Psychological Topics 21 (2012), 455-472

Original scientific paper – UDC – 159.923.3.072

The Effects of Faking on the Construct Validity of Personality Questionnaires: A Direct Faking Measure Approach

Maša Tonković Grabovac, Željko Jerneić, Zvonimir Galić

Department of Psychology, Faculty of Humanities and Social Sciences, University of Zagreb, Croatia

Abstract

Some authors clearly showed that faking reduces the construct validity of personality questionnaires, whilst many others found no such effect. A possible explanation for mixed results could be searched for in a variety of methodological strategies in forming comparison groups supposed to differ in the level of faking: candidates vs. non-candidates; groups of individuals with "high" vs. "low" social desirability score; and groups given instructions to respond honestly vs. instructions to "fake good". All three strategies may be criticized for addressing the faking problem indirectly – assuming that comparison groups really differ in the level of response distortion, which might not be true. Therefore, in a within-subject design study we examined how faking affects the construct validity of personality inventories using a direct measure of faking. The results suggest that faking reduces the construct validity of personality questionnaires gradually – the effect was stronger in the subsample of participants who distorted their responses to a greater extent.

Keywords: faking, personality questionnaire, construct validity, direct measure of faking

Introduction

After decades of research conducted in simulated and real selection contexts, it is now clear that people can distort their responses on personality scales and that applicants in personnel selection do that in order to present themselves in a positive light (e.g. Donovan, Dwight, & Hurtz, 2003; Galić & Jerneić, 2006; Mueller-Hanson, Heggestad, & Thornton III, 2003; Rosse, Stecher, Miller, & Levin, 1998). Further research has shown that, due to differences in ability and motivation to fake, not all people distort their answers to the same extent (Griffith, Chmielowski, & Yoshita, 2007; Jerneić et al., 2010; Snell, Sydell, & Lueke, 1999). Hence,

✉ Maša Tonković Grabovac, Department of Psychology, Faculty of Humanities and Social Sciences, Ivana Lučića 3, 10000 Zagreb, Croatia. E-mail: *masa.tonkovic@ffzg.hr*.

individual differences in faking could contaminate personality scores and distort the personality-based rank order of applicants – "people who inflate scores by even a moderate degree can displace honest candidates at the top of the pool" (Kim, 2011, p. 258). Consequences of this distortion are lower quality personnel decisions (Rosse et al., 1998; Stewart, Darnold, Zimmerman, Parks, & Dustin, 2010), and potentially a decrease of personality inventory properties such as construct and criterion validity (e.g. Tett & Christiansen, 2007).

The present study focuses on the effect of faking on the construct validity of personality questionnaires. This issue was investigated in numerous studies typically comparing personality inventory responses of those responding honestly with those faking (Ellingson, Smith, & Sackett, 2001). However, studies have been yielding contradictory results. In some, the construct validity was preserved (e.g. Ellingson et al., 2001; Henry & Raju, 2006; Marshall, De Fruyt, Rolland, & Bagby, 2005; Robie, Zickar, & Schmit, 2001) whilst in others it was seriously affected (e.g. Biderman & Nguyen, 2006; Schmit & Ryan, 1993; Zickar & Robie, 1999). A possible explanation for mixed findings could be searched for in the shortcomings of methodological strategies in forming comparative groups which are supposed to differ in the level of response distortion.

Three main comparison strategies can be identified in previous studies. One strategy was to compare job applicants' responses to those of incumbents (Marshall et al., 2005; Smith, Hanges, & Dickson, 2001; Weekley, Ployhart, & Harold, 2004), or of other non-applicants, such as students (Schmit & Ryan, 1993; Smith & Ellingson, 2002; Stark, Chernyshenko, Chan, Lee, & Drasgow, 2001). However, studies comparing these naturally occurring groups yielded inconsistent results. For example, Schmit and Ryan (1993) examined students' and applicants' responses on NEO-FFI and found that the five-factor structure fit the students' responses but not the applicants' responses. A six-factor model, with additional "ideal employee factor", fit the applicant data significantly better. On the other hand, Smith et al. (2001), comparing the NEO PI-R structure between job applicants, job incumbents and students, found that faking did not reduce the construct validity of personality questionnaires - the NEO PI-R factor structure remained stable across all samples. Such inconsistent findings regarding the effects of faking on construct validity of personality questionnaires could be attributed to the wrong assumption that all applicants fake, as well as that incumbents and other non-applicants do not. Incumbents could also be motivated to present themselves in a more favorable light, in order to get a promotion or some other benefit. In addition, they represent a preselected population in respect to their personality and abilities, thus the level of response distortion might not be the only relevant difference between the two comparison groups.

Another approach was to split a sample into "high"- and "low"-fakers groups on the basis of social desirability scale scores. For example, Henry and Raju (2006) examined the item-level and scale-level responses on six empirically-derived facets of conscientiousness from the California Psychological Inventory between "high" and "low" social desirability groups. Measurement equivalence across the two groups was tested by using the Item response theory based DFIT (*differential functioning of items and tests*) framework. Since analyses revealed that DFIT was present in negligible amount, Henry and Raju concluded that faking did not affect the construct validity of personality questionnaires. The same conclusion was drawn in many other studies using scores on a socially desirable responding scale as a basis for comparison groups formation (Ellingson et al., 2001; Flanagan & Raju, 1997; Marshall et al., 2005; Stark et al., 2001). However, Griffith and Peterson (2008) argued that using social desirability scores in studies on effects of faking was not valid, because "…little empirical evidence had supported the linkage between social desirability measures and actual faking behavior" (p. 308). For example, Galić and Jerneić (in press) concluded that at the individual level social desirability scores did not represent a good measure of faking, and Ellingson, Sackett and Hough (1999) showed that a social desirability-based correction of faked results failed to produce a corrected score that approximated an honest score.

The third, and perhaps the most common, strategy in forming comparison groups was to manipulate instruction sets using either between- (Douglas, McDaniel, & Snell; 1996; Griffith, 1997) or within-group designs (Ellingson et al., 1999). Regularly, the baseline instruction set was instructing participants to "respond honestly", while response distortion was encouraged by instructions to "fake good", i.e. present themselves as an extremely virtuous person (Douglas et al., 1996; Griffith, 1997; Pauls & Crost, 2005). This strategy was used by Biderman and Nguyen (2006) when modeling faking via structural equations models. The results strongly suggested that response tendencies that might be best characterized as faking changed the factor structure of the test scores completed under "fake-good" conditions. The main critique of this approach is that response distortion elicited by instructions does not reflect the real-life applicants' faking. Real-life applicants might fake their responses only on the personality dimensions relevant for the specific job (Martin, Bowen, & Hunt, 2002; Pauls & Crost, 2005), compared to raising their results on the whole inventory when instructed to "fake good".

Taken together, all three strategies can be criticized for addressing the faking problem indirectly, only assuming that comparative groups really differ in the level of response distortion. This inference might not be justified. Two respondents can be tested in different conditions (selection vs. non-selection), have unequal social desirability scale results ("high" vs. "low"), as well as having been given different instructions ("respond honestly" vs. "fake-good"), and still distort answers to the same extent. Thus, the aim of the present study was to examine how faking affects the construct validity of personality inventories using a direct measure of faking. We calculated the direct measure of faking as an individual difference score between two different instructional sets in the within-group research design. One instructional set was designed to elicit as little distortion as possible ("honest" responses), while the other instructional set was designed to elicit distortion comparable to that in selection situations. To make the faking condition more realistic, we used simulated selection instead of "fake-good" instructions. Accordingly, participants responded to a personality questionnaire twice – the first time they were asked to respond honestly and the second time to respond imagining that they were in a process of a specific selection program. The difference between scores on the personality questionnaire, filled in by the same respondent on two occasions, directly represented the individual amount of faking. This was not the first time in faking literature that the difference score was used to represent faking. It had been used in several studies on determinants of faking behavior (e.g. McFarland & Ryan, 2000). However, as far as we know, difference scores had never been used when exploring the effect of faking on construct validity. Thus, the novelty of this research is a comparison of personality responses of empirically detected "high"-fakers and "low"-fakers, instead of comparing groups only supposing to differ in the level of response distortion.

In order to test the effect of faking on the construct validity of personality questionnaires, we examined the factorial validity of the Five-factor model inventory IPIP-300 (*International Personality Item Pool*, Goldberg et al., 2006) by comparing the factor structure of its facets in groups and contexts in which the degree of faking varied. Based on the findings of some previous research on the effect of faking on factorial validity of personality questionnaires (Biderman & Nguyen, 2006; Cellar, Miller, Doverspike, & Klawsky, 1996; Ellingson et al., 1999; Schmit & Ryan, 1993; Ziegler & Buehner, 2009), we put forward the following hypotheses:

Hypothesis 1. Factorial validity in simulated selection condition will be reduced compared to factorial validity of the same personality questionnaire in condition when participants were instructed to respond honestly. The factor structure distortion will be manifested in a different arrangement of factor loadings, in comparison to the expected structure of five factors loading on their lower-ordered theoretically corresponding facets.

Hypothesis 2. The effect of factor structure distortion will be more pronounced in the subsample of participants who faked their responses to a greater extent. Thus, we expect that in "Applicant" condition the theoretical Five-factor model structure will be more distorted in the "High"-fakers subsample than in the "Low"-fakers subsample.

Method

Participants

Participants were 412 students (91% of the sample) and newly graduated alumni (9%) of the University of Zagreb, Croatia. Participants' age ranged from 18 to 30 years, averaging 23.0 (*SD*=2.17); 56% of the sample was female. Psychology students and alumni were excluded.

Measures and Procedures

We used the Croatian translation of Goldberg's IPIP-300 questionnaire (Jerneić, Galić, & Parmač, 2007) to measure five personality dimensions: Neuroticism, Extraversion, Openness, Agreeableness and Conscientiousness. Every personality dimension consists of six facets (see Table 2 for the list of facets), which were represented with 10 items each. Responses were made on a 7-point scale from *strongly disagree* to *strongly agree*. Participants anonymously responded to the IPIP-300 twice. The first time they were asked to respond honestly while the second time they were instructed to respond to the questionnaire as if they were in the process of a specific student scholarship selection program – a random half of the participants were imagining a selection for an abundant financial grant and the other half a selection for an attractive free student educational excursion. Since we found no differences in factor structures of responses collected under different programs, the data from both simulated selection programs were merged together to form "Applicant" condition.

Scale reliabilities of all five dimensions were high, with Alpha coefficients varying between .88 for Openness and .94 for Conscientiousness in "Honest" condition, and between .86 and .96 for the same two traits in "Applicant" condition. At the subscale level, most facets showed acceptable levels of internal consistency (>.70) across conditions. Only for Self-consciousness in both conditions (.59 and .52, respectively), and Activity (.55) and Openness to Values (.59) in "Applicant" condition, Alpha coefficients below .60 were obtained.

Statistical Methods and Analyses

Since we were interested in examining differential effects of the intensity of faking on factor structure of personality questionnaire, we divided participants into two subsamples according to the individual levels of faking. The faking score for each individual was calculated as the difference between personality scores in "fake good" and honest conditions. To get only one measure of faking, we summed up the absolute values of scale differences on five dimensions. The internal consistency of this composite measure was .69. Based on the median difference score, the sample was split into groups of "Low"- (N=207) and "High"-fakers (N=205).

The effects of faking were tested by a series of exploratory factor analyses. The choice of exploratory factor analysis over the confirmatory factor analysis was based on the previous research conclusions that confirmatory factor analysis seems to be too restrictive in testing the Five-factor model of personality (Borkenau & Ostendorf, 1990; Church & Burke, 1994, McCrae, Zonderman, Costa, Bond, & Paunonen, 1996). The exploratory analyses were conducted on both subsamples for each condition separately. The principal axis factor method (with varimax rotation) was applied to 30 facets by restricting the number of factors to five. We decided to fix the number of factors to five because of the Five-factor model background of the IPIP-300, which was in previous research empirically shown to be robust with respect to testing context (e.g. Johnson, 2005; McCrae et al., 1996). Another reason for fixing the number of factors was to enable a direct comparison of the similarity between factor structures as well as factors extracted in different subsamples and conditions, by using Tucker's (1951) congruence coefficients (r_c). We chose to factorize the facets instead of questionnaire items, because facets, as more reliable measures, should include a relatively lower amount of error variance.

Results

Group Mean Differences

In order to evaluate the effects of faking instructions, we computed descriptive statistics for "Honest" and "Applicant" conditions (Table 1). In both subsamples, the mean scores on Neuroticism in the "Applicant" conditions were lower compared to "Honest" condition, while means of other dimensions were generally higher compared to "Honest" condition. These results showed that in the "Applicant" condition, participants distorted their results in the socially desirable direction. The average effect sizes (*d*-indices, Cohen, 1988) were found in both subsamples – 0.97 in the "Low"-fakers subsample and 2.30 in the "High"-fakers subsample. Even though the average effect size was more than one standard deviation larger for "High"-fakers, the pattern of *d*-indices remained the same as in the "Low"-fakers subsample, with scores on Conscientiousness and Neuroticism being the most distorted. Within both groups, the pattern of results was similar to meta-analysis of fakability estimates (Viswesvaran & Ones, 1999) – the largest effect sizes were observed on Neuroticism and Conscientiousness and the lowest on Agreeableness and Openness.

Table 1. *Means, Standard Deviations and Effect Sizes for IPIP Dimensions, for Subsamples of "Low"- and "High"-Fakers in "Honest" and "Applicant" Conditions.*

Scale	"Low"-fakers (*N*=207)					"High"-fakers (*N*=205)				
	"Honest" Condition		"Applicant" Condition		"Honest" vs. "Applicant" Condition	"Honest" Condition		"Applicant" Condition		"Honest" vs. "Applicant" Condition
	M	*SD*	*M*	*SD*	Cohen's *d*	*M*	*SD*	*M*	*SD*	Cohen's *d*
Neuroticism	202.7	37.89	146.6	35.70	1.52	232.9	39.49	121.5	23.12	3.44
Extraversion	291.3	34.01	325.2	30.29	-1.05	268.0	37.18	341.3	25.24	-2.31
Openness	301.0	32.41	309.5	29.00	-0.28	296.9	36.10	314.7	26.31	-0.56
Agreeableness	302.2	30.82	311.8	32.18	-0.30	288.1	35.55	320.2	31.59	-0.95
Conscientiousness	301.7	36.07	365.2	38.62	-1.70	262.0	38.08	392.4	20.84	-4.25

Factorial Validity

Following the procedure explained in the Method section, we first factor-analyzed the facets scores in "Honest" condition by fixing the number of factors to five. The expected five factor solution was supported by the Scree test (Cattell, 1966) and parallel analysis (Horn, 1965). Tucker's congruence coefficient values of .95 and higher indicated that the five factors extracted on "Low"-fakers were highly similar to the corresponding five factors on the "High"-fakers subsample. Therefore, in this condition we merged the two subsamples and performed factor-analysis on the whole sample. The extracted factors explained 7-14% (cumulative 58%) of the total variance of facets scores (Table 2). The factor loadings suggested that the five-factor structure of the IPIP-300 was reproduced. Even though the simple structure, as expected by the Five-factor model of personality, was not found, the amount of cross-loadings was comparable to those found in other factor-analytic studies on the same inventory (e.g. Johnson, 2005; Mottus, Pullmann, & Allik, 2006).

Table 2. *Results of Factor Analysis of IPIP-300 Facets in "Honest" Condition ("Low"- and "High"-Fakers Together) (N=412)*

Scale	Factor					h^2
	N	*E*	*O*	*A*	*C*	
N1 Anxiety	**.88**	-.25	-.03	.01	-.04	.83
N2 Angry Hostility	**.72**	-.09	-.07	-.30	.14	.64
N3 Depression	**.60**	**-.42**	.15	-.12	-.22	.62
N4 Self-Consciousness	**.61**	**-.33**	-.11	.26	**-.41**	.73
N5 Impulsiveness	**.39**	.23	.14	**-.51**	-.25	.55
N6 Vulnerability	**.82**	-.02	-.03	.01	-.27	.75
E1 Warmth	-.13	**.81**	.01	.08	.30	.76
E2 Gregariousness	-.04	**.73**	-.02	-.01	.20	.58
E3 Assertiveness	-.10	**.37**	.14	-.21	**.66**	.65
E4 Activity	-.01	.18	-.00	.01	**.69**	.50
E5 Excitement-Seeking	-.14	**.37**	**.37**	**-.35**	-.05	.42
E6 Positive Emotions	-.18	**.73**	.08	.02	.18	.61
O1 Fantasy	**.37**	.09	**.51**	-.10	-.07	.42
O2 Aesthetics	.16	.24	**.45**	**.40**	.15	.47
O3 Feelings	**.47**	.20	**.50**	.21	.20	.60
O4 Actions	**-.32**	.27	**.50**	.01	.15	.45
O5 Ideas	-.12	.00	**.56**	.00	**.42**	.51
O6 Values	-.05	-.13	**.47**	-.04	-.13	.26
A1 Trust	-.16	**.54**	.11	**.33**	-.06	.44
A2 Straightforwardness	-.01	.07	.02	**.76**	.05	.59
A3 Altruism	.05	**.51**	.21	**.58**	.22	.70
A4 Compliance	-.22	.13	.00	**.73**	-.23	.65
A5 Modesty	.11	-.23	-.06	**.45**	**-.40**	.44
A6 Tender-Mindedness	.14	.29	.27	**.61**	.01	.56
C1 Competence	**-.41**	.15	.12	.07	**.70**	.70
C2 Order	.10	-.00	-.21	**.39**	**.39**	.36
C3 Dutifulness	-.12	.05	-.13	**.74**	**.34**	.70
C4 Achievement Striving	.00	.11	.08	.18	**.77**	.65
C5 Self-Discipline	-.16	.14	-.07	**.37**	**.71**	.69
C6 Deliberation	-.20	-.23	-.20	**.53**	**.41**	.59
% of Explained Variance	12.70	11.60	6.60	13.80	13.40	58.00

Note. These are Varimax-rotated principal axes, extracted when the number of factors was fixed to 5. Loadings greater than .30 in absolute magnitude are in boldface. N = Neuroticism; E = Extraversion; O = Openness; A = Agreeableness; C = Conscientiousness; h^2 = Communalities.

In the next step we factor-analyzed the facets scores in the "Applicant" condition for each subsample separately (Table 3). To have a comparable number of factors with the "Honest" condition, we retained the five factors albeit results of parallel analysis and scree plot suggested discarding the fifth factor in both subsamples. In the subsample of "Low"-fakers, five extracted factors explained 65% of the total variance of 30 IPIP facets – seven percent more than in the "Honest" condition. Further, in this condition the amount of variance explained by each factor (4-21%) varied more. Factor matrix with a greater number of cross-loadings revealed that a more complex solution was produced, compared to the five factor structure in the "Honest" condition (Table 2).

Table 3. *Results of Factor Analysis of IPIP-300 Facets for "Low"-Fakers (N=207) and "High"-Fakers (N=205) in "Applicant" Condition*

Scale	"Low"-fakers						"High"-fakers					
	Factor					h^2	Factor					h^2
	1_L	2_L	3_L	4_L	5_L		1_H	2_H	3_H	4_H	5_H	
N1 Anxiety	**-.34**	-.20	**.76**	-.22	.06	.78	**-.41**	-.27	**-.45**	.28	.11	.54
N2 Angry/ Hostility	-.21	-.29	**.67**	-.07	.05	.58	**-.33**	**-.37**	-.19	**.38**	-.10	.43
N3 Depression	**-.41**	**-.40**	**.57**	-.20	.13	.70	**-.67**	-.24	**-.33**	.14	-.04	.63
N4 Self-Consciousness	**-.41**	-.04	**.71**	-.14	-.03	.69	**-.50**	.05	**-.36**	.14	.03	.40
N5 Impulsiveness	**-.52**	-.26	**.45**	.26	.12	.63	**-.47**	**-.36**	-.02	**.40**	-.09	.52
N6 Vulnerability	**-.55**	-.06	**.72**	-.01	-.00	.82	**-.62**	-.01	-.23	**.37**	-.01	.57
E1 Warmth	**.33**	**.66**	-.27	**.41**	-.04	.76	.22	**.56**	**.61**	.02	-.00	.73
E2 Gregariousness	.24	**.45**	-.29	**.53**	.03	.62	.14	.27	**.68**	.06	.09	.57
E3 Assertiveness	**.71**	.04	-.26	.30	.05	.67	**.57**	-.18	**.34**	.16	.10	.51
E4 Activity	**.63**	-.11	-.26	-.09	-.06	.49	**.47**	-.15	.03	-.21	**.44**	.48
E5 Excitement-Seeking	-.06	.04	-.11	**.58**	.14	.37	-.10	-.23	**.70**	**.30**	-.09	.65
E6 Positive Emotions	.20	**.58**	-.15	**.59**	.06	.75	.17	.23	**.61**	**.32**	-.06	.56
O1 Fantasy	-.06	.18	**.32**	**.34**	**.48**	.48	-.18	-.01	.23	**.72**	-.05	.61
O2 Aesthetics	.27	**.59**	-.03	.25	.15	.50	.26	**.38**	**.53**	.23	.12	.56
O3 Feelings	.07	**.48**	**.32**	.18	**.53**	.65	-.03	.24	.21	**.72**	.06	.63
O4 Actions	.28	**.34**	**-.48**	.26	.17	.53	.22	.19	**.59**	-.08	.20	.48
O5 Ideas	**.53**	**.32**	**-.34**	.16	**.40**	.68	**.45**	.17	**.38**	.17	**.32**	.50
O6 Values	-.23	.04	-.08	.02	**.44**	.25	**-.37**	.13	.26	-.00	.26	.28
A1 Trust	.00	**.71**	-.22	.17	.13	.60	.11	**.57**	**.31**	.08	.01	.44
A2 Straight-forwardness	.24	**.75**	-.14	-.19	-.06	.67	.11	**.72**	.10	-.08	.04	.55
A3 Altruism	.27	**.81**	-.14	.18	.18	.81	.23	**.63**	**.42**	.17	.13	.68
A4 Compliance	-.02	**.76**	-.21	-.06	-.07	.63	-.07	**.77**	.01	-.04	**-.32**	.71
A5 Modesty	**-.50**	.26	.06	**-.33**	.02	.43	**-.44**	**.49**	**-.31**	-.03	.13	.54
A6 Tender-Mindedness	-.13	**.71**	.04	.06	**.36**	.65	-.16	**.67**	.16	.23	**.35**	.67
C1 Competence	**.76**	.23	**-.40**	.10	-.00	.80	**.73**	.10	.29	-.02	-.03	.63
C2 Order	**.71**	.29	-.03	-.02	-.27	.66	**.74**	.06	-.08	-.03	-.14	.58
C3 Dutifulness	**.58**	**.65**	-.27	-.02	-.06	.83	**.48**	**.63**	.11	.02	-.08	.64
C4 Achievement/ Striving	**.80**	.19	-.27	.11	-.08	.77	**.66**	.02	.08	-.05	.17	.47
C5 Self-Discipline	**.80**	.24	**-.34**	.09	-.08	.82	**.76**	.26	.15	-.06	.10	.68
C6 Deliberation	**.73**	**.34**	-.26	-.15	-.07	.74	**.66**	**.34**	-.05	-.17	.07	.58
% of Explained Variance	21.00	19.30	13.60	6.60	4.20	64.60	19.20	14.50	12.70	7.00	2.70	56.00

Note. These are Varimax-rotated principal axes, extracted when the number of factors was fixed to 5. Loadings greater than .30 in absolute magnitude are in boldface. N = Neuroticism; E = Extraversion; O = Openness; A = Agreeableness; C = Conscientiousness; h^2 = Communalities.

The first extracted factor in the "Low"-fakers subsample showed good similarity to Conscientiousness in the "Honest" condition (r_c=.96; Table 4). Also, it saturated Extraversion and Neuroticism facets which showed the greatest response distortion effect. The second factor resembled the Agreeableness factor the most, but Tucker's congruence coefficient was below .85, which is considered a threshold for fair similarity (Lorenzo-Seva & ten Berge, 2006). Based on factor loadings, the third factor most resembled Neuroticism from honest responding condition, although the congruence coefficient value of .93 indicated that similarity was only moderate. The remaining two factors, with only seven and four percent of explained variance, were most similar to Extraversion and Openness, respectively. Even though the fifth factor was not supported by parallel analysis, the congruence coefficient of both factors reached the threshold of fair similarity with factors extracted in the "Honest" condition. The total matrix congruence coefficient, comparing similarity between whole matrices (McCrae et al., 1996) equaled .89, which indicated only fair similarity. Hence, in the subsample of "Low"-fakers we confirmed our first hypothesis – faking in the "Applicant" condition to a certain degree changed the construct validity of IPIP-300 inventory.

Table 4. *Tucker's Congruence Coefficients Expressing Correlations between Factors Extracted in "Honest" (N=412) and "Applicant" Condition in "Low"-Fakers Subsample (N=207)*

		"Applicant" Condition				
		1_L	2_L	3_L	4_L	5_L
"Honest" Condition	N	-.51	-.24	**.93**	-.20	.25
	E	.39	.66	-.46	**.85**	.22
	O	.08	.36	-.09	.54	**.89**
	A	.38	**.80**	-.27	-.15	-.01
	C	**.96**	.36	-.58	.34	-.03

Note. L = Low fakers; N = Neuroticism; E = Extraversion; O = Openness; A = Agreeableness; C = Conscientiousness. The highest values for each of the factors extracted in "Applicant" condition are in boldface.

In the subsample of "High"-fakers, the five extracted factors in the "Applicant" condition explained 56% of IPIP-300 facets variance, almost the same amount as in the "Honest" condition (58%). Similarly to the subsample of "Low"-fakers in the same condition, the first factor explained much more variance (19%) than the fifth (3%). The pattern of factor loadings (Table 3) did not resemble the five-factor structure found in the "Honest" condition (Table 2). Indeed, the matrix of Tucker's coefficients did not reveal clear comparative factors (Table 5).

Table 5. *Tucker's Congruence Coefficients Expressing Correlations between Factors Extracted in "Honest" (N=412) and "Applicant" Condition in "High"-Fakers Subsample (N=205)*

		"Applicant" Condition				
		1_H	2_H	3_H	4_H	5_H
"Honest" Condition	N	-.61	-.26	-.44	**.63**	.01
	E	.38	.46	**.86**	.29	.14
	O	-.01	.20	.61	.59	**.44**
	A	.35	**.89**	.12	-.07	.16
	C	**.89**	.18	.46	-.06	.40

Note. H = High fakers; N = Neuroticism; E = Extraversion; O = Openness; A = Agreeableness; C = Conscientiousness. The highest values for each of the factors extracted in "Applicant" condition are in boldface.

Factor 1 in the "Applicant" condition was the most similar factor (.89) to Conscientiousness in the "Honest" condition. Nevertheless, it was more similar (.98) to Factor 1 in the "Low"-fakers subsample in the "Applicant" condition, with almost the same pattern of factor loadings. It saturated all the Conscientiousness facets and negatively Neuroticism facets, together with Assertiveness, Activity, Openness to Ideas, and negatively Modesty. Factors 2 and 3 were most similar to Agreeableness (.89) and Extraversion (.86) factors in the "Honest" condition, respectively. Again, these factors were, in absolute terms, more similar to factors in the "Low"-fakers subsample in the "Applicant" condition: Factor 2 (.94) and Factor 4 (.89), respectively. Finally, the remaining two extracted factors could not be clearly interpreted, having saturated too few facets. The fourth factor saturated Openness to Feelings and Fantasy, and Impulsiveness, while the rejected fifth factor saturated Activity only. The deviation from the five-factor structure found in the "Honest" condition was clearly evidenced by the fact that Neuroticism and Openness factors did not reproduce. In addition, the total congruence coefficient, comparing similarity between factor matrices in "Honest" and "Applicant" conditions in the "High"-fakers subsample, equaled .75. This value was lower than the congruence coefficient value in the "Low"-fakers subsample (.89), and, in absolute terms, below the threshold of fair similarity. From these results we concluded not only that the factorial validity in the "Applicant" condition in the subsample of "High"-fakers was reduced, but also that this effect was greater compared to the effect in the subsample of "Low"-fakers. This finding confirmed both our hypotheses.

Discussion

In order to examine whether faking reduces the factorial validity of the personality questionnaire, we compared its factor structure in subsamples and conditions in which the degree of faking varied. We had hypothesized that factorial validity of a personality questionnaire in a simulated selection condition would be reduced in comparison to the condition when participants were instructed to respond honestly. In addition, we expected that distortion would be more pronounced in the subsample of participants who faked their responses to a greater extent. A within-subjects design enabled us to calculate a direct measure of faking, as a difference between scores on the same personality questionnaire filled in by the same respondent in simulated selection and honest conditions. The results of our analyses revealed that faking reduced the construct validity of the personality questionnaire. Moreover, the effect was stronger in the subsample of participants who faked their responses to a greater extent ("High"-fakers) in comparison to the "Low" fakers group. This was evident from two findings: first, within the "High"-fakers subsample there were more cross-loadings, resulting in some factors not being interpretable, and second, in the same subsample, Tucker's congruence coefficients showed that the factors structure was less similar to the expected five-factor structure.

The finding that faking changes the construct validity of personality questionnaires is inconsistent with some previous investigations examining the invariance of personality inventories factor structure in honest and faking conditions that showed that the factor structure was preserved in faking conditions (e.g. Ellingson et al., 2001; Henry & Raju, 2006; Marshall et al., 2005; Robie et al., 2001). However, in these studies the formation of comparison groups was based on indirect indicators of faking: social desirability scale scores, membership in certain groups (job applicants, incumbents, or other non-applicants), or motivational context (instruction to respond honestly or "fake-good"). The limitation of these strategies is the assumption that comparative groups really differ in the level of faking, which might not be true. For this reason, we introduced a strategy of forming comparison groups based on a direct measure of faking.

In our study, the baseline factor structure was the one found in the condition when participants were instructed to respond honestly. As expected, the factor-analyses of personality facets in this condition replicated a Five-factor model structure of the IPIP-300 questionnaire. However, the same analyses in simulated selection program revealed that the five-factor structure had been distorted. Even though the Tucker's congruence coefficients showed that some factors were similar to the five personality factors, the pattern of factor loadings offered potential alternative interpretations. In both "High"- and "Low"-fakers subsamples, the two largest factors might be seen as social desirability factors. The first described a competent, dominant, emotionally stable and intellectual person, resembling factors known in literature as "ideal-employee factor" (Schmit & Ryan, 1993), "agentic factor" (Wiggins & Trapnell, 1996) and "egoistic bias" (Paulhus, 2002). The second factor described an agreeable, warm and reliable person, which are characteristics of "Factor β" (Digman, 1997), "communal" factor (Wiggins & Trapnell, 1996) and "moralistic bias" (Paulhus, 2002). The finding that parallel analysis and scree test suggested retaining only four factors is in line with evidence that faking inflates the inter-correlations among trait dimensions (Pauls & Crost, 2005; Ziegler & Buehner, 2009), which often results in the extraction of fewer, social desirability, factors when compared to honest responders (Douglas et al., 1996; Ellingson et al., 1999).

The methodology of using instructions to experimentally induce faking behavior was often criticized for resulting in more response distortion in comparison to actual selection context, where job applicants fake more cautiously to prevent getting caught. In order to overcome this limitation and create a response set that is representative of the mind sets of actual applicants (Biderman & Nguyen, 2006), we used simulated selection instead of commonly used "fake-good" instructions. Furthermore, by dividing the sample based on the amount of faking, we managed to isolate "low"-fakers who distorted their answers to a lesser extent, potentially comparable to that obtained on actual applicants. Our results clearly showed that even relatively low amount of faking could reduce the construct validity of personality questionnaires. However, it is still possible that responses collected in a natural environment, such as an actual applicant condition, would yield different results. In order to enhance the external validity, i.e. generalizability of the present findings, future studies should consider replication of this research in other samples and conditions. Specifically, future research might compare responses collected in a condition where participants are actual selection candidates to the condition where selection program is simulated. The most challenging aspect of such within-subject research would be getting applicants' honest responses. Also, it would be interesting to examine potential moderator variables such as job need or desire, which could raise applicants' motivation to fake and therefore impact the effect on construct validity changes. In selection situations which elicit more motivation to fake (e.g. attractive job position), we would expect a more severe decrease in validity, compared to situations which elicit less motivation to fake and consequently less faking behavior.

In addition, it should be noted that all participants responded to the IPIP-300 twice and in the same order: the first time participants were asked to respond honestly while the second time they were instructed to respond to the questionnaire as if they were in the process of a specific student scholarship selection program. Hence, it might be questionable whether the effects could be attributed only to manipulation of the experimental condition and not to the order of the conditions. However, we believe that our conclusions regarding differences between honest responding and experimentally induced faking situations are valid. This is because the condition effect sizes (see Table 1 for Cohen *d*-values) on personality scales were robust in our analyses and are very well documented in the literature for within- and between-subject designs (e.g., Viswesvaran & Ones, 1999). Moreover, the order effect, if any, would be a small one and substantially outperformed by the condition effect size (Hough, Eaton, Dunnette, Kamp, & McCloy, 1990).

This study has a strong implication for the personnel selection practice. The finding that faking reduces the construct validity of the IPIP-300 personality questionnaire suggests that the Five-factor personality model may not fit applicant population responses as well as it fits honest responses. In support of this, there are many studies carried out in actual selection contexts that have found distorted construct validity of personality measures in applicants' samples (e.g. Cellar et al., 1996; Cox, 2010; Schmit & Ryan, 1993; Stark et al., 2001; Weekley et al., 2004). In addition, our results imply that the more the applicants fake, the more severe the decrease in construct validity could be. The most straightforward implication of our research is the awareness that by using the Five-factor model inventories in personnel selection context we might not measure constructs we intend to. Thus, descriptions of candidates' personality based on dimensions scale scores might not represent the degree to which individuals actually possess the personality traits being measured (Ellingson et al., 2001). To prevent lower quality personnel decisions, existing personality scales scores should be interpreted with caution.

References

Biderman, M.D., & Nguyen, N.T. (2006, February). *Measuring response distortion using structural equation models.* Paper presented at the conference New Directions in Psychological Measurement with Model-Based Approaches, Georgia Institute of Technology, Atlanta, GA, United States of America. Retrieved from http://web2.utc.edu/~pbv184/RecentPapers/Biderman,Nguyen_060216_Measuring%20response%20distortion%20using%20structural%20equation%20models_NDPsyMeasConf_2006.pdf

Borkenau, P., & Ostendorf, F. (1990). Comparing exploratory and confirmatory factor analysis: a study of the 5-factor model of personality. *Personality and Individual Differences, 11,* 515-524.

Cattell, R.B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1,* 245-276.

Cellar, D.F., Miller, M.L., Doverspike, D.D., & Klawsky, J.D. (1996). Comparison of factor structures and criterion-related validity coefficients for two measures of personality based on the five-factor model. *Journal of Applied Psychology, 81*, 694-704.

Church, A.T., & Burke, P.T. (1994). Exploratory and confirmatory tests of the Big Five and Tellegen's three- and four-dimensional models. *Journal of Personality and Social Psychology, 66,* 93-114.

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

Cox, B.D. (2010). *Differential functioning by high and low impression management groups on a Big Five applicant screening tool*. (Unpublished doctoral dissertation). Graduate Faculty of Auburn University, Auburn, Alabama, United States of America.

Digman, J.M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology*, *73*(6), 1246-1256.

Donovan, J.J., Dwight, S.A., & Hurtz, G.M. (2003). An assessment of the prevalence, severity and verifiability of entry-level applicant faking using the randomized response technique. *Human Performance, 16*, 81-106.

Douglas, E.F., McDaniel, M.A., & Snell, A.F. (1996, August). *The validity of non-cognitive measures decays when applicants fake.* Paper presented at the annual meeting of the Academy of Management, Cincinnati, Ohio, United States of America. Abstract retrieved from http://www.ipacweb.org/files/noncog.pdf

Ellingson, J.E., Sackett, P.R., & Hough, L.M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology, 84*, 155-166.

Ellingson, J.E., Smith, D.B., & Sackett, P.R. (2001). Investigating the influence of social desirability on personality factor structure. *Journal of Applied Psychology, 86*, 122-133.

Flanagan, W.J., & Raju, N.S. (1997, April). *Measurement equivalence between high and average impression management groups: An IRT analysis of personality dimensions.* Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, St. Louis.

Galić, Z., & Jerneić, Ž. (2006). Socijalno poželjno odgovaranje u različitim motivacijskim kontekstima: Provjera Paulhusova modela socijalne poželjnosti. [Socially desirable responding in different motivational contexts: Evaluation of the Paulhus' social desirability model]. *Suvremena psihologija, 9*(2), 155-170.

Galić, Z., & Jerneić, Ž. (in press). Usefulness of Communal and Agentic Management Scale for measuring faking on Five Factor personality questionnaires. *Personnel Psychology*.

Goldberg, L.R., Johnson, J.A., Eber, H.W., Hogan, R., Ashton, M.C., Cloninger, C.R., & Gough, H.C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*, 84-96.

Griffith, R. (1997). *Faking of non-cognitive selection devices: Red herring is hard to swallow*. (Unpublished doctoral dissertation). University of Akron, Akron, Ohio, United States of America.

Griffith, R.L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review, 36*, 341-355.

Griffith, R.L., & Peterson, M.H. (2008). The failure of social desirability measure to capture applicant faking behavior. *Industrial and Organizational Psychology, 1*, 308-311.

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

Horn, J.L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30,* 179-185.

Hough, L.M., Eaton, N.K., Dunnette, M.D., Kamp, J.D., & McCloy, R.A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75*(5), 581-595.

Jerneić, Ž, Galić, Z., & Parmač, M. (2007). *Croatian translation and adaptation of the International Personality Item Pool questionnaire - IPIP-300* (author L. R. Goldberg). Department of Psychology, Faculty of Humanities and Social Sciences, University of Zagreb.

Jerneić, Ž., Galić, Z., Parmač Kovačić, M., Tonković, M., Prevendar, T., Bogdanić, A., … Zlatar, A. (2010). *Ličnost i radno ponašanje: priroda i mjerenje socijalno poželjnog odgovaranja [Personality and work behavior: Nature and measurement of socially desirable responding]*. Zagreb: FF Press.

Johnson, J. (2005). Ascertaining the validity of individual protocols from Web-based personality inventories. *Journal of Research in Personality, 39,* 103-129.

Kim, B.H. (2011). Deception and applicant faking: Putting the pieces together. *International Review of Industrial and Organizational Psychology, 26,* 239-292.

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.

Marshall, M.B., De Fruyt, F., Rolland, J.P., & Bagby, R.M. (2005). Socially desirable responding and the factorial stability of the NEO PI-R. *Psychological Assessment, 17*, 379-384.

Martin, B.A., Bowen, C.C., & Hunt, S.T. (2002). How effective are people at faking on personality questionnaires? *Personality and Individual Differences, 32,* 247-256.

McCrae, R.R., Zonderman, A.B., Costa, P.T., Bond, M.H., & Paunonen, S.V. (1996). Evaluating replicability of factors in the Revised NEO Personality Inventory: Confirmatory Factor Analysis versus Procrustes Rotation. *Journal of Personality and Social Psychology, 70*(3), 552-566.

McFarland, L.A., & Ryan, A.M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology, 85*(5), 812-821.

Mottus, R., Pullmann, H., & Allik, J. (2006). Toward more readable Big Five personality inventories. *European Journal of Psychological Assessment, 22*(3), 149-157.

Mueller-Hanson, R., Heggestad, E.D., & Thornton III, G.C. (2003). Individual differences in impression management: An exploration of the psychological processes underlying faking. *Psychology Science, 48,* 288-312.

Paulhus, D.L. (2002). Socially desirable responding: The evolution of a construct. In H.I. Braun, D.N. Jackson, & D.E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49-69). Mahwah: NJ Erlbaum.

Pauls, C.A., & Crost, N.W. (2005). Effects of different instructional sets on the construct validity of the NEO-PI-R. *Personality and Individual Differences*, *39*, 297-308.

Robie, C., Zickar, M.J., & Schmit, M.J. (2001). Measurement equivalence between applicant and incumbent groups: An IRT analysis of personality scales. *Human Performance, 14,* 187-207.

Rosse, J.G., Stecher, M.D., Miller, J.L., & Levin, R.A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology, 83,* 634-644.

Schmit, M.J., & Ryan, A.M. (1993). The Big Five in personnel selection: Factor structure in applicant and nonapplicant populations. *Journal of Applied Psychology, 78,* 966-974.

Smith, D.B., & Ellingson, J.E. (2002). Substance versus style: A new look at social desirability in motivating contexts. *Journal of Applied Psychology, 87,* 211-219.

Smith, D.B., Hanges, P.J., & Dickson, M.W. (2001). Personnel selection and the five-factor model: Reexamining the effects of applicant's frame of reference. *Journal of Applied Psychology*, *86*, 304-315.

Snell, A.F., Sydell, E.J., & Lueke, S.B. (1999). Towards a theory of applicant faking: Integrating studies of deception. *Human Resource Management Review*, *9*, 219-242.

Stark, S., Chernyshenko, O.S., Chan, K., Lee, W., & Drasgow, F. (2001). Effects of the testing situation on item responding: Cause for concern. *Journal of Applied Psychology*, *86*, 943-953.

Stewart, G.L., Darnold, T.C., Zimmerman, R.D., Parks, L., & Dustin, S.L. (2010). Exploring how response distortion of personality measures affects individuals. *Personality and Individual Differences, 49*(6), 622-628.

Tett, R.P., & Christiansen, N.D. (2007). Personality tests at the crossroads: A response to Morgeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt. *Personnel Psychology*, *60*, 967-993.

Tucker, L.R. (1951). *A method for synthesis of factor analysis studies.* Personnel Research Section Report No. 984. Washington, DC: Department of the Army.

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

Weekley, J.A., Ployhart, R.E., & Harold, C.M. (2004). Personality and situational judgment tests across applicant and incumbent settings: An examination of validity, measurement, and subgroup differences. *Human Performance*, *17*, 433-461.

Wiggins, J.S., & Trapnell, P.D. (1996). A dyadic-interactional perspective on the Five Factor Model. In J.S. Wiggins (Ed.), *The Five Factor model of personality: Theoretical perspectives*, (pp. 88-162). New York: Guilford Press.

Zickar, M.J., & Robie, C. (1999). Modeling faking good on personality items: An item-level analysis. *Journal of Applied Psychology, 84*, 551-563.

Ziegler, M., & Buehner, M. (2009). Modeling socially desirable responding and its effects, *Educational and Psychological Measurement*, *69*(4), 548-565.

Received: February 27, 2012