

# Current Psychology

## Psychometric properties of an Italian version of the Dirty Dozen: Factor structure, reliability, validity, measurement invariance across gender

--Manuscript Draft--

<b>Manuscript Number:</b>	
<b>Full Title:</b>	Psychometric properties of an Italian version of the Dirty Dozen: Factor structure, reliability, validity, measurement invariance across gender
<b>Article Type:</b>	Original Article
<b>Funding Information:</b>	
<b>Abstract:</b>	<p>The Dark Triad is a constellation of three socially undesirable personality traits: narcissism, psychopathy, and Machiavellianism. Previous research has shown that men tend to score higher than women on measures of the Dark Triad, but the validity of these results is unclear as no evidence of the measurement invariance of measures in the adult population has been provided. Here we present four studies that aimed: (a) to test the psychometric properties of an Italian version of a recently developed, concise measure of the Dark Triad: Jonason and Webster's Dirty Dozen (DD); (b) to examine the measurement invariance of the DD across gender. Results showed that the psychometric properties of the Italian DD overlap those of the original US version. We also found support for the measurement invariance of the DD over gender, with men scoring significantly higher than women on psychopathy and Machiavellianism and, to a lesser extent, on narcissism. Results were consistent with a biosocial approach to social role theory that assumes that being agentic and not being communal is considered desirable for men but not for women.</p>
<b>Corresponding Author:</b>	Carlo Chiorri Universita degli Studi di Genova Genova, ITALY
<b>Corresponding Author Secondary Information:</b>	
<b>Corresponding Author's Institution:</b>	Universita degli Studi di Genova
<b>Corresponding Author's Secondary Institution:</b>	
<b>First Author:</b>	Carlo Chiorri
<b>First Author Secondary Information:</b>	
<b>Order of Authors:</b>	Carlo Chiorri Patrizia Velotti Carlo Garofalo
<b>Order of Authors Secondary Information:</b>	
<b>Author Comments:</b>	<p>Genova, August 31, 2016</p> <p>Editor-in-Chief Matthias Ziegler Department of Psychology Humboldt University Berlin Rudower Chaussee 18 D-12480 Berlin Germany Tel. +49 30 2093-9447 Fax +49 30 2093-9361 zieglema(at)hu-berlin.deOpens window for sending email</p> <p>Dear Prof. Ferraro,</p>

I am enclosing a submission to Current Psychology entitled "Psychometric properties of an Italian version of the Dirty Dozen: Factor structure, reliability, validity, measurement invariance across gender". The manuscript is 27 pages long (excluding the title page) and includes four tables. The total word count is 7534. A file with supplementary information and analyses is also provided.

The paper addresses the issue of the measurement invariance across gender of a recently developed, concise measure of the Dark Triad: Jonason and Webster's Dirty Dozen (DD). As we acknowledge in the paper, the measurement invariance of the DD has already been investigated by Klimstra et al. (2014), but their results were of somehow limited generalizability, as they were obtained on samples of Dutch adolescents. We present here a study in which the measurement invariance of the DD across gender was tested in a large community adult sample. The results supported the measurement invariance of the DD over gender and showed that men scored significantly higher than women on psychopathy and Machiavellianism and, to a lesser extent, on narcissism.

As we performed the study on Italian participants, we first needed to adapt the DD into Italian. A thorough investigation of the psychometric properties of the Italian DD has been carried out and is reported in the paper. The results supported the overlapping of the psychometric properties of the Italian DD with the original.

My coauthors (Carlo Garofalo and Patrizia Velotti) and I wish for the manuscript to be given a masked review. However, we wish the paper not to be reviewed by Peter K. Jonason. We contacted him in March, 2012 to ask him about other ongoing Italian studies on the DD and he answered that there were none. He also supported us in the translation process. In 2013 he notified us that he was working at a similar project with another Italian group from the University of Enna and in 2015 he accepted to review a previous version of this manuscript for Personality and Individual Differences, apparently ignoring a clear conflict of interest - although he claimed he lost track of all these projects. Another Italian adaptation study of the DD is under revision at the Journal of Personality Assessment, and Jonason is among the authors. Notably, the other Italian colleagues (Adriano Schimmenti is the leading author) were informed by Jonason about our project only a few days ago, after we contacted him in July, 2015 to invite him to join in our paper.

None of the data from this paper were previously published, nor the manuscript is under review for publication in other journals. Please note that part of the results of the Italian adaptation of the DD were presented at a poster session at the XIX Italian Congress of Experimental Psychology, Rome, September 16-18, 2013.

My coauthors and I do not have any conflicts of interest that might be interpreted as influencing the research, and American Psychological Association ethical standards were followed in the conduct of the study. I will be serving as the corresponding author for this manuscript. All of the authors listed in the byline have agreed to the byline order and to submission of the manuscript in this form. I have assumed responsibility for keeping my coauthors informed of our progress through the editorial review process, the content of the reviews, and any revisions made. I understand that, if the manuscript is accepted for publication, the authors agree to automatic transfer of the copyright to the publisher; that the manuscript will not be published elsewhere in any language without the consent of the copyright holders; that written permission of the copyright holder is obtained by the authors for material used from other copyrighted sources; and that any costs associated with obtaining this permission are the authors' responsibility.

Sincerely,

Carlo Chiorri, PhD  
Department of Educational Sciences, University of Genova  
Corso Podestà, 2 - 16128 Genova (Italy)  
E-mail address: carlo.chiorri@unige.it  
Telephone +39 010 209 53709  
Fax +39 010 209 53728

**Suggested Reviewers:**

## Introduction

1  
2 In the last decade or so, three distinct, albeit overlapping, personality styles have been considered to  
3  
4 represent the "dark side" of human nature: psychopathy, narcissism, and Machiavellianism (Paulhus  
5  
6 & Williams, 2002). Among these personality traits, psychopathy is characterized by a constellation  
7  
8 of affective, interpersonal and behavioral features, including: egocentricity; impulsivity and thrill-  
9  
10 seeking; irresponsibility; shallow affectivity; lack of empathy, guilt, or remorse; pathological lying;  
11  
12 manipulateness; early, persistent, and versatile violation of social norms and expectations (Hare &  
13  
14 Neumann, 2008).  
15  
16  
17  
18

19  
20 Narcissism is defined by a pattern of grandiosity and inflated sense of self, entitlement,  
21  
22 dominance, exhibitionism and superiority (Morf & Rhodewalt, 2001). Narcissism also shares with  
23  
24 psychopathy a tendency to interpersonal exploitativeness and callousness. However, there is  
25  
26 substantial evidence that narcissism also encompasses more vulnerable features like fragile or  
27  
28 contingent self-esteem, emotion dysregulation, hypersensitivity to rejection and consequent social  
29  
30 avoidance (Cain, Pincus, & Ansell, 2008). The vulnerable side of narcissism also involves the  
31  
32 conscious experience of negative feelings like helplessness, emptiness, and shame (Velotti, Elison,  
33  
34 & Garofalo, 2014).  
35  
36  
37  
38

39  
40 Finally, Machiavellianism is defined as a duplicitous interpersonal style assumed to emerge  
41  
42 from a broader network of cynical beliefs and pragmatic morality (Jones & Paulhus, 2009).  
43  
44 Machiavellian individuals give higher priority to money, power, and competition than to  
45  
46 community building, self-love, and family concerns and use manipulative interpersonal strategies  
47  
48 such as flattery and lying to achieve their goals. Machiavellianism has been conceptualized as  
49  
50 consisting of four basic dispositions which are likely to lead to successful interpersonal  
51  
52 manipulation: affective detachment in interpersonal relationships; a lack of concern for  
53  
54 conventional morality; an intact reality contact (i.e., an absence of clear psychopathology); and low  
55  
56 ideological dedication (Jones & Paulhus, 2009).  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

Although these traits have been initially studied in separate fields (forensic, clinical, and social/organizational psychology, respectively), it has been highlighted the advantage of studying them simultaneously, coining the expression "Dark Triad" (DT; Paulhus & Williams, 2002). Ever since, there has been a progressive and ever-increasing growth in the amount of studies investigating psychopathic, narcissistic, and Machiavellian traits in the general population (i.e., at sub-clinical levels).

One of the most consistent findings on DT personality traits, regardless of which measures are used (DT measures or separate measures for each construct), pertains sex differences in scores on all three dimensions, with men usually scoring higher than women (Furnham & Trickey, 2011). This result has been consistently replicated for narcissism (Grijalva et al., 2015) and psychopathy scores (Cale & Lilienfeld, 2002) across different populations and with various assessment measures. Sex differences on Machiavellianism have instead been somehow inconsistent: However, even when no sex difference emerged on Machiavellianism, males reported overall higher levels of the composite DT score (Jonason & Webster, 2010; Klimstra, Sijtsema, Henrichs, & Cima, 2014).

One potentially problematic feature of these comparisons of DT scores across gender is that they have been carried out assuming that there is measurement invariance of the measures between women and men. Unless the underlying DT factors are measuring the same construct in the same way, and the measurements themselves are operating in the same way across gender, manifest mean comparisons are likely to be invalid (Millsap, 2011). Moreover, a statistically significant difference in latent means might no longer be statistically significant after correcting for measurement non-invariance.

To the best of our knowledge, the only paper that addressed this issue is a study by Klimstra et al. (2014). As a measure of DT they used the Dirty Dozen (DD, Jonason & Webster, 2010). Despite some limitations in its discriminant validity (Maples, Lamkin, & Miller, 2014; Miller et al., 2012), the DD have demonstrated a replicable 3-factor structure, adequate internal consistency, test-retest reliability and convergent validity (Jonason & Luévano, 2013; Webster & Jonason, 2013). So

1 far tests for sex differences on observed DD scores have always shown that men tend to score  
2 higher than women in all scales (Cohen's  $d$  ranging from 0.09 to 0.79), although this pattern seem to  
3 be more stable for psychopathy (see also Section 1 of the Electronic Supplementary Materials  
4 [ESM]).  
5  
6  
7

8  
9 Klimstra et al. (2014) found evidence of measurement invariance of the DD across gender in  
10 two samples of Dutch adolescents, and reported that boys tended to score consistently higher than  
11 girls on psychopathy, while somewhat less convincing evidence for sex differences in  
12 Machiavellianism and narcissism was found. Although this study shed some light on the  
13 measurement invariance of the DD over gender, its results are limited to an adolescent population.  
14 Hence, the issue is yet to be address on an adult population.  
15  
16  
17  
18  
19  
20  
21  
22

23  
24 Since when we started this research project no validated Italian version of the DD was  
25 available, we first needed to translate it into Italian and test its psychometric properties, including  
26 factor structure, internal consistency, temporal stability of scores and construct validity, in  
27 community and student samples. Then, we tested the measurement invariance of the DD across  
28 gender in a large community adult sample. We extended previous results using a larger taxonomy of  
29 invariance models that included models that also examine the invariance of residual variances,  
30 factor variances and factor correlations.  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40

## 41 **Study 1 – Translation and investigation of the factor structure of the Italian DD**

### 42 *Translation of the Dirty Dozen into Italian*

43  
44 The Italian translation of the Dirty Dozen was carried out through a mixed forward- and back-  
45 translation procedure. Before being used in this study, the newly developed Italian version of the  
46 DD was administered to ten naïve participants in order to check the clarity and readability of the  
47 items, which were all found to be easy to understand and score. The Italian DD (DD-I) is reported  
48 in Section 2 of the ESM.  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

### *Participants and procedure*

1  
2 The Italian DD (DD-I) was administered to three independent samples of community participants.  
3  
4 These participants were recruited through snowball sampling by psychology students as part of their  
5  
6  
7 dissertation or research training project. Sample 1 included 102 participants (mean age  $40.04 \pm 14.45$   
8  
9  
10 years, range 18-69, Females 53%), Sample 2 included 128 participants (mean age  $35.75 \pm 14.96$   
11  
12 years, range 18-80, Females 57%), Sample 3 included 305 participants (mean age  $37.34 \pm 13.30$   
13  
14 years, range 18-74, Females 61%). All participants volunteered to participate after being presented  
15  
16  
17 with a detailed description of the procedure, and all were treated in accordance with the *Ethical*  
18  
19  
20 *Principles of Psychologists and Code of Conduct* (American Psychological Association, 2010). In  
21  
22 order to be included in the study, participants had to be at least 18 years old and report never to  
23  
24  
25 have been diagnosed with a psychiatric disorder. They did not receive any compensation for their  
26  
27 participation. Administration of the DD-I took place at the premises of a psychology department in  
28  
29 North-Western Italy.  
30

### *Data analysis*

31  
32  
33  
34 In testing the factor structure of the DD-I using data from Samples 1 and 2 we thus chose to use an  
35  
36  
37 *exploratory* approach, rather than a confirmatory one. CFA requires each indicator to load on only  
38  
39 one factor, but, as shown by recent studies (Asparouhov & Muthén, 2009), this assumption might  
40  
41  
42 be too restrictive for personality research, because indicators may have secondary loadings  
43  
44 significantly different from zero. The presence of these secondary loadings is a critical issue: It  
45  
46  
47 would imply that the item(s) have a weak discriminant validity, since an item that is considered an  
48  
49 indicator of a specific construct can also be an indicator of another construct. In a CFA, the more  
50  
51  
52 the secondary loadings depart from zero, the more the correlations among the factors will be  
53  
54 inflated to account for non-zero secondary loadings restricted to zero, thus yielding: biased  
55  
56  
57 loadings, overestimated factor correlations, distorted structural relations, and lack of fit  
58  
59 (Asparouhov & Muthén, 2009). In their studies Jonason and Webster (2010) found some evidence  
60  
61 of substantial (i.e., larger than  $|\cdot30|$ ) cross-loadings in the DD (see their Table 2, p. 423).  
62  
63  
64  
65

1 The factor structure of the DD-I was first tested in the two smaller samples using Maximum  
2 Likelihood Exploratory Factor Analysis (ML-EFA) with the *fa* function in the R package *psych*  
3 (Revelle, 2015). An oblique Promax rotation was applied. According to de Winter, Doudou and  
4 Wieringa (2009), our sample sizes were adequate: for a 3-factor structure, 12 items and factor  
5 loadings in the .60s, they recommend a minimum of 67 participants (de Winter et al., 2009, p. 155).  
6  
7  
8  
9  
10

11 The optimal number of factors to extract was investigated through Parallel Analysis (PA,  
12 Horn, 1965) and Minimum Average Partial Correlation Statistic (MAP, Velicer, 1976). Analyses  
13 were performed with the *psych* (Revelle, 2015) package in R.  
14  
15  
16  
17  
18

19 In order to test the similarity of the factor solutions in the two samples, we computed  
20 congruence coefficients (CCs; Tucker, 1951). In principle, we could have carried out a  
21 measurement invariance analysis in a structural equation modeling (SEM) framework (see Marsh et  
22 al. 2010), but one requirement for these methods is sufficient statistical power to adequately  
23 estimate all the model parameters and, above all, their standard errors (Muthén & Muthén, 2002).  
24 Grounding on results of preliminary factor analyses, we performed a Monte Carlo analysis as  
25 described in Muthén and Muthén (2002), and found that the criteria suggested by these authors to  
26 achieve a power of .80 power could be met only with at least 300 participants per group (Section 3  
27 of the ESM)  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40

41 CCs are a measure of factor similarity advised when data do not meet the requirements of  
42 SEM (Lorenzo-Seva & ten Berge 2006). CCs can be interpreted as a standardized measure of  
43 proportionality of elements in factor loading matrices of different samples, and they measure factor  
44 similarity independently of the mean absolute size of the loadings. CCs range from  $-1$  to  $1$ , with  
45 values in the range .85–.94 suggesting adequate similarity and values higher than .95 suggesting  
46 substantial equality of factor loading matrices (Lorenzo-Seva & ten Berge 2006). Although CCs are  
47 not a thorough test of measurement invariance as it would be a set of nested SEM models with  
48 different degrees of invariance, they can provide evidence of at least configural invariance  
49 (similarity of the overall pattern of parameters) across samples.  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 Since Sample 3 afforded sufficient statistical power, we then performed confirmatory factor  
2 analyses (CFA) on data from this sample. As more parsimonious alternatives to the 3-correlated-  
3 factor model, we tested the fit of a 1-factor model and a 3-independent-factor models. Moreover, we  
4 tested the fit of bifactor model, that outperformed the other measurement models for the DD items  
5 in a recent study (Jonason & Luévano, 2013). In this model items loaded on two types of latent  
6 factors: a latent "general", Dark Triad factor and three latent factors associated with the Dark Triad  
7 traits. For model identification purposes, all latent factors are left uncorrelated. Diagrams of these  
8 models are reported in Section 4 of the ESM.  
9

10 CFA was performed with Mplus 7 (Muthén & Muthén, 1998–2012). We used the Mplus  
11 robust maximum likelihood estimator (MLR), with standard errors and tests of fit that were robust  
12 in relation to the nonnormality of observations. The goodness of fit of the CFA models was  
13 evaluated considering the comparative fit index (CFI), the Tucker-Lewis index (TLI), and the root-  
14 mean-square error of approximation (RMSEA), as operationalized in Mplus in association with the  
15 MLR estimator. Following Marsh, Hau, and Wen (2004), we considered values  $\geq .90$  as acceptable  
16 and  $\geq .95$  as optimal for TLI and CFI, and values  $\leq .08$  as acceptable and  $\leq .06$  as optimal for  
17 RMSEA.  
18

### 19 *Results and discussion*

20 Item descriptive statistics showed that items had moderate levels of positive skewness (Sample 1:  
21 Median = 1.04, range: 0.26-1.51; Sample 2: Median = 0.67, range: 0.03-1.81) and negative kurtosis  
22 (Sample 1: Median = -0.19, range: -1.17-1.46; Sample 2: Median = -0.45, range: -0.95-2.92). More  
23 details are reported in Section 5 of the ESM.  
24

25 Dimensionality analyses showed convincing evidence of the adequacy of a three-factor  
26 structure, as in both samples the scree-plot began to level off after the third factor, only the first  
27 three observed eigenvalues were higher than the simulated ones and the MAP reached its minimum  
28 at three components (Section 6 of the ESM). Hence, when performing EFAs, we set to three the  
29 number of factors to extract.  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65



The 3-factor solution account for 55% of variance in Sample 1 and 62% of variance in Sample 2. Factor loadings and factor correlations are reported in Table 1. All items substantially loaded on the expected factor (Sample 1: Median target loading: .71, range .48-.96; Sample 2: Median target loading: .79, range .51-.94), with minimal cross-loadings (Sample 1: Median cross-loading: .01, range -.27-.33; Sample 2: Median cross-loading: .04, range -.19-.26). Congruence coefficients for the three factors were all .95.

[Table 1]

Factor correlations of Machiavellianism with the other factors was in the .50s, while the correlation of narcissism with psychopathy was somehow lower (.40 in Sample 1 and .29 in Sample 2), consistent with previous studies (Jonason & Webster, 2010). The coefficients in the two correlation matrices did not significantly differ ( $X^2(6) = 3.24, p = .222$ ), suggesting that the pattern of association of factor scores was stable. Taken together, these results seemed to support the replicability of the expected 3-factor structure for the DD-I.

Results of the CFAs are reported in Table 2. Results suggested that the best fitting model was the bifactor model, but factor loading estimates of items 1, 4 and 5 were not statistically significant. The 3-correlated-factor model was second best-fitting model. Parameter estimates of factor loadings and factor correlations were all statistically significant (Section 7 of the ESM). The size of the factor correlations was consistent with results on samples 1 and 2.

[Table 2]

We then performed an item analysis on observed scores. We computed Cronbach's alphas, mean inter-item correlations, corrected-item total correlations, items' squared multiple correlations and alpha-if-item-deleted indices for all DD-I scales in all samples. Detailed results are reported in Section 8 of the ESM and show that, despite the relatively low number of items, DD-I scales have a high degree of internal consistency: all Cronbach's alphas were equal to or larger than .80, except the one of Psychopathy in Sample 3, which however could be considered adequate (.73). Corrected

1 item-total correlations were always well above .30, indicating an adequate discriminativity of DD-I  
2 items between high and low levels of the traits.  
3

## 4 **Study 2 -Temporal stability of the scores on the DD-I**

### 5 *Participants and procedure*

6  
7 The DD-I was administered twice at a 3-week interval to 164 undergraduate psychology students  
8  
9 (mean age  $22.68 \pm 5.50$  years, range 19-59, Females 77%). Students were informed that the  
10  
11 completion of the DD-I was not compulsory, that participation could not affect their final evaluation  
12  
13 and that they could decide to retire their consent to participate at any moment without  
14  
15 consequences.  
16  
17  
18  
19  
20

### 21 *Results and discussion*

22  
23 Cronbach's alphas were always equal to or higher than .80. As measures of test-retest reliability we  
24  
25 computed intraclass correlation coefficients (ICCs). ICCs were computed as single measure using a  
26  
27 two-way random effect model with an absolute agreement definition. The findings indicated that  
28  
29 DD-I observed scale scores were stable over the 3-week interval (ICCs ranging from .83 to .86).  
30  
31 Paired-sample *t*-tests also revealed that the mean difference of scores from Time 1 to Time 2 was  
32  
33 not statistically different from zero (Section 9 of the ESM). Taken together, the results of Study 2  
34  
35 indicated an adequate temporal stability of DD-I scores, at least in a student population.  
36  
37  
38  
39  
40

## 41 **Study 3 – Construct validity of the DD-I**

42  
43 The construct validity of the DD-I was tested using a strategy similar to the one adopted by Jonason  
44  
45 and Webster (2010). We thus administered a battery of questionnaires that included other measures  
46  
47 of the same constructs measured by the DD-I (convergent validity) and measures of other constructs  
48  
49 in the same nomological network of the Dark Triad (discriminant validity).  
50  
51

52  
53 The non-Dark Triad measures included the Big Five, aggressiveness, socio-sexual  
54  
55 orientation, self-esteem, social desirability and impression management. An association of higher  
56  
57 scores on the DD scales with lower scores on agreeableness (Paulhus & Williams, 2002) and  
58  
59 conscientiousness (Jonason, Li & Teicher, 2010) has been consistently replicated by previous  
60  
61  
62  
63  
64  
65

1 studies (Paulhus & Williams, 2002; Jonason, Li & Teicher, 2010). Associations of the other Big  
2 Five constructs with DD scores were found, but did not show a consistent pattern (Jonason &  
3 Webster, 2010).  
4  
5

6  
7 Higher levels of Dark Triad traits have been found to be associated with higher levels of  
8 aggressiveness (e.g., Paulhus & Williams, 2002) and Jonason and Webster (2010) reported positive  
9 significant correlations of DD scores with scores on the Aggression Questionnaire (Buss & Perry,  
10 1992).  
11  
12

13  
14  
15  
16  
17 Jonason and Webster (2010) did not find convincing evidence of an association of DD  
18 scores with self-esteem. Hence, we did not expect significant correlations of DD-I scores with self-  
19 esteem.  
20  
21

22  
23  
24 It has been found that Dark Triad traits are associated with short-term mating, especially in  
25 men (Webster & Bryan, 2007). Consistent with these results, Jonason and Webster (2010) found  
26 that DD scores were more positively associated with short-term mating than with long-term mating.  
27  
28

29  
30  
31 Reported forms of self-presentation as perfectionistic self-promotion, nondisclosure of  
32 imperfection, and non-display of imperfection (Sherry et al., 2006) are common in individuals with  
33 higher levels of Machiavellianism (Lopes & Fletcher, 2004) and narcissism (Rauthmann, 2011).  
34 Hence, we expected that higher scores on the DD scales measuring these constructs were associated  
35 with higher levels of impression management (IM), while it has been reported that the association  
36 of psychopathy scores with IM is weak (see, e.g., Ray et al. 2013). On the other hand, the moralistic  
37 bias, i.e., the tendency to exaggerate communion-related traits such as duty, agreeableness, and  
38 impulse control (Paulhus, 2002), should be negatively correlated with the DD scores, since it  
39 assumes high levels of ego control, achievement via conformity, nurturance, social closeness,  
40 interpersonal sensitivity, restraint, and socialization (Paulhus & John, 1998).  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

### *Participants*

Sixty-six participants (mean age  $32.06 \pm 12.50$  years, range 20-59, Females 67%) took part to Study 3. These participants were recruited through snowball sampling by a psychology student as part of their dissertation project. Inclusion and exclusion criteria were the same of Study 1.

### *Measures*

*Italian Dirty Dozen (DD-I)*. As described above. Descriptive statistics and Cronbach's alpha of the DD-I and of all the other measures used in this study are reported in Table 3.

[Table 3]

*Machiavellianism subscale of the Multidimensional Personality Profile (MPP-Machiavellianism; Caprara, Barbaranelli, De Carlo & Robusto, 2006)*. The MPP is an Italian measure of personality traits. It comprises several subscales scales, among which Machiavellianism, Social Desirability and Impression Management. The Machiavellianism subscale assesses the tendency to put one's own needs ahead of others and to use manipulation, deceit and tactics (e.g., bending the rules) to achieve one's goals.

*Psychopathic Deviate scale of the Minnesota Multiphasic Personality Inventory-2 (MMPI-PD, Butcher et al., 1989; Italian version in Pancheri & Sirigatti, 1995)*. The PD is a subscale of the MMPI that contains 50 true/false items and assesses the individual's degree of social deviation, lack of acceptance of authority, and amorality.

*Narcissistic Personality Inventory (NPI, Raskin & Hall, 1988; Italian version in Fossati & Borroni, 2008a)*. The NPI is a widely used measure of narcissism. It includes 40 items and for each item participants are asked to choose one of two statements they felt applied to them more. One of the two statements reflects a narcissistic attitude more than the other.

*Big Five Questionnaire (BFI, John, Donahue & Kentle, 1991; Italian version in Ubbiali, Chiorri, Hampton, & Donati, 2013)*. The BFI is a 44-item self-report measure of the Big Five (Extraversion [8 items], Agreeableness [9 items], Conscientiousness [9 items], Neuroticism [8

items] and Openness [10 items]) consisting of short phrases that include trait adjectives known to be prototypical markers of the Big Five.

*Aggression Questionnaire* (AQ, Buss & Perry, 1992; Italian version in Fossati & Borroni, 2008b). The AQ is a 29-item self-report measure of perceived levels of anger and aggression. The AQ provides a total score and scores in 4 scales: Physical Aggression (9 items), Verbal Aggression (5 items), Anger (7 items) and Hostility (8 items).

*Rosenberg's Self-Esteem Scale* (RSES, Rosenberg, 1965; Italian version in Prezza, Trombaccia, & Armento, 1997). RSES is a 10-item self-report measure of global self-esteem.

*Sociosexual Orientation Inventory-Revised* (SOI-R, Penke & Asendorpf, 2008; Italian version available at: <http://www.larspenke.eu/en/translated-soi-r.html>). The SOI-R is a measure of sociosexual orientation, with high scores indicating an "unrestricted" sociosexual orientation (i.e., an overall more promiscuous behavioral tendency) and low scores indicating a "restricted" sociosexual orientation. It provides a total score and three subscale scores: Behavior (3 items), Attitude (3 items), Desire (3 items).

*Social Desirability subscale of the Multidimensional Personality Profile* (MPP-Social Desirability; Caprara et al., 2006). The MPP-Social Desirability is an 8-item measure of the moralistic bias, i.e., a self-deceptive tendency to deny socially deviant impulses and behaviors and to claim "saint-like" attributes (Paulhus & John, 1998).

*Impression Management subscale of the Multidimensional Personality Profile* (MPP-Impression Management; Caprara et al., 2006). The MPP-Impression Management is a 8-item measure of the egoistic bias, i.e., a self-deceptive tendency to exaggerate one's social and intellectual status (Paulhus & John, 1998).

### *Procedure*

All participants were tested individually and anonymously. The scales included in the battery were administered in counterbalanced fashion to control for order and sequence effects.

### *Results and discussion*

1  
2 Results are reported in Table 3. As expected, the DD-I scales correlated significantly and  
3  
4 substantially (i.e., correlations in the .40s) with measures of the same constructs, partially  
5  
6 supporting their convergent validity, since, especially for Machiavellianism and psychopathy,  
7  
8 similar correlations were found also with other DT traits. They also showed significant negative  
9  
10 correlations with agreeableness, but not with conscientiousness, albeit the effect sizes of  
11  
12 correlations with the psychopathy and narcissism scales was comparable to those of Jonason and  
13  
14 Webster (2010, Table 4, p. 424). Significant positive correlations were found between extraversion,  
15  
16 on the one hand, and Machiavellianism and narcissism, on the other, consistent with Jonason et al.  
17  
18 (2010). DD-I scales were also positively and significantly associated with AQ scales and SOI-R  
19  
20 scores (except SOI-R-Behavior), consistent with expectations and Jonason and Webster's (2010)  
21  
22 results. Finally, the correlations of DD-I scale scores with measures social desirability and  
23  
24 impression management were consistent with the hypotheses, since higher Machiavellianism and  
25  
26 narcissism scores were associated with lower levels of moralistic bias and higher levels of egoistic  
27  
28 bias.  
29  
30  
31  
32  
33  
34  
35

36 While the results of these study substantially met the expectations with respect to convergent  
37  
38 validity, they seem to provide little support for the discriminant validity of the DD-I, since  
39  
40 correlations with other constructs often were as large as the correlations with the same constructs.  
41  
42 This is a known limitation of the DD, as other studies (e.g., Maples et al., 2014; Miller et al., 2012)  
43  
44 raised some concerns about the construct validity of the DD, especially for the psychopathy  
45  
46 subscale. However, it must be noted that the results presented here are consistent with the ones  
47  
48 reported in the above-mentioned and other studies about the DD (e.g., Czarna, Jonason, Dufner, &  
49  
50 Kossowska, 2016; Jonason et al., 2010; Küfner, Dufner, & Back, 2014). As argued by Czarna et al.  
51  
52 (2016), such findings might be due to the fact that the original Dark Triad measures assess multiple  
53  
54 facets of the respective constructs while the DD-I might only measure the core aspects.  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

## Study 4 Measurement invariance of the DD-I across gender

### Participants

Participants were recruited in Central Italy through a snowball sampling procedure in which students were given the DD-I to pass on to members of their families and acquaintances. The total number of participants was 974 (56.9% females). Mean age was 36.45 years (SD=13.21, range 18–80). The gender subgroups were adequately balanced on these and other background characteristics (Section 10 of the ESM). Inclusion and exclusion criteria were the same of Study 1.

### Statistical Analyses

Measurement invariance of the DD-I was tested using Multi-Group Confirmatory Factor Analysis (MG-CFA). We first tested the ability of the *a priori* three-correlated-factor model to fit the data in the total sample and, separately, in the groups defined by gender. We then tested measurement invariance across gender. The hypothesized factor structure was estimated simultaneously in the two groups defined by gender (*configural invariance model*, M0). This model tested whether the same factor structure was maintained across groups. We then imposed equality on factor loadings (*weak invariance model*, M1) to test whether items showed a proportional amount of increase between women and men for the same amount of increase on the latent factor. Latent scores could be compared only if women and men with similar levels on the construct presented comparable scores on the items reflecting the construct; hence, item intercepts were constrained to be invariant (*strong invariance model*, M2). Manifest scores could be compared if the constructs were assessed with similar levels of measurement errors in women and men; hence, item residual variances were also constrained to be invariant (*strict invariance model*, M3).

We also tested models in which latent factor variances (M4) and covariances (M5) were constrained to be invariant. M4 implied that women and men used the same range on the factor continuum to report their levels of DT traits and that same items had equal levels of reliability across gender. M5 assumed that the correlation between the same factor pairs was the same for women and men (see Section 11 of the ESM).

The goodness of fit (GOF) of the CFA models was evaluated with the same criteria of Study

1. Measurement invariance models were also compared using fit indices. We considered as evidence of invariance a change in CFI of less than .01 or a change in RMSEA of less than .015 (Chen, 2007).

### *Results*

We first tested the fit of a three-correlated-factor CFA model in the total sample of participants and in the groups defined by gender. Results showed that the hypothesized model had an adequate fit (Table 4).

[Table 4]

We then tested the measurement invariance of the DD-I across gender. Factor loadings, item intercepts, residual variances and factor correlations of the configural invariance model (M0) are reported in Section 12 of the ESM. All parameter estimates were statistically different from zero ( $p < .001$ ).

The inspection of the GOF indices for invariance models reported in Table 4 revealed that the measurement invariance of the DD-I factor model across gender was fully supported for all models. As shown in the rightmost columns of Table 4, invariance models in which the factor mean differences were estimated revealed that men scored higher than women in all factors, albeit in narcissism the difference had a smaller effect size – note that the standardized mean difference estimates reported in Table 4 are in Cohen's  $d$  metric.

### **Discussion**

Previous studies had tested sex differences on the Dirty Dozen (DD) using observed scores, possibly overlooking that these comparisons might have been biased by a lack of measurement invariance of the scale. The only exception was the Klimstra et al. (2014)'s study, that found support for the measurement invariance of the DD across gender and reported that males tended to endorse higher scores in all scales. However, these results were of somehow limited generalizability, as they



1 were obtained on samples of adolescents. In this study we were able to replicate these results on a  
2 large adult sample and extended them by using a larger taxonomy of invariance models.  
3

4 Before addressing the issue of the measurement invariance of the DD, we had to adapt the  
5 DD into Italian. We thus carried out three studies, whose results showed that the Italian DD has  
6  
7 psychometric properties that overlap those of the US version.  
8  
9

10 The results of the measurement invariance analyses revealed that the measurement model of  
11 the DD-I and its parameters are invariant across gender. Specifically, we found that also factor  
12 variances and covariances are invariant between women and men. This means that (1) women and  
13 men used the same range on the factor continuum to report their levels of DT traits (i.e., narcissism,  
14 psychopathy, and Machiavellianism) and that the same items have the same reliability across  
15  
16 gender; (2) the correlation between the same factor pairs for women is statistically equivalent to the  
17 correlation between the same factors pairs for men.  
18  
19

20 This result implies that mean score differences actually reflect differences in the amount of  
21 constructs as they are operationalized by the DD. Consistent with previous studies (Furnham &  
22 Trickey, 2011), we found that men's scores on all DD-I scales were statistically higher than  
23 women's, with higher effect sizes for Machiavellianism and psychopathy (0.40s) than for narcissism  
24 (0.15s). One possible explanation of this result can be the multidimensional nature of narcissism, as  
25 suggested by recent research that supports the existence of two phenotypic expressions of  
26 narcissism, characterized by grandiose and vulnerable features (Cain et al., 2008). Men usually  
27 score higher on the grandiose dimension of narcissism, whereas smaller, or even null, differences  
28 are usually found on vulnerable narcissism features (Grijalva et al., 2015). The narcissism scale of  
29 the DD has apparently the merit to capture both faces of narcissism (Maples et al., 2014), thus the  
30 smaller differences on the narcissism scale (when compared with the other two scales) could  
31 resemble the presence of both grandiose and vulnerable features which are likely to differ across  
32  
33 gender. However, the lack of distinction between overt vs covert narcissism (Wink, 1991) is  
34 potentially problematic and deserves further investigation.  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 As argued by Grijalva et al. (2015), sex differences can be explained in terms of a biosocial  
2 approach to social role theory. Many of the correlates of DT traits reflect high levels of agentic  
3 characteristics rather than communal characteristics (Jones & Paulhus, 2010). Rudman, Moss-  
4 Racusin, Phelan, and Nauts (2012) found that being agentic and not being communal is considered  
5 desirable for men and undesirable for women.  
6  
7  
8  
9  
10

11 As individuals might be socially penalized for deviating from gender role expectations,  
12 women may experience societal pressure for communal behaviors and face disapproval for  
13 displaying agentic behaviors. Hence, they may be less likely *to report* DT traits. Not surprisingly,  
14 "the entire construct of Machiavellianism [is considered] more appropriate for men than for  
15 women" (Wilson, Near & Miller, 1996, p. 293).  
16  
17  
18  
19  
20  
21  
22  
23

24 Bivariate associations of the DT traits with external correlates were meaningful and largely  
25 consistent with prior studies (Jonason & Webster, 2010). Machiavellianism, psychopathy, and  
26 narcissism were negatively related to agreeableness, confirming that a disagreeable attitude toward  
27 other might represent a shared feature of DT traits. Furthermore, Machiavellianism and narcissism,  
28 but not psychopathy, were positively associated with extraversion, paralleling findings with  
29 adolescents (Klimstra et al., 2014). However, it should be noted that the association of  
30 Machiavellianism with extraversion is not consistently found in other works (see e.g., O'Boyle,  
31 Forsyth, Banks, Story, & White, 2015). Aggression dimensions and unrestricted sexual orientation  
32 were also positively correlated with DT traits, confirming and expanding current knowledge of the  
33 potentially risky interpersonal consequences of sub-clinical levels of Machiavellianism,  
34 psychopathy, and narcissism. Finally, Machiavellianism and narcissism – but not psychopathy –  
35 were positively linked with impression management and negatively with social desirability. This  
36 could suggest that higher levels of these dark traits are associated with an increased tendency to  
37 exaggerate personal attributes and status, but not to deny antagonistic impulses.  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57

58 Some limitations need to be pointed out. First, social desirability tends to be positively  
59 related to age and negatively related to undesirable self-report characteristics. Since we used a self-  
60  
61  
62  
63  
64  
65

report measure, this might have elicited the manipulative tendencies of people high DT traits.

Second, the conciseness of the DD-I might have limited the possibility to cover the full breadth of traits and components characterizing the DT (Miller et al., 2012) and thus the possibility to disentangle specific facets of each personality style within the DT construct. However, our results are consistent with those obtained with longer measures.

Despite these limitations, this study showed that each DD-I factor actually captures the same construct among adult women and men, suggesting that sex differences in scale scores found by previous studies cannot be considered as artifacts due to measurement error. Machiavellianism, psychopathy and narcissism seem thus to have the same "faces" in both women and men, with men reporting higher levels of all traits.

### References

- American Psychological Association. (2010). *Ethical principles of psychologists and code of conduct*. Retrieved from <http://apa.org/ethics/code/index.aspx>
- Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, 16(3), 397–438. doi:10.1080/10705510903008204
- Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology*, 63(3), 452–459. doi:10.1037/0022-3514.63.3.452
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *MMPI-2: Manual for administration and scoring*. Minneapolis, MN: University of Minnesota Press.
- Cain, N. M., Pincus, A. L., & Ansell, E. B. (2008). Narcissism at the crossroads: phenotypic description of pathological narcissism across clinical theory, social/personality psychology, and psychiatric diagnosis. *Clinical Psychology Review*, 28(4), 638–656. doi:10.1016/j.cpr.2007.09.006
- Cale, E. M., & Lilienfeld, S. O. (2002). Sex differences in psychopathy and antisocial personality disorder: A review and integration. *Clinical Psychology Review*, 22(8), 1179–1207. doi:10.1016/S0272-7358(01)00125-8

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65
- Caprara, G. V., Barbaranelli, C., De Carlo, N. A., & Robusto, E. (Eds.). (2006). *Multidimensional Personality Profile*. Milano: Franco Angeli.
- Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling, 14*(3), 464–504. doi:10.1080/10705510701301834
- Czarna, A. Z., Jonason, P. K., Dufner, M., & Kossowska, M. (2016). The Dirty Dozen Scale: Validation of a Polish version and extension of the nomological net. *Frontiers in Psychology, 7*:445. doi: 10.3389/fpsyg.2016.00445
- De Winter, J. C. F., Dodou, D., & Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research, 44*(2), 147–181. doi:10.1080/00273170902794206
- Fossati, A., & Borroni, S. (2008a). Versione italiana del Narcissistic Personality Inventory [Italian version of the Narcissistic Personality Inventory]. In C. Maffei (Ed.), *Borderline* (pp. 387–413). Milano: Raffaello Cortina.
- Fossati, A., & Borroni, S. (2008b). Versione italiana dell'Aggression Questionnaire [Italian version of the Aggression Questionnaire]. In C. Maffei (Ed.), *Borderline* (pp. 279–307). Milano: Raffaello Cortina.
- Furnham, A., & Trickey, G. (2011). Sex differences in the dark side traits. *Personality and Individual Differences, 50*(4), 517–522. doi:10.1016/j.paid.2010.11.021
- Grijalva, E., Newman, D. A., Tay, L., Donnellan, M. B., Harms, P. D., Robins, R. W., & Yan, T. (2015). Sex differences in narcissism: A meta-analytic review. *Psychological Bulletin, 141*(2), 261–310. doi:10.1037/a0038231
- Hare, R. D., & Neumann, C. S. (2008). Psychopathy as a clinical and empirical construct. *Annual Review of Clinical Psychology, 4*, 217–246. doi:10.1146/annurev.clinpsy.3.022806.091452
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179–185. doi:10.1007/BF02289447

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65
- John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory-Versions 4a and 54*. Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.
- Jonason, P. K., & Luévano, V. X. (2013). Walking the thin line between efficiency and accuracy: Validity and structural properties of the Dirty Dozen. *Personality and Individual Differences*, 55(1), 76–81. doi:10.1016/j.paid.2013.02.010
- Jonason, P. K., & Webster, G. D. (2010). The Dirty Dozen: A concise measure of the dark triad. *Psychological Assessment*, 22(2), 420–432. doi:10.1037/a0019265
- Jonason, P. K., Li, N. P., & Teicher, E. A. (2010). Who is James Bond?: The dark triad as an agentic social style. *Individual Differences Research*, 8(2), 111–120.
- Jones, D. N., & Paulhus, D. L. (2009). Machiavellianism. In M. R. Leary & R. H. Hoyle (Eds.), *Handbook of individual differences in social behavior* (pp. 93–108). New York: Guilford Press.
- Jones, D. N., & Paulhus, D. L. (2010). Differentiating the Dark Triad within the interpersonal circumplex. In L. M. Horowitz & S. N. Strack (Eds.), *Handbook of interpersonal theory and research* (pp. 249–267). New York: Guilford Press.
- Klimstra, T. A., Sijtsma, J. J., Henrichs, J., & Cima, M. (2014). The Dark Triad of personality in adolescence: Psychometric properties of a concise measure and associations with adolescent adjustment from a multi-informant perspective. *Journal of Research in Personality*, 53, 84–92. doi:10.1016/j.jrp.2014.09.001
- Küfner, C. P., Dufner, M., and Back, M. D. (2014). Das Dreckige Dutzend und die Niederträchtigen Neun – Kurzskalen zur Erfassung von Narzissmus, Machiavellismus und Psychopathie. *Diagnostica*, 61, 79–91. doi: 10.1026/0012-1924/a000124
- Lopes, J., & Fletcher, C. (2004). Fairness of impression management in employment interviews: A cross-country study of the role of equity and Machiavellianism. *Social Behavior and Personality*, 32(8), 747–768. doi:10.2224/sbp.2004.32.8.747

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65
- Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology*, 2(2), 57–64. doi:10.1027/1614-2241.2.2.57
- 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(1), 326–331. doi:10.1037/a0035084
- Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, 11(3), 320–341. doi:10.1207/s15328007sem1103\_2
- 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(3), 471–91. doi:10.1037/a0019227
- Miller, J. D., Few, L. R., Seibert, L. A., Watts, A., Zeichner, A., & Lynam, D. R. (2012). An examination of the Dirty Dozen Measure of psychopathy: A cautionary tale about the costs of brief measures. *Psychological Assessment*, 24(4), 1048–1053. doi:10.1037/a0028583.
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York: Taylor & Francis Ltd.
- Morf, C. C., & Rhodewalt, F. (2001). Expanding the dynamic self-regulatory processing model of narcissism: Research directions for the future. *Psychological Inquiry*, 12(4), 243–251. doi:10.1207/S15327965PLI1204\_3
- Muthén, B., & Muthén, L. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Muthén, L., & Muthén, B. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 9(4), 599–620. doi:10.1207/S15328007SEM0904\_8

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65
- 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. *Journal of Personality*, 83(6), 644–664. doi:10.1111/jopy.12126
- Pancheri, P., & Sirigatti, S. (Eds.). (1995). *MMPI-2 - Minnesota Multiphasic Personality Inventory - 2. Manuale*. Firenze, Italy: Giunti O.S.
- Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In D. N. Jackson & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality*, 66(6), 1025. doi:10.1111/1467-6494.00041
- Paulhus, D. L., & Williams, K. M. (2002). The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, 36(6), 556–563. doi:10.1016/S0092-6566(02)00505-6
- Penke, L., & Asendorpf, J. B. (2008). Beyond global sociosexual orientations: a more differentiated look at sociosexuality and its effects on courtship and romantic relationships. *Journal of Personality and Social Psychology*, 95(5), 1113–1135. doi:10.1037/0022-3514.95.5.1113
- Prezza, M., Trombaccia, F. R., & Armento, L. (1997). La scala dell'autostima di Rosenberg: traduzione e validazione italiana [The Rosenberg Self-Esteem Scale: Italian translation and validation]. *Bollettino Di Psicologia Applicata*, 223, 35–44.
- Raskin, R. N., & Hall, C. S. (1979). A narcissistic personality inventory. *Psychological Reports*, 45(2), 590. doi:10.2466/pr0.1979.45.2.590
- Rauthmann, J. F. (2011). Acquisitive or protective self-presentation of dark personalities? Associations among the Dark Triad and self-monitoring. *Personality and Individual Differences*, 51(4), 502–508. doi:10.1016/j.paid.2011.05.008

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65
- Ray, J. V., Hall, J., Rivera-Hudson, N., Poythress, N. G., Lilienfeld, S. O., & Morano, M. (2013).  
The relation between self-reported psychopathic traits and distorted response styles: A meta-  
analytic review. *Personality Disorders: Theory, Research, and Treatment*, 4(1), 1–14.  
doi:10.1037/a0026482
- Revelle, W. (2015). *psych: Procedures for Personality and Psychological Research*. Northwestern  
University, Evanston, Illinois, USA, <http://CRAN.R-project.org/package=psych> Version =  
[1.5.4](#).
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University  
Press.
- Rudman, L. A., Moss-Racusin, C. A., Phelan, J. E., & Nauts, S. (2012). Status incongruity and  
backlash effects: Defending the gender hierarchy motivates prejudice against female leaders.  
*Journal of Experimental Social Psychology*, 48(1), 165–179. doi:10.1016/j.jesp.2011.10.008
- Sherry, S. B., Hewitt, P. L., Besser, A., Flett, G. L., & Klein, C. (2006). Machiavellianism, trait  
perfectionism, and perfectionistic self-presentation. *Personality and Individual Differences*,  
40(4), 829–839. doi:10.1016/j.paid.2005.09.010
- Tucker, L. R. (1951). *A method for synthesis of factor analysis studies (Personnel Research Section.  
Report No. 984)*. Washington, DC.
- Ubbiali, A., Chiorri, C., Hampton, P., & Donati, D. (2013). Psychometric properties of the Italian  
adaptation of the Big Five Inventory (BFI). *Bollettino di Psicologia Applicata*, 266, 37–48.
- Velicer, W. (1976). Determining the number of components from the matrix of partial correlations.  
*Psychometrika*, 41(3), 321–327. doi:10.1007/BF02293557
- Velotti, P., Elison, J., & Garofalo, C. (2014). Shame and aggression: Different trajectories and  
implications. *Aggression and Violent Behavior*, 19(4), 454–461.  
doi:10.1016/j.avb.2014.04.011



1 Webster, G. D., & Bryan, A. (2007). Sociosexual attitudes and behaviors: Why two factors are  
2 better than one. *Journal of Research in Personality*, 41, 917–922.

3  
4 doi:10.1016/j.jrp.2006.08.007

5  
6  
7 Webster, G. D., & Jonason, P. K. (2013). Putting the “IRT” in “Dirty”: Item response theory

8 analyses of the Dark Triad Dirty Dozen-An efficient measure of narcissism, psychopathy, and  
9 Machiavellianism. *Personality and Individual Differences*, 54(2), 302–306.

10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
doi:10.1016/j.paid.2012.08.027

17 Wilson, D. S., Near, D., & Miller, R. R. (1996). Machiavellianism: a synthesis of the evolutionary  
18 and psychological literatures. *Psychological Bulletin*, 119(2), 285–299. doi:10.1037/0033-

19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
2909.119.2.285

24 Wink, P. (1991). Two faces of narcissism. *Journal of Personality and Social Psychology*, 61(4),

25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
590–597. doi:10.1037/0022-3514.61.4.590

1 Table 1 Results of exploratory factor analyses on the Italian Dirty Dozen in samples 1 and 2 in  
 2 Study 1  
 3

Item	Sample 1 ( $n = 102$ )			Sample 2 ( $n = 128$ )		
	M	P	N	M	P	N
DD01	.85	.01	-.09	.62	.07	.14
DD02	.68	.03	.11	.89	.04	-.19
DD03	.87	-.27	-.04	.81	-.13	.05
DD04	.77	.09	.02	.79	-.01	.09
DD05	-.19	.96	-.09	-.17	.70	.04
DD06	.07	.63	.00	-.05	.94	-.09
DD07	.04	.69	.00	.11	.71	.05
DD08	.33	.48	.00	.26	.51	.03
DD09	-.04	-.15	.86	.01	.00	.78
DD10	.07	.00	.65	.04	-.11	.84
DD11	.06	.06	.54	-.15	.04	.86
DD12	-.11	.05	.73	.07	.00	.75
<i>r</i> with P	.54			.49		
<i>r</i> with N	.53	.40		.56	.29	

25  
 26 Note: M = Machiavellianism; P = Psychopathy; N = Narcissism; *r* = Pearson's correlation  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65

Table 2 Goodness-of-fit statistics of confirmatory factor analyses performed on data from Sample 3 ( $n = 305$ ) in Study 1

Model and description	$\chi^2$	df	CFI	TLI	RMSEA
1-factor model	370.77	54	.674	.601	.139
3-independent-factor model	225.88	54	.823	.784	.102
3-correlated-factor model	103.639	51	.946	.930	.058
Bifactor model	65.43	42	.976	.962	.043

Note.  $\chi^2$  = Chi-square; df = degrees of freedom; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation

Table 3 Descriptive statistics and correlations of the Italian Dirty Dozen scale scores with scores on measures of Machiavellianism, psychopathy, narcissism, personality, aggression, self-esteem, social desirability, impression management and sociosexual orientation (Study 3)

Variable	M	P	N	$\alpha$	M	SD
M	1.00			.80	2.57	1.30
P	.50***	1.00		.85	2.58	1.28
N	.66***	.24	1.00	.89	3.94	1.90
Convergent validity						
MPP - Machiavellianism	.48***	.33**	.28*	.75	2.45	0.51
MMPI-PD Total Score	.25*	.41**	.15	.62	20.98	4.92
NPI	.44***	.37**	.54***	.76	11.02	5.32
Discriminant validity						
BFI - Extraversion	.28*	.17	.28*	.83	3.44	0.65
BFI - Agreeableness	-.31*	-.34**	-.29*	.72	3.73	0.51
BFI - Conscientiousness	-.07	-.22	-.23	.86	3.51	0.76
BFI - Neuroticism	.02	-.17	.20	.83	2.99	0.64
BFI - Openness	.03	-.02	.15	.87	3.84	0.69
AQ - Physical Aggression	.49***	.54***	.37**	.75	2.53	0.67
AQ - Verbal Aggression	.36**	.33**	.31*	.72	2.16	0.71
AQ - Anger	.42***	.37**	.45***	.72	2.38	0.57
AQ - Hostility	.54***	.33**	.57***	.77	2.01	0.60
Self-esteem	.08	.14	-.07	.87	3.04	0.53
SOI-R - Behavior	.18	.13	.23	.75	1.87	0.92
SOI-R - Attitude	.41**	.40**	.29*	.87	3.70	2.22
SOI-R - Desire	.37**	.34**	.43***	.88	3.14	2.02
SOI-R - Total score	.44***	.42**	.42***	.84	2.90	1.36
MPP - Social Desirability	-.34**	-.10	-.35**	.76	2.71	0.58
MPP - Impression Management	.47***	-.03	.58***	.71	3.44	0.44

Note:  $n = 66$ ; M = Machiavellianism; P = Psychopathy; N = Narcissism;  $\alpha$  = Cronbach's alpha; M = Mean; SD = Standard deviation; MPP = Multidimensional Personality Profile; MMPI-PD = Minnesota Multiphasic Personality Inventory – Psychopathic Deviate; NPI = Narcissistic Personality Inventory; BFI = Big Five Inventory; AQ = Aggression Questionnaire; SOI-R = Sociosexual Orientation Inventory – Revised; \* =  $p < .05$ ; \*\* =  $p < .01$ ; \*\*\* =  $p < .001$ ;

Table 4 Goodness-of-fit statistics of confirmatory factor analytic and measurement invariance models in Study 4

Model and description	$\chi^2$	df	CFI	TLI	RMSEA	M	P	N
Total group	196.303	51	.956	.943	.054	-	-	-
Women	133.011	51	.953	.939	.054	-	-	-
Men	121.359	51	.953	.939	.057	-	-	-
Invariance models								
M0 Configural	255.136	102	.953	.939	.056	-	-	-
M1 Weak	260.863	111	.954	.945	.053	-	-	-
M2 Strong	288.541	120	.949	.944	.054	.400***	.445***	.154*
M3 Strict	330.714	132	.942	.942	.056	.397***	.442***	.154*
M4 Factor variances invariant	347.337	135	.939	.940	.057	.438***	.481***	.150*
M5 Factor covariances invariant	350.252	138	.939	.940	.056	.438***	.481***	.151*

*Note.*  $\chi^2$  = Chi-square; df = degrees of freedom; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation; M = standardized factor mean difference for Machiavellianism; P = standardized factor mean difference for Psychopathy; N = standardized factor mean difference for Narcissism. Positive standardized factor mean differences indicate higher scores in men. \*\*\*  $p < .001$ . \*  $p < .05$

RUNNING HEAD: ITALIAN VERSION OF THE DD

Psychometric properties of an Italian version of the Dirty Dozen: Factor structure, reliability,  
validity, measurement invariance across gender

Carlo Chiorri<sup>1</sup>, Carlo Garofalo<sup>2</sup>, Patrizia Velotti<sup>1</sup>

<sup>1</sup> Department of Educational Sciences, University of Genova, Corso A. Podestà, 2, 16128 Genova  
(Italy)

<sup>2</sup> Department of Developmental Psychology, Tilburg University, PO Box 90153, 5000 LE Tilburg  
(The Netherlands)

Correspondence to:

Carlo Chiorri, PhD

Department of Educational Sciences

University of Genova

Corso A. Podestà, 2

16128 Genova (Italy)

E-mail: [carlo.chiorri@unige.it](mailto:carlo.chiorri@unige.it)

Tel: +39 010 20953709

Fax: +39 010 20953728

**Electronic Supplementary Materials for the paper *Psychometric properties of an Italian version of the Dirty Dozen: Factor structure, reliability, validity, measurement invariance across gender***

**1. Review of sex differences (Cohen's *d*) in scores on the Dirty Dozen (studies are in chronological order)**

Source	Participants	M	P	N
Jonason & Webster (2010)				
Study 1	273 psychology students (90 men, 183 women) aged 18-47 years (M=20.08, SD=3.79)	0.49	0.40	0.62
Study 2	246 psychology students (101 men, 145 women) aged 18-42 years (M=20.69, SD=3.76)	0.21	0.42	0.62
Study 3	96 undergraduate students (36 men, 60 women), aged 18-25 years (M=20.44, SD=1.43)	0.79	0.35	0.34
Study 4	470 psychology students (157 men, 312 women) aged 17-26+ years (mode=18, Median=19, M=19.00, SD=1.30)	0.05	0.46	0.09
Jonason & Krause (2013)	320 online participants (78 men, 242 women), aged 17-56 years (M=24.24, SD=7.33)	0.75	0.49	0.51
Muris, Meesters & Timmermans (2013)	117 adolescents (51 men, 66 women), aged 12-18 years (M=13.90, SD=0.96)	0.45	0.56	0.01
Webster & Jonason (2013)	544 undergraduate students (169 men, 375 women), aged 17-50 (M=20.25, SD=4.70)	0.31	0.41	0.40
Aghababaei, Mohammadtabar, & Saffarinia (2014)	223 Iranian employees (90 men, 133 women), aged 18-57 (M=31.24, SD=8.94).	0.42	0.42	0.18
Jonason, Baughman, Carter & Parker (2015)	1,389 undergraduate students (458 men, 931 women), aged 18-50 years (M=18.88, SD=2.15)	0.22	0.52	0.27
Czarna, Jonason, Dufner, & Kossowska (2016)				
Study 1	304 undergraduate students (111 men, 193 women) aged 18-54 years (M=22.24, SD=4.69)	0.44	0.54	0.02
Study 2	136 undergraduate students (53 men, 83 women), aged 18-48 years (M=24.40, SD=6.60)	0.45	0.36	0.01

Note: M = Machiavellianism; P = Psychopathy; N = Narcissism. Positive *ds* indicate higher scores in males.

**References**

- Aghababaei, N., Mohammadtabar, S., & Saffarinia, M. (2014). Dirty Dozen vs. the H factor: Comparison of the Dark Triad and Honesty-Humility in prosociality, religiosity, and happiness. *Personality and Individual Differences*, *67*, 6–10. doi:10.1016/j.paid.2014.03.026
- Czarna, A. Z., Jonason, P. K., Dufner, M., & Kossowska, M. (2016). The Dirty Dozen Scale: Validation of a Polish version and extension of the nomological net. *Frontiers in Psychology*, *7*:445. doi: 10.3389/fpsyg.2016.00445
- Jonason, P. K., & Krause, L. (2013). The emotional deficits associated with the Dark Triad traits: Cognitive empathy, affective empathy, and alexithymia. *Personality and Individual Differences*, *55*, 532–537. <http://dx.doi.org/10.1016/j.paid.2013.04.027>
- Jonason, P. K., & Webster, G. D. (2010). The Dirty Dozen: A concise measure of the dark triad. *Psychological Assessment*, *22*(2), 420–432. doi:10.1037/a0019265
- Jonason, P. K., Baughman, H. M., Carter, G. L., & Parker, P. (2015). Dorian Gray without his portrait: Psychological, social, and physical health costs associated with the Dark Triad. *Personality and Individual Differences*, *78*, 5–13. doi:10.1016/j.paid.2015.01.008
- Muris, P., Meesters, C., & Timmermans, A. (2013). Some youths have a gloomy side: Correlates of the dark triad personality traits in non-clinical adolescents. *Child Psychiatry & Human Development*, *44*(5), 658–665. doi:10.1007/s10578-013-0359-9
- 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*(2), 302–306. doi:10.1016/j.paid.2012.08.027



## 2. Italian version of the Dirty Dozen

### DD-I

Troverai qui di seguito alcune affermazioni che possono o meno descrivere il tuo modo di essere, di pensare e di comportarti. Indica per ogni affermazione il tuo grado di accordo, ossia quanto l'affermazione ti sembra appropriata a descrivere la tua personalità, ricordando che 1 = fortemente in disaccordo e 7 = fortemente d'accordo

Non ci sono risposte "giuste" o "sbagliate". la migliore risposta è sempre quella che per prima ti viene in mente, in quanto è quella che ha la maggiore probabilità di essere la più sincera e quella che più si avvicina alla tua esperienza.

1. Tendo a manipolare gli altri per ottenere ciò che voglio	1	2	3	4	5	6	7
2. Ho ingannato o mentito per ottenere ciò che volevo	1	2	3	4	5	6	7
3. Ho fatto ricorso all'adulazione per ottenere ciò che volevo	1	2	3	4	5	6	7
4. Tendo a sfruttare gli altri per raggiungere i miei scopi	1	2	3	4	5	6	7
5. Tendo a non provare rimorso	1	2	3	4	5	6	7
6. Tendo a non preoccuparmi della moralità delle mie azioni	1	2	3	4	5	6	7
7. Tendo a essere duro o insensibile	1	2	3	4	5	6	7
8. Tendo a essere cinico	1	2	3	4	5	6	7
9. Tendo a volere l'ammirazione degli altri	1	2	3	4	5	6	7
10. Tendo ad esigere che gli altri mi prestino attenzione	1	2	3	4	5	6	7
11. Tendo a ricercare il prestigio ed un elevato status sociale	1	2	3	4	5	6	7
12. Tendo ad aspettarmi un trattamento speciale da parte degli altri	1	2	3	4	5	6	7

### 3. Power Analysis Study 1

The power analysis for confirmatory factor analysis (CFA) models of Study 1 was carried out using the procedure described in Muthén and Muthén (2002). The method relies on Monte Carlo simulations in which data are generated from a population with hypothesized parameter values. Ten thousand samples are drawn, and a model is estimated for each sample. Parameter values and standard errors are averaged over the samples and the following criteria are examined: parameter estimate bias, standard error bias, and coverage. In this case we followed the guidelines provided by the *Mplus User's Guide* (Muthén & Muthén, 1998–2010), Example 12.12, with the following settings for starting values:

- .80 for target loadings
- .00 for cross-loadings
- 2.5 for intercepts
- 1.00 for factor variances
- .50 for factor correlations in one group
- .20 for uniquenesses (residual variances)
- .00 for factor means

Muthén and Muthén (2002) suggest considering, as a first criterion, that parameter and standard error biases do not exceed 10% for any parameter in the model. The second criterion is that the standard error bias for the parameter for which power is being assessed does not exceed 5%. The third criterion is that coverage (i.e., the proportion of the replications where a 95% confidence interval covers the true parameter value) remains between .91 and .98. Once these three conditions are satisfied, the sample size is considered to keep power close to 0.80, a commonly accepted value for sufficient power.

We tested the power achieved by 5 different sample sizes: 100, 150, 200, 350 and 300. Results are reported in Table 1 and suggested that only Sample 3 afforded a sufficient statistical power to test the expected 3-correlated-factor CFA model.

Table 1 Parameter and standard error highest absolute bias and coverage for five different sample sizes to test the factor structure of the Dirty Dozen

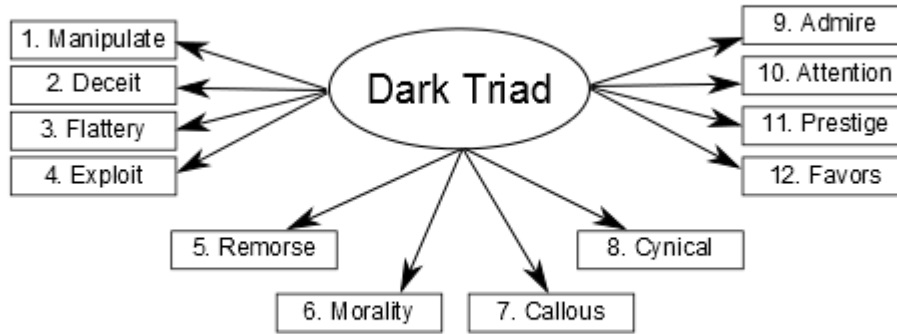
Criteria	n= 100	n =150	n = 200	n = 250	n = 300
Parameter bias	6.78%	4.68%	3.60%	2.88%	2.54%
Standard error bias	18.39%	14.90%	13.21%	11.19%	9.51%
Coverage range	90.90-97.60%	92.30-97.30%	93.00-97.60%	93.40-97.50%	93.50-97.60%

### References

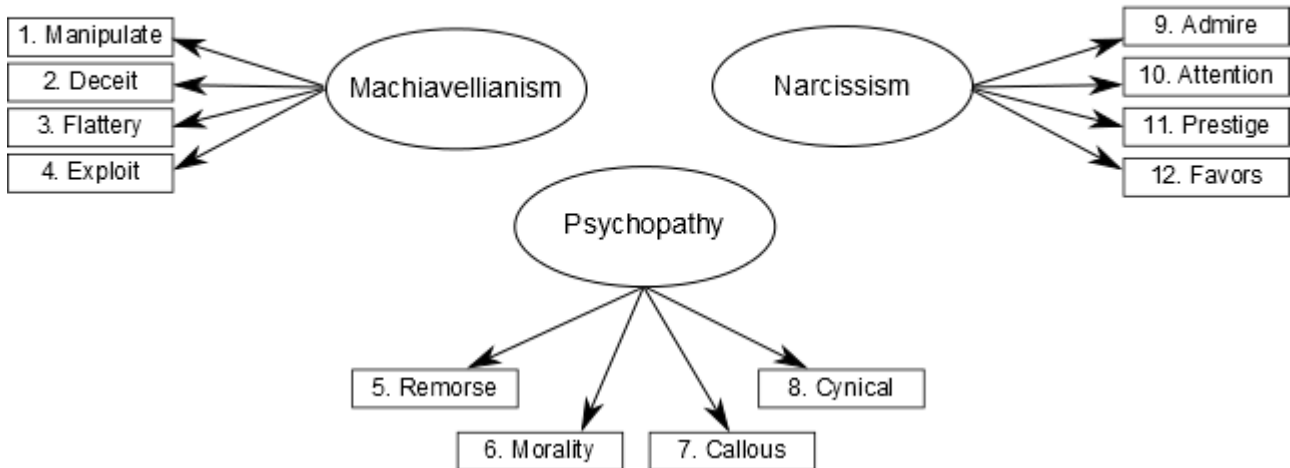
- Muthén, B. & Muthén, L. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 4, 599–620: doi: 10.1207/S15328007SEM0904\_8
- Muthén, L. K., & Muthén, B. (1998–2010). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.

4. Confirmatory factor analysis models in Study 1

1-factor model



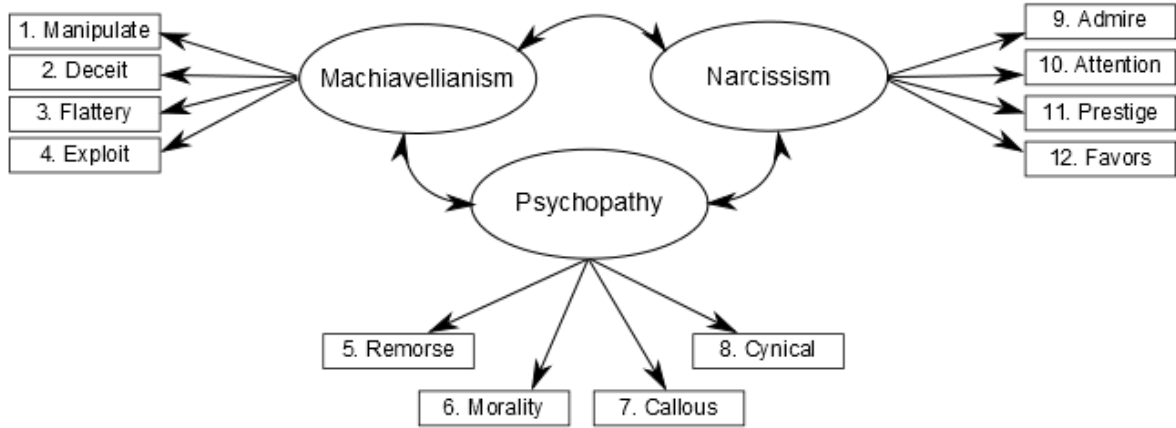
3-independent-factor model



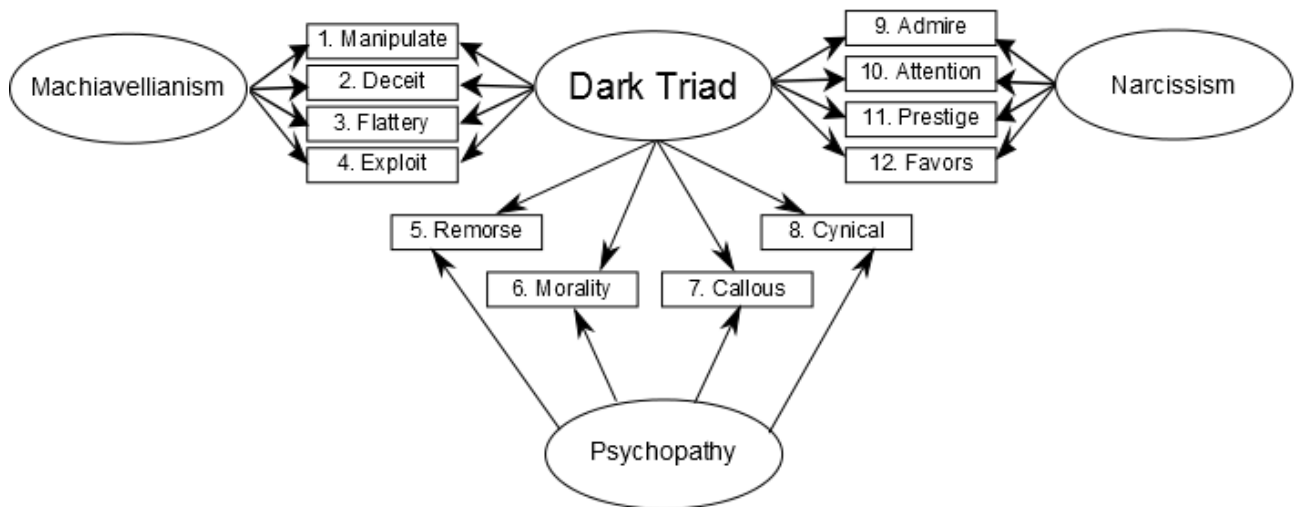
Continue

4. Confirmatory factor analysis models in Study 1 (continued)

3-correlated-factor model



Bifactor model



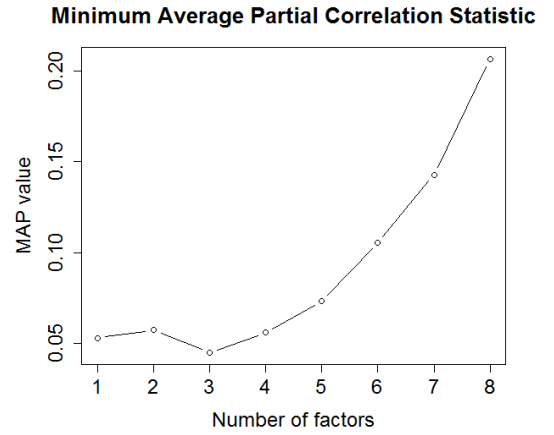
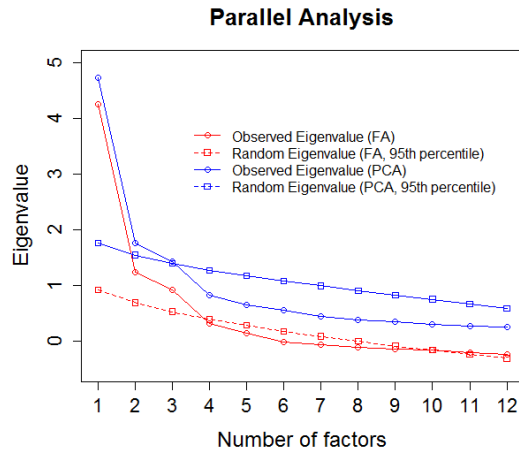
**5. Item descriptive statistics for the Italian Dirty Dozen Items in all samples of Study 1**

Item	Min	Max	M	SD	SK	KU
Sample 1 ( <i>n</i> =102)						
DD01	1	7	2.52	1.88	1.00	-0.21
DD02	1	7	2.34	1.72	1.14	0.26
DD03	1	7	2.25	1.78	1.33	0.59
DD04	1	7	1.93	1.42	1.51	1.46
DD05	1	7	2.50	2.00	1.07	-0.19
DD06	1	7	2.10	1.63	1.47	1.20
DD07	1	7	2.38	1.72	1.14	0.35
DD08	1	7	2.59	1.88	0.97	-0.19
DD09	1	7	3.41	2.09	0.32	-1.17
DD10	1	7	3.49	2.02	0.26	-1.14
DD11	1	7	2.99	1.98	0.53	-1.08
DD12	1	7	3.05	2.12	0.62	-1.03
Sample 2 ( <i>n</i> =128)						
DD01	1	7	2.27	1.49	1.12	0.43
DD02	1	7	2.23	1.46	1.37	1.42
DD03	1	7	2.21	1.43	1.17	0.74
DD04	1	7	1.86	1.30	1.81	2.92
DD05	1	7	2.67	1.53	0.56	-0.45
DD06	1	7	2.30	1.54	1.03	0.18
DD07	1	7	2.41	1.53	0.77	-0.44
DD08	1	7	2.66	1.56	0.56	-0.65
DD09	1	7	3.72	1.80	0.03	-0.95
DD10	1	7	3.59	1.72	0.09	-0.94
DD11	1	7	3.17	1.68	0.38	-0.72
DD12	1	7	3.07	1.80	0.54	-0.72
Sample 3 ( <i>n</i> =305)						
DD01	1	7	2.14	1.76	1.36	0.53
DD02	1	7	1.97	1.62	1.60	1.45
DD03	1	7	2.05	1.62	1.44	1.02
DD04	1	7	1.66	1.36	2.12	3.60
DD05	1	7	2.40	1.96	1.18	0.02
DD06	1	7	1.95	1.74	1.85	2.21
DD07	1	7	2.09	1.67	1.41	0.80
DD08	1	7	2.18	1.83	1.44	0.83
DD09	1	7	3.32	2.07	0.37	-1.24
DD10	1	7	3.17	1.91	0.38	-1.10
DD11	1	7	2.67	1.91	0.84	-0.57
DD12	1	7	2.57	1.88	0.88	-0.53

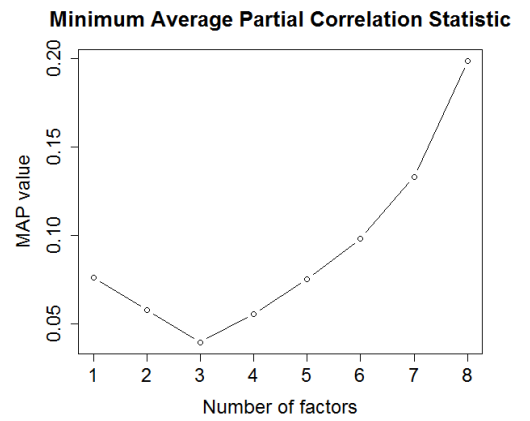
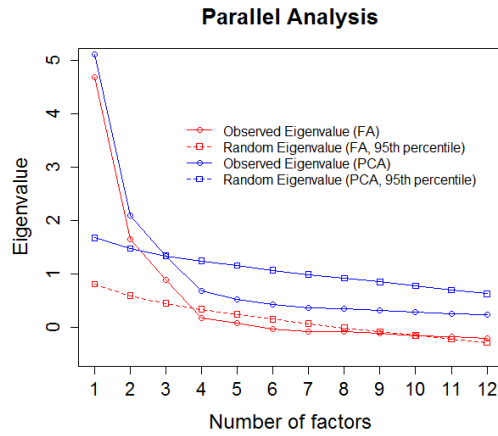
*Note:* Min = minimum; Max = maximum; M = mean; SD =standard deviation; SK = Skewness; KU = Kurtosis

**6. Analysis of the dimensionality of the item pool of the Italian Dirty Dozen in two independent samples of community participants (Study 1). FA = Factor Analysis; PCA = Principal Component Analysis**

Sample 1  
(n=102)



Sample 2  
(n=128)



**7. Parameter estimates in the 3-correlated-factor Confirmatory Factor Analysis (CFA) model (n = 305), Study 1**

	CFA			
	M	P	N	RV
DD01	0.69***	0.00	0.00	.52***
DD02	0.83***	0.00	0.00	.31***
DD03	0.79***	0.00	0.00	.38***
DD04	0.77***	0.00	0.00	.41***
DD05	0.00	0.42***	0.00	.83***
DD06	0.00	0.51***	0.00	.74***
DD07	0.00	0.84***	0.00	.29***
DD08	0.00	0.81***	0.00	.34***
DD09	0.00	0.00	0.73***	.47***
DD10	0.00	0.00	0.75***	.44***
DD11	0.00	0.00	0.66***	.57***
DD12	0.00	0.00	0.73***	.47***

*r* with P 0.53\*\*\*

*r* with N 0.60\*\*\* 0.39\*\*\*

Note: M = Machiavellianism; P = Psychopathy; N = Narcissism; RV = Residual Variance; *r* = Pearson's correlation; \* =  $p < .05$ ; \*\* =  $p < .01$ ; \*\*\* =  $p < .001$ ;

**8. Results of item analyses on the Italian Dirty Dozen in Study 1**

Statistic	Sample 1 ( <i>n</i> =102)	Sample 2 ( <i>n</i> =128)	Sample 3 ( <i>n</i> =305)
$\alpha$			
M	.85	.86	.84
P	.80	.82	.73
N	.78	.88	.81
$Mr_{ii}$ (range)			
M	.59 (.53-.67)	.62 (.53-.69)	.59 (.50-.70)
P	.50 (.36-.61)	.53 (.32-.66)	.42 (.29-.70)
N	.47 (.41-.59)	.64 (.60-.68)	.51 (.42-.61)
$Mr_{it}$ (range)			
M	.69 (.64-.74)	.71 (.67-.78)	.69 (.63-.74)
P	.62 (.57-.66)	.64 (.53-.75)	.53 (.40-.62)
N	.59 (.52-.65)	.73 (.72-.76)	.63 (.58-.66)
$M_{SMC}$ (range)			
M	.50 (.41-.57)	.53 (.50-.62)	.50 (.41-.57)
P	.43 (.39-.48)	.47 (.37-.58)	.36 (.18-.52)
N	.37 (.28-.46)	.54 (.52-.58)	.41 (.36-.46)
$\alpha$ w/o (higher)			
M	.83	.84	.84
P	.77	.81	.72
N	.76	.85	.78
Scale score <i>rs</i>			
M with P	.46***	.45***	.46***
M with N	.42***	.50***	.50***
N with P	.31**	.26**	.33***

Note:  $\alpha$  = Cronbach's alpha;  $Mr_{ii}$  = Mean inter-item correlation;  $Mr_{it}$  = Mean corrected item-total correlation;  $M_{SMC}$  = Mean squared multiple correlation;  $\alpha$  w/o = alpha-if-item-deleted index; *r*: Pearson correlation; \* =  $p < .05$ ; \*\* =  $p < .01$ ; \*\*\* =  $p < .001$ .



**9. Cronbach's alpha, descriptive statistics and intraclass correlation coefficients for Italian Dirty Dozen observed scale scores in Study 2 (n = 164)**

Scale	Time 1			Time 2			<i>p</i>	ICC
	M	SD	$\alpha$	M	SD	$\alpha$		
M	2.52	1.26	.84	2.60	1.35	.88	.209	.83
P	2.38	1.28	.81	2.38	1.34	.83	.955	.86
N	3.59	1.36	.80	3.55	.139	.83	.533	.83

*Note:* M = Mean; SD = Standard deviation;  $\alpha$  = Cronbach's alpha; M = Machiavellianism; P = Psychopathy; N = Narcissism; *p*: p-value of the paired-sample t-test (df=163); ICC = intraclass correlation coefficient

## 10. Socio-demographic differences between women and men in the sample of the paper

We tested whether the gender subgroups of the paper differed with respect to demographical variables. As shown in the table, some significant differences were found, but effect sizes were at best in the small range, suggesting that the comparisons of mean scores on the DD could have negligibly been biased (if ever) by the confounding effect of these variables.

Variable	Category	Women (n=554)	Men (n=420)	p	ES
Age (M±DS)		35.60±13.19	37.57±13.17	.021 <sup>a</sup>	0.15 <sup>b</sup>
Years of education (M±DS)		14.08±3.35	13.69±3.39	.065 <sup>a</sup>	0.12 <sup>b</sup>
Marital Status (proportion)	Single	.28	.23	.061 <sup>c</sup>	.09 <sup>d</sup>
	Married/Living together	.23	.18		
	Divorced/Separated	.04	.02		
	Widow/er	.01*	<.01*		
Occupation (proportion)	Unoccupied	.07*	.02*	<.001 <sup>c</sup>	.19 <sup>d</sup>
	Employed	.28*	.25*		
	Professional	.05	.06		
	Student	.15*	.08*		
	Retired	.01*	.02*		

*Note:* p = p-value for the statistical test; ES = effect size; M=mean; SD=standard deviation; <sup>a</sup>: independent sample t-test p-value; <sup>b</sup>: Cohen's *d*; <sup>c</sup>: chi-square test for the independence of categorical variables p-value; <sup>d</sup>:Cramer's *V*; \*: column proportions statistically different ( $p < .05$ ) after Bonferroni correction for multiple comparisons. Cohen's *d* is considered as negligible if  $d < .20$ , small if  $.20 < d < .50$ , moderate if  $.50 < d < .80$ , large if  $d > .80$ . Cramer's *V* is considered as negligible if  $V < .10$ , small if  $.10 < V < .30$ , moderate if  $.30 < V < .50$ , large if  $V > .50$ .

## 11. Measurement invariance models

Measurement invariance is usually tested with a sequence of models that impose equality constraints on the model parameters. Following Meredith (1964, 1993), the sequence of invariance testing begins with a model of *configural invariance* (M0), that is, with no invariance of any parameter estimates (i.e., all parameters are freely estimated), such that only similarity of the overall pattern of parameters is evaluated. This model tests whether the same factor structure is maintained across groups. Note that this model does not require any estimated parameters to be the same, hence it cannot be considered an actual invariance model. However, its fit has to be evaluated in order to provide both a test of the ability of the *a priori* model to fit the data in each group without invariance constraints and a baseline for comparing the other models that do impose equality constraints on the parameter estimates across groups.

The first step in invariance testing is to impose equality on factor loadings, i.e., specify a *weak* (or *scalar*) *invariance* model (M1). If identical items have statistically equivalent loadings, then the identical items show the same (if factor variances are fixed to 1 or constrained to be equal) or proportional (if variances are unequal) amount of increase between women and men for the same amount of increase on the latent factor (i.e., equality of scaling units; Millsap, 2011). This invariance is a prerequisite to comparisons of latent variances or relations among latent constructs. However, this model does not allow a test of differences in latent factor means, since mean differences based on latent constructs must be reflected in each of the individual items used to infer the latent constructs. It must then be shown that not only factor loadings, but also item intercepts (i.e., mean scores of individual items) are invariant over groups (*strong* or *scalar invariance* model, M2). If factor loadings and item intercepts are invariant over groups, then at all points along the factor continuum the same level of the latent factor results in statistically equivalent average scores on identical items between groups. This means that changes in the latent factor means can legitimately be interpreted as changes in the latent constructs. However, in models with freely estimated item intercepts and freely estimated latent means are not identified. Hence, the latent means are constrained to be zero in one group and freely estimated in the second group. This means that the freely estimated latent mean and its statistical significance reflect the differences between the two groups (Sörbom, 1974).

If one wants to compare (manifest) scale scores across groups, then an equality constraint must be posed also on item residual (or unique) variances (*strict measurement invariance* model, M3). This model assumes that same items have similar amounts of residual variance for both groups. We also tested models in which latent factor variances (M4) and covariances (M5) were constrained to be invariant. If latent factor variances are equal, in this case it would indicate that women and men used the same range on the factor continuum to report their levels of machiavellianism, psychopathy and narcissism and that the same items have equal levels of precision (reliability) across groups. If covariances are also invariant, the correlation between the same factor pairs for one group is statistically equivalent to the correlation between the same factors pairs for the other group.

### References

- Meredith, W. (1964). Notes on factorial invariance. *Psychometrika*, 29(2), 177–185.  
doi:10.1007/BF02289699
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525–543. doi:10.1007/BF02294825
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York, USA: Taylor & Francis Ltd.
- Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology*, 27(2), 229–239.  
doi:10.1111/j.2044-8317.1974.tb00543.x

**12. Standardized coefficients from the configural invariance model**

Items	Women ( <i>n</i> = 554)				Men ( <i>n</i> = 420)			
	M	P	N	RV	M	P	N	RV
DD01	.75			.44	.74			.46
DD02	.76			.42	.76			.43
DD03	.73			.47	.73			.46
DD04	.78			.39	.79			.38
DD05		.48		.77		.54		.71
DD06		.57		.68		.65		.57
DD07		.84		.30		.81		.34
DD08		.77		.41		.78		.40
DD09			.77	.41			.77	.42
DD10			.80	.37			.77	.41
DD11			.73	.47			.62	.61
DD12			.75	.44			.70	.51
Correlation with P	.51				.57			
Correlation with N	.59	.41			.59	.37		
Factor score determinacy <sup>°</sup>	.93	.91	.93		.93	.91	.91	
Reliability (McDonald's $\omega$ ) <sup>°</sup>	.81	.82	.71		.74	.79	.67	

*Note.* M = Machiavellianism; P = Psychopathy; N = Narcissism; I = intercept; RV = residual variance. All parameters are significant at  $p < .001$ ; <sup>°</sup>: see SM for details.

