degree of narcissism.

Statistical Analysis

As in earlier studies, there was a relatively large skewness in the psychopathy scores and kurtosis in the narcissism scores. This has, however, been shown to not have a negative effect on subsequent statistical analysis when the sample size reaches the thousands (Lumley, Diehr, Emerson, & Chen, 2002). First, using Classic Test Theory, we used CFA for testing two contending models, one with only the latent dark triad core branching into three dark traits, and second, a bi-factor model with the latent dark triad core connecting directly with all items, while the three dark traits connecting only to their respective items. We used Structural Equation Modeling (SEM) in the software Amos v.22 for these calculations. Second, using SPSS v. 22, we conducted convergent correlational analyses with the collected contiguous single dark traits scales and Extraversion and Neuroticism (i.e., Mach-IV, Eysenck’s Personality Questionnaire Revised, and the Narcissism Personality Inventory). Third, with the purpose of exploring the Dirty Dozen content, we utilized the much in-demand method of IRT using the R package MIRT version 1.10 (Chalmers, 2012) in R version 3.2.1 (R Core Team, 2015). This is a methodology for modeling how test items contribute to one latent, scalable trait. We used a graded response
model (analogous to the 2PL for dichotomous items) which basically generates two defining characteristics for each item: a slope coefficient, or discrimination parameter alpha ($a$), and a discrimination coefficient, or threshold parameter beta ($b$). The $a$ parameter shows how strongly an item relates to a given latent construct theta ($\theta$; which in this study the Dark Triad core as measured by the Dirty Dozen scale). The $a$ parameter can be analogized as a factor loading, whereas the threshold parameters $b_{1-6}$ relates to the level of the latent trait at which the next highest response category has at least 50% probability of being endorsed. For more information about IRT see Morizot, Ainsworth, and Reise (2007).

### Results

The first purpose was to replicate the original factor-structure of the Dirty Dozen using Classic Test Theory. Two CFA-models were tested. The first model, a hierarchical structure, with the Dark Triad core “above” the three dark traits, DD Machiavellianism ($\lambda = .75$), DD psychopathy ($\lambda = .86$), and DD narcissism ($\lambda = .56$), was not optimal ($\chi^2(40) = 1530.24$, $p < .001$) and with non-satisfactory fit indices as well (NFI = .92, CFI = .92, and RMSEA = .10). The second model tested was a bi-factor structure, which proved more successful ($\chi^2(28) = 360.19$, $p < .01$) with sufficient fit indices (NFI = .98, CFI = .98, and RMSEA = .05). The RMSEA of this specific model was slightly better than in previous studies (RMSEA = .07 in Jonason & Luévano, 2013; RMSEA = .06 in Jonason et al., 2013). Furthermore, our model showed that 3 out of the 4 items in the DD narcissism cluster had very weak relationships with the Dark Triad core and that the DD Machiavellianism-items demonstrated the strongest relationships. The full model with all items’ regression coefficients is reported in Figure 1.

Figure 1 should be here
In addition, with the large sample at hand, there were sufficient respondents for an exploratory distribution analysis, which as far as we know has not been reported before. Figure 2 depicts a strong bimodality of the distribution (peaks on both Likert-categories 1 and 5) found in all three dark traits. DD Psychopathy showed the largest overrepresentation in the lowest scale-category (Likert-category 1), followed by DD Machiavellianism and last DD narcissism. Females were overrepresented in the lowest scale-category (Likert-category 1) compared to males, $N_{\text{females}} = 280, N_{\text{males}} = 151$ (DD Machiavellianism), $N_{\text{females}} = 486, N_{\text{males}} = 280$ (DD psychopathy), and $N_{\text{females}} = 188, N_{\text{males}} = 114$ (DD narcissism). In small sample-studies, distributions such as these, might strongly affect statistical validity, as well as external validity, indicating strong social undesirability with the items.

The second purpose was to analyze convergent validity of the dark traits as measured with the Dirty Dozen scale. We were simply looking for the expected, conjoining relationships between the Dark Triad traits and the contiguous, single long scales of the dark traits and both neuroticism and extraversion. Table 3 summarizes the correlations found, which overall did show relatively weak (all $r < .50$) converging relationships (Machiavellianism $r = .49$; psychopathy $r = .41$, narcissism $r = .47$). That is, the three dark traits measured with Dirty Dozen showed that DD Machiavellianism and DD psychopathy showed similar correlations, while DD narcissism related less well with the corresponding scales measured using the single long scales. Additionally, the Dirty Dozen Dark Triad composite (Table 3, row 4) showed the smaller relationships with the dark traits measured using the single long scales (correlations between the Dirty Dozen dark traits and the Dark Triad core were between .75-.88, correlations between the dark traits measured with the single instruments and the Dark Triad core were between .31-.53).
In addition, there are some discrepancies between the correlations between neuroticism and extraversion and the dark traits depending on how the Dark Triad was measured. For example, while there was a weak significantly positive correlation between DD narcissism and neuroticism ($r = .20, \ p < .01$), there was no significant correlation to neuroticism when narcissism was measured using the Narcissism Personality Inventory ($r = -.07$). In contrast the relationship between narcissism and extraversion was almost twice as large when narcissism was measured using the single instrument ($r = .40, \ p < .01$) than when measured using the Dirty Dozen ($r = .25, \ p < .01$).

The third and last purpose was to extend the discussion on the construct of the Dark Triad as measured by the Dirty Dozen scale, using IRT. The results from both the CFA model 1 (NFI = .92) and 2 (NFI = .98) indicated adequate unidimensionality, which is the basic assumption for IRT. We ran a polytomous graded-response model on the 12 Dirty Dozen items, allowing items to load on a latent Dark Triad core. The Total Information Curve reported in Figure 3 shows that the core of the Dark Triad ($\theta$) was revealed in a maximized way only when a participant has close to average levels (-0.5) of this latent trait (see Figure 4 for the Information Curve for each one of the dark traits measured using the Dirty Dozen). Hence, the Dirty Dozen scale functions well for capturing average and higher levels of the core Dark Triad, but not the lower levels. This once again leads to the question what constitutes the dark core, or more specifically to what the Dirty Dozen scale actually measures. See Figure 5 in the supplementary material for Scale Information Curves for each of the 12 items of the Dirty Dozen.

Table 3 should be here

Figure 3 should be here

Figure 4 should be here
In Table 4 the items’ ability to differentiate ($a$ parameter) between people with similar levels of the same latent trait are ranked, starting with the item yielding the most information (item 4: Exploit). The $a$-parameter typically ranges from 0.5 to 2.0 in personality scales (Morizot, Ainsworth, & Reise, 2007). As can be seen, three of the four DD narcissism items (item 10: Attend, item 9: Admire, and item 11: Status) contributed least to differentiation between individuals. The difficulties ($b$) for each item are listed in rows, and reflect the threshold levels of the latent trait necessary to have at least 50% chance of endorsing the next scale-step (e.g., $b_1$ denotes answering Option 1 vs. 2, 3, 4, 5, 6, 7). The $b$-parameters are scaled on the same metric as the latent trait ($\theta$) and falls in the range of -3 to +3 SD, thus 0 is approximated to be of average difficulty (at the mid-point of the distribution). At the highest scale-step ($b_6$), all sub-factor items showed extreme difficulty, close to 3 SD. At the lower end ($b_1$), the items on DD Machiavellianism (items 1-4) and DD psychopathy (items 5-8), still showed much difficulty, close to average 0, while narcissism (items 9-12), showed much less difficulty, close to -2SD. These results imply that DD narcissism is not contributing as much information to the core constitution of the dark triad construct, since the probability of filling out steps on the DD narcissism-items are much higher than on the DD Machiavellianism and DD psychopathy-items (cf. first, the skewness in distributions in Figure 2 and second, that DD narcissism correlated the least with the contiguous scales in Table 3). In other words, when a respondent does fill out high numbers on scale-items on DD Machiavellianism and DD psychopathy, this rapidly predicts the latent level of participant’s core dark personality (e.g., item 1: Manipulate, $a = 2.73$ or item 6: Amoral, $a = 1.91$), but not for DD narcissism (e.g., item 10: Attend, $a = .91$). The separate item information curves are found in Figure 4.
The item with the highest $a$ (item 4: Exploit, $a = 3.33$), was of particular interest, due to its superior discriminatory ability compared to the others. We surmised that this item in itself would be able to capture the entire Dark Triad core, as measured by the Dirty Dozen scale. This exploitation-item correlated with the summed Dark Triad ($r = .77$), to the same degree that the three dark traits did, DD Machiavellianism ($r = .88$), DD psychopathy ($r = .75$), and DD narcissism ($r = .76$). When exchanging the summed Dark Triad for the single exploit-item, the internal reliability between the constructs was only marginally lowered: inter-item .61 to .51; Cronbach’s Alpha .85 to .80. Finally, comparing the single item with the convergence coefficients of the summed Dark Triad composite (Table 3), the single item performed as well or better as a substitute.

Table 4 should be here

Discussion

This was a replication study of the popular and much used Dirty Dozen, based on the largest and most diverse sample to date. All previous research results on the bi-factor structure and convergent validity were confirmed. Concerning the previous varying findings on one-factor, bi-factor, or three-factor solutions (cf. Carter et al., 2015), the large sample in the present study warrants to overall lean towards a bi-factor solution. However, a new (old) problem was brought to the surface with the reporting of a strong bi-modal distribution of all three sub-factors, Machiavellianism, psychopathy, and narcissism. This has not been much emphasized in previous publications and a contribution of the present study is to highlight the scope of this problem and admonish for large samples sizes when researching the Dark Triad, which is known to compensate for unwanted distribution skewness in statistical analyses (cf. Lumley et al., 2002).
In the wake of the Mokken analysis by Carter et al. (2015), one of the attempts of this paper was to further the discussion on what the Dark Triad trait consists of and what the Dirty Dozen seeks to measure. First, the distribution analyses (Figure 2) can be interpreted as an inertia to filling out Machiavellianism- and psychopathy-items, while narcissism-items showed normal, unskewed distribution, and consequently, not adding as much to the prediction of the core construct. A second clue to the latent core of the Dark Triad is the convergence analysis in Table 3 indicating that narcissism was the sub-factor that least correlated with the adjacent constructs of sub-factors to Mach-IV and EPQ, again showing that Machiavellianism-psychopathy is at the center of the construct. Nevertheless, although we used convergent analyses as in many other validation studies; recent research suggests that short scales should not be validated using these type of analysis (Olaru, Witthöft & Wilhelm, 2015). Third, the IRT-analysis showed that narcissism-items (e.g., need for admiration, attention, and status) had the least difficulty and the least discriminating power, not contributing to the total information on the latent dark trait. We conclude and submit for future research that the Dirty Dozen is a measurement consisting of a core found in Machiavellianism-psychopathy (e.g., manipulation, deceit, amorality, and callousness, with no remorse, as seen in Table 4).

Our proposal is that the Dark Triad, at least as measured by the Dirty Dozen, might be the product of a hasty grouping of two “difficult” sub-factors, Machiavellianism and psychopathy, together with one ‘easy’ sub-factor, narcissism. Being narcissistic is considered more normal in these days and times (Twenge, Campbell, & Freeman, 2012), not as undesirable to fill out in questionnaires, and does not add to the core of the construct. This grouping of three is unfortunate both from a social desirability- (method artifact) and a subclinical perspective (how to find people with real problems). If one wants to quickly find the core of the Dark Triad, a one-
item of “I exploit others” might be the proper one-item ultra-brief scale to use. A similar approach has been taken recently with narcissism, compressing the original 40-item scale into a Single Item Narcissism Scale (SINS), and demonstrating sufficient reliability and validity in initial studies (Konrath, Meier, & Bushman, 2014; van der Linden & Rosenthal, 2015).

**Future research and conclusions**

The varying social undesirability with all sub-factors measured by the Dirty Dozen should be further explored. It is not clear if the bi-modality of distribution is a reflection of this, or if it is a certain group of people with for instance very high Big Five-agreeableness, thus virtually hitting zero on all sub-factors of the Dirty Dozen. In other words, it is not clear if this indicates a genuine difference in the Dark Triad as a construct that is not instrument-specific. Additional studies have to be carried out using different methods and measures in order to assess whether or not such a difference is a method artifact or a real difference.

Another problem is that it is not apparent to what extent a short Dark Triad scale taps into and is confounded by clinical populations. In a large replication study such as the present, statistically 1-5% will be eligible for personality disorders. The results from IRT implicates that item-difficulties are sufficient on Machiavellianism and psychopathy to be able to distinguish problematic levels of the dark personality core, but not narcissism.

The conclusion on our part is that the Dirty Dozen has its advantages by being short, intuitive, and even fun, containing high face validity, but also has drawbacks by being highly differing in item-difficulties. The mismeasurement of the Dirty Dozen seems to be that it actually measures two constructs, narcissism and an anti-social trait. This specific conclusion is supported by the fact of what we choose to call a Single Item Dirty Dark Triad (SIDDT), the “exploit”
item. In situations of restrained research time and space, the SIDDT captures the essence of what the Dirty Dozen actually measures.

“Show me again, the power of the darkness, and I'll let nothing stand in our way. Show me, grandfather, and I will finish what you started.“

From Star Wars: The Force Awakens

References


Table 1

The Dark Triad Dirty Dozen Scale’ traits, item numbers, statements and the key word in each one of the statements.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Item No.</th>
<th>Statement</th>
<th>Key Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>MACHIAVELLIANISM</td>
<td>1</td>
<td>I tend to manipulate others to get my way.</td>
<td>Manipulate</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>I have used deceit or lied to get my way.</td>
<td>Deceit</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>I have use flattery to get my way.</td>
<td>Flatter</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>I tend to exploit others towards my own end.</td>
<td>Exploit</td>
</tr>
<tr>
<td>PSYCHOPATHY</td>
<td>5</td>
<td>I tend to lack remorse.</td>
<td>Remorse</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>I tend to be unconcerned with the morality of my actions.</td>
<td>Amoral</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>I tend to be callous or insensitive.</td>
<td>Callous</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>I tend to be cynical.</td>
<td>Cynical</td>
</tr>
<tr>
<td>NARCISSISM</td>
<td>9</td>
<td>I tend to want others to admire me.</td>
<td>Admire</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>I tend to want others to pay attention to me.</td>
<td>Attend</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>I tend to seek prestige or status.</td>
<td>Status</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>I tend to expect special favors from others.</td>
<td>Favors</td>
</tr>
</tbody>
</table>

Table 2

Descriptive Analysis of the Dark Triad traits as measured by the Dark Triad Dirty Dozen.

<table>
<thead>
<tr>
<th></th>
<th>$M$</th>
<th>$SD$</th>
<th>$\alpha$</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>$M_{\text{male}}$</th>
<th>$SD_{\text{male}}$</th>
<th>$\alpha_{\text{male}}$</th>
<th>$M_{\text{female}}$</th>
<th>$SD_{\text{female}}$</th>
<th>$\alpha_{\text{female}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 DD Machiavellianism</td>
<td>3.00</td>
<td>1.41</td>
<td>0.80</td>
<td>0.38</td>
<td>-0.66</td>
<td>3.23</td>
<td>1.45</td>
<td>.84</td>
<td>2.79</td>
<td>1.35</td>
<td>.85</td>
</tr>
<tr>
<td>2 DD Psychopathy</td>
<td>2.42</td>
<td>1.26</td>
<td>0.76</td>
<td>0.95</td>
<td>0.43</td>
<td>2.74</td>
<td>1.30</td>
<td>.81</td>
<td>2.13</td>
<td>1.14</td>
<td>.79</td>
</tr>
<tr>
<td>3 DD Narcissism</td>
<td>3.55</td>
<td>1.44</td>
<td>0.81</td>
<td>-0.15</td>
<td>-0.81</td>
<td>3.71</td>
<td>1.43</td>
<td>.75</td>
<td>3.41</td>
<td>1.44</td>
<td>.74</td>
</tr>
<tr>
<td>4 Dark Triad$^+$</td>
<td>2.99</td>
<td>1.08</td>
<td>0.85</td>
<td>0.26</td>
<td>-0.27</td>
<td>3.23</td>
<td>1.08</td>
<td>.80</td>
<td>2.78</td>
<td>1.04</td>
<td>.81</td>
</tr>
</tbody>
</table>

Note. $N = 3,698$; $N_{\text{males}} = 1726$; $N_{\text{females}} = 1972$; $^+$ composite score of the three dark traits.
**Table 3**

**Convergent analysis (Persons’ r) of the Dark Triad Dirty Dozen traits and the dark traits and Extraversion, and Neuroticism as measured by the Mach-IV, Eysenck’s Personality Questionnaire Revised, and the Neuroticism Personality Inventory.**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Item 4:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 DD Machiavellianism</td>
<td>—</td>
<td>.58</td>
<td>.50</td>
<td>.88</td>
<td>.49</td>
<td>.27</td>
<td>.37</td>
<td>.26</td>
<td>.10</td>
<td>.80</td>
</tr>
<tr>
<td>2 DD Psychopathy</td>
<td>—</td>
<td>.27</td>
<td>.75</td>
<td>.57</td>
<td>.41</td>
<td>.30</td>
<td>.25</td>
<td>-.11</td>
<td>.60</td>
<td></td>
</tr>
<tr>
<td>3 DD Narcissism</td>
<td>—</td>
<td>.76</td>
<td>.22</td>
<td>.09</td>
<td>.47</td>
<td>.20</td>
<td>.25</td>
<td>.43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Dark Triad +</td>
<td>—</td>
<td>.53</td>
<td>.31</td>
<td>.49</td>
<td>.30</td>
<td>.11</td>
<td>.77</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Machiavellianism</td>
<td>—</td>
<td>.40</td>
<td>.34</td>
<td>.28</td>
<td>-.06</td>
<td>.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Psychopathy</td>
<td>—</td>
<td>.35</td>
<td>.05</td>
<td>.05</td>
<td>.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 Narcissism</td>
<td>—</td>
<td>-.07</td>
<td>.40</td>
<td>.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Neuroticism</td>
<td>—</td>
<td>-.24</td>
<td>.18</td>
<td>.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 Extraversion</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Item 4: Exploit**

*Note. N = 500. All r coefficients > .12 are significant at p < .01. + Summarized composite score of the three dark traits.*

Yellow fields: intra-relationships within the Dark Triad Dirty Dozen traits and the Dark Triad composite.

Blue fields: intra-relationships within the Dark Triad traits measured by the single instruments (i.e., MACH-IV, Eysenck’s Personality Questionnaire Revised, and the Narcissism Personality Inventory).

Green fields: relationships between corresponding dark traits measured using the Dirty Dozen scale and the single instruments.

Black fields: relationships between dark traits as measured by the Dirty Dozen and Neuroticism and Extraversion.

Grey fields: relationships between dark traits as measured by the single instruments and Neuroticism and Extraversion.
Table 4

*Item Response Theory Analysis of the Dirty Dozen*

<table>
<thead>
<tr>
<th>Item</th>
<th>$a$</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$b_3$</th>
<th>$b_4$</th>
<th>$b_5$</th>
<th>$b_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 Exploit</td>
<td>3.33</td>
<td>-0.03</td>
<td>0.63</td>
<td>0.97</td>
<td>1.31</td>
<td>1.99</td>
<td>2.58</td>
</tr>
<tr>
<td>1 Manipulate</td>
<td>2.73</td>
<td>-0.24</td>
<td>0.42</td>
<td>0.75</td>
<td>0.94</td>
<td>1.78</td>
<td>2.54</td>
</tr>
<tr>
<td>2 Deceit</td>
<td>1.96</td>
<td>-0.71</td>
<td>-0.07</td>
<td>0.27</td>
<td>0.44</td>
<td>1.24</td>
<td>2.35</td>
</tr>
<tr>
<td>6 Amoral</td>
<td>1.91</td>
<td>0.36</td>
<td>1.11</td>
<td>1.56</td>
<td>1.87</td>
<td>2.44</td>
<td>3.14</td>
</tr>
<tr>
<td>5 Remorse</td>
<td>1.84</td>
<td>0.25</td>
<td>1.00</td>
<td>1.36</td>
<td>1.68</td>
<td>2.29</td>
<td>3.10</td>
</tr>
<tr>
<td>7 Callous</td>
<td>1.82</td>
<td>-0.03</td>
<td>0.75</td>
<td>1.15</td>
<td>1.45</td>
<td>2.18</td>
<td>3.05</td>
</tr>
<tr>
<td>12 Favors</td>
<td>1.71</td>
<td>-0.47</td>
<td>0.42</td>
<td>0.82</td>
<td>1.37</td>
<td>2.21</td>
<td>3.35</td>
</tr>
<tr>
<td>3 Flatter</td>
<td>1.41</td>
<td>-1.43</td>
<td>-0.79</td>
<td>-0.35</td>
<td>-0.07</td>
<td>1.00</td>
<td>2.46</td>
</tr>
<tr>
<td>11 Status</td>
<td>1.19</td>
<td>-1.23</td>
<td>-0.45</td>
<td>0.02</td>
<td>0.52</td>
<td>1.59</td>
<td>2.76</td>
</tr>
<tr>
<td>8 Cynical</td>
<td>1.07</td>
<td>-1.13</td>
<td>-0.41</td>
<td>0.05</td>
<td>0.43</td>
<td>1.41</td>
<td>2.77</td>
</tr>
<tr>
<td>9 Admire</td>
<td>1.02</td>
<td>-2.12</td>
<td>-1.38</td>
<td>-0.89</td>
<td>-0.18</td>
<td>1.07</td>
<td>2.83</td>
</tr>
<tr>
<td>10 Attend</td>
<td>0.91</td>
<td>-1.99</td>
<td>-1.12</td>
<td>-0.54</td>
<td>0.21</td>
<td>1.72</td>
<td>3.75</td>
</tr>
</tbody>
</table>

*Note.* Items are ranked according to item’s ability to discriminate ($a$) levels of the latent trait (the core of the Dark Triad) and are numbered according to their positions in the original questionnaire (DD Machiavellianism, 1-4, DD psychopathy, 5-8, and DD narcissism, 9-12). $b_{1-6}$ reports the item difficulties, reflecting the threshold level (-3 to +3 SD) of the latent trait necessary to have at least a 50% chance of endorsing the next scale-steps.
Figure 1. Bi-factor model of Dirty Dozen. $N = 3,698$. NFI = .98, CFI = .98, RMSEA = .05.
Figure 2. Frequency distributions showing the bi-modality of the three Dirty Dozen Dark Triad traits. N = 3,698. The numbers on the y-axis represent the proportion of replies for each Likert-category (1–7) on the x-axis. For instance, 47% of total replies on DD psychopathy items were placed on the lowest option (1) “strongly disagree”, which depicts the skewness in response pattern.
Figure 3. Total Information Curve on the latent core of the Dark Triad Dirty Dozen. Overall, the Dirty Dozen instrument delivers the most predictive value with persons with average and high Dark Triad traits (Theta > -0.5).
Figure 4. Scale Information Curves depicting the information content in each respective subfactor.
Figure 5. Item Information Curves for each of the 12 items of the Dirty Dozen. Note that item 1 (Manipulate) and 4 (Exploit) deliver the most information on the latent Dark Triad (labeled Theta). Item 8 (Cynicism) delivers the least information. Furthermore, the three first items on narcissism (9-11) don’t deliver much predictive information (flat curves, cf. earlier CFA bifactor model) to the overall Dark Triad trait – However, an IRT with only the narcissism items confirms that these predict the latent trait (Theta, i.e., narcissism) very satisfactorily (cf. earlier CFA hierarchical model).
SUPPLEMENTAL MATERIAL

The (mis)measurement of the Dark Triad Dirty Dozen: Exploitation at the core of the scale

Petri J. Kajonius\textsuperscript{1, 2, 3}, Björn N. Persson\textsuperscript{3, 4}, Patricia Rosenberg\textsuperscript{3}, Danilo Garcia\textsuperscript{1, 3, 5, 6*}

\textsuperscript{1}Department of Psychology, University of Gothenburg, Sweden

\textsuperscript{2}Department of Social Psychology, University of Skövde, Sweden

\textsuperscript{3}Network for Empowerment and Well-Being, Gothenburg, Sweden

\textsuperscript{4}Department of Cognitive Neuroscience and Philosophy, University of Skövde, Sweden

\textsuperscript{5}Blekinge Center of Competence, Blekinge County Council, Karlskrona, Sweden

\textsuperscript{6}Centre for Ethics, Law, and Mental Health (CELAM), University of Gothenburg, Sweden
Correspondence concerning this article should be addressed to D. Garcia, Network for Empowerment and Well-Being, Axel W. Andersson 8A, SE 371 62 Lyckeby, Sweden. E-mail: danilo.garcia@icloud.com.

Table S1

**Item Response Theory Rank and Exploratory Factor Analysis of Dirty Dozen Items.**

<table>
<thead>
<tr>
<th></th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Info</td>
<td>Machiavellianism</td>
</tr>
<tr>
<td>1 Manipulate*</td>
<td>2</td>
<td>.81</td>
</tr>
<tr>
<td>2 Deceit</td>
<td>3</td>
<td>.73</td>
</tr>
<tr>
<td>3 Flatter</td>
<td>8</td>
<td>.59</td>
</tr>
<tr>
<td>4 Exploit*</td>
<td>1</td>
<td>.78</td>
</tr>
<tr>
<td>5 Remorse</td>
<td>5</td>
<td>.44</td>
</tr>
<tr>
<td>6 Amoral</td>
<td>4</td>
<td>.46</td>
</tr>
<tr>
<td>7 Callous*</td>
<td>6</td>
<td>.47</td>
</tr>
<tr>
<td>8 Cynical*</td>
<td>11</td>
<td>.36</td>
</tr>
<tr>
<td>9 Admire</td>
<td>10</td>
<td>.35</td>
</tr>
<tr>
<td>10 Attend</td>
<td>12</td>
<td>.33</td>
</tr>
<tr>
<td>11 Status</td>
<td>9</td>
<td>.38</td>
</tr>
<tr>
<td>12 Favors*</td>
<td>7</td>
<td>.52</td>
</tr>
</tbody>
</table>

*Note. * = Potentially conflicting items in regards to double-loadings. Bold figures = Loadings according to three-factor Dark Triad theory. Cursive figures = Loadings deviating from three-factor Dark Triad theory (> .50). Item 8, underscored, “I tend to be cynical” does not load well enough on the psychopathy-factor and deviates from theory by aligning also with Machiavellianism (especially for females). Items 1, “I tend to manipulate others to get my way”,
4, “I tend to exploit others toward my own end”, and 12, “I tend to expect special favors from others” double-load on two separate factors. Items 4, 5, and 7 have somewhat varying loadings for males and females.