Review of Psychology
2003, Vol. 10, No. 2, 65-74

UDC 159.9

Response-order effect in some self-reported personality questionnaires: The cases of 16PF and BFQ

VALENTIN BUCIK

Different personality questionnaires, in most cases belonging to different theoretical models, are usually compared with each other with the purpose of confirmation, or rejection, of their convergence to the specific theory or model of the structure of personality. Nevertheless, up to now only few studies investigated the stability of measurement with regard to different response styles or response sets, which self-reported personality questionnaires are usually liable to. It has been shown that the specific method factor - the response-order effect (ROE - positively scored or reverse scored items) subsequently distorts the results in 16PF. It was reasonable to expect that similar pattern would emerge in other questionnaires, which try to control answering strategies with positive-negative balance in items. The aim of this study was to examine comparatively the effect of this systematic method factor in two personality questionnaires, the Slovenian adaptations of 16PF and the Big Five Questionnaire. The latter, like 16PF, consists of half of positively stated and half of negatively stated items. The study was conducted on the homogeneous sample of 386 high-school participants and different proposed models were analysed with the confirmatory factor analysis. The results are discussed not only in terms of their convergence with the specific theoretical model, but rather with respect to the ascendance of the "positive" or "reversed" items in these instruments. The results indicate that the method factor (ROE) affects the answers in both questionnaires importantly, and that positively scored items represent the constructs more reliably.

Self-reported questionnaires are within most frequently used instruments for exploring the structure of personality. Different questionnaires are usually compared to each other on order to confirm, or reject, their convergence to specific model of the structure of personality. One should admit, however, that answering the questions about typical behaviour in such instruments often gives the final result that doesn't seem to represent very reliably the so-called true value of the characteristics of personality or "the true score" in terms of classical test theory (Cronbach, 1990). There are various reasons for this. Most frequently the result (or "the observed score") does not meet the true value because of (and in the amount of) measurement error(s), within which different cognitive and non-cognitive mediating factors can be found. Usually these factors have nothing in common with the true value of the personality trait of interest, but their influences are so strong that they should not be neglected in the process of measurement.

Valentin Bucik, University of Ljubljana, Department of Psychology, Aškerčeva 2, 1000 Ljubljana, Slovenia. E-mail: tine.bucik@ff.uni-lj.si. (Correspondence concerning this article should be sent to this address.)
Author wishes to thank the anonymous reviewers for insightful comments on the first draft of the manuscript.

There are several difficulties regarding answering questions in self-reported personality questionnaires. Psychologists usually point out different factors that influence the discrepancy between the true and the observed score in such instruments (see Cohen, Montague, Nathanson & Swerdlik, 1988; Cronbach, 1990; Lienert & Raatz, 1994; Murphy & Davidshofer, 1988). Among them are the following (and the list is far from being complete):

- Sampling the items in the final version of instrument. The set of items should be the best representation of the universe of all possible items that can be stated regarding the particular personality trait, but the instrument constructor never knows exactly how successful she/he was in sampling the final set of typical items in the questionnaire from that universe (Cronbach, 1990).
- Sampling the participants in the standardization study. The successfulness of the standardization of the instrument will depend on the quality of planning the structure of the sample (or samples) used for building the norms. The stratification of the basic sample, which is representative with regard to population, is tricky because one does not know exactly which standardization criteria are correlated with personality.

- The different interpretational value that each item in the questionnaire has for different respondents. The meaning of each item (and this is true for the items that seem very structured, too) is not always expected to be the same for all the participants involved in answering the questionnaire. Usually the respondent has troubles reasoning if the answer he/she intends to give is really "the right one" regarding true behavioural characteristics of the respondent. The standardization of personality questionnaires is therefore usually performed on large heterogeneous samples.
- Determining the validity of the instrument. This is perhaps the most serious problem in building a personality questionnaire, as it is very difficult to determine the external criteria, which could help examining its predictive and concurrent validity. The process of analysing the internal structure of the instrument and determining the construct validity is not at all an easy task, too. In addition, dilemmas concerning factor structure of many of the personality theories or models and their accompanying questionnaires are not completely resolved (see, for example, the discussion about the Cattell's 16 Personality Factors questionnaire (16PF) – e.g. Barrett & Kline, 1982; Gerbing & Tuley, 1991; Krug & Johns, 1986; Marsh & Bailey, 1991; Matthews, 1989 – or about the five-factor personality model – e.g. Brand, 1994a, 1994b; Buyst, De Fruyt & Mervielde, 1994; Draycott & Kline, 1995; McCrae, 1989; Mervielde, 1994; Ostendorf & Angleitner, 1994).
- Response sets or response strategies of different respondents when answering the same items in the questionnaire. They can be defined as (not necessarily willing and conscious) sources of variability linked to particular types of answering available in a certain type of items in a questionnaire (Wilde, 1977). Many different forms of possible response sets can be found in the literature (Cohen, Kim & Wollack, 1996; Furnham, 1990c; Greene, 1996; Newton, Stein & Lucey, 1998; Patz & Junker, 1999; Prieto & Delgado, 1999a; Rogers & Harley, 1999; Rorer, 1965; Shankar, 1999; Wang & Calhoun, 1997): (i) the so-called speed-accuracy trade-off in responding (where the respondent "sacrifices" the accurate answer on account of the fast accomplishing of the questionnaire, or on the contrary, he/she spends too much time thinking about the answers to questions in a questionnaire), (ii) the tendency to guess the "right" answer, (iii) agreeableness, (iv) the "central tendency" (which means that the respondent avoids choosing a categorical or extreme answer), (v) the "black-white" response set (which is the tendency to give only categorical answers), (vi) the tendency to respond randomly, (vii) the tendency to give socially desirable answers (which is particularly uneasy, as it is not necessary that the respondent in fact seized the "real" face validity of the instrument and of the certain item she/he is giving answer to). There are many other additional response sets.
- The nature of the scale that determines how the respondent will act when answering. Some studies have shown that the scale being used for answering the question is a very important variable affecting the observed score. It is also important whether the items are expressed in positive or negative fashion, or maybe even in form of double negation, which – according to all recommendations – should be avoided in personality questionnaires (Cronbach, 1990; Gordon & Holden, 1996, 1998; Harvil & Davis, 1997; Kline, 1993). Widely accepted and frequently used questionnaires try to diminish the influence of answering strategies with positive-negative balance in items. One of them is Cattell's 16PF questionnaire (Cattell, Eber & Tatsuoka, 1970).

It is well known that the personality structure of the respondent influences very much the nature and features of his/her answers to items in personality questionnaires. The personality of participants plays a very important role when they are in the situation to choose (or not to choose) one of the alternative answers to questions which are too difficult for them to answer, and they do not know the right answer to (Brenk & Bucik, 1994; Bucik, 2001; Bucik & Brenk, 1992; Dahlbach, 1990, 1991; Nazor, 1983; Petz & Žužul, 1987; Prieto & Delgado, 1999b; Weitzman, 1996; Zaleski, 1980, Zarevski & Rijavec, 1996). Therefore the participant's personality should not be neglected when discussing the factors influencing the possible discrepancies between the true and observed score in the examination of the personality profile. When developing a personality questionnaire, authors usually examine different strategies for establishing the sensitivity of instruments for "biased" answering of the respondents (see e.g. Furnham, 1986, 1990a, 1990b, 1990c for studies examining the role of facial validity of the instruments), but only few studies investigated the stability of measurements with regard to different response styles or response sets that self-reported personality questionnaires are usually liable to (see e.g. Ben-Simon, Budescu & Nevo, 1997; Deinzer, Steyer, Eid, Notz, Schwenkmetzger, Ostendorf & Neubauer, 1995; Furnham, 1990c; Lammers & Frankenfeld, 1999; Lim & Butcher, 1996). In some of the studies (e.g. Bucik, 1999; Bucik & Brenk, 1997) it has been distinctly shown that the specific method factor, the response-order effect (ROE), which appears in answering positively scored or reverse scored

items, distorts the inter-item correlation matrix and, consequently, also the results in the self-report personality questionnaire 16PF. There may as well be two factors – one responsible for bias in answering positive items and another, different, connected to answering negative ones. Although 16PF has a balanced design that tries to cancel out ROE, it had been proven (Vodopivec, Brenk & Bucik, 1992) that ROE introduces error to the covariance matrix and subsequently distorts the results of factor analysis, which is carried out to get the latent structure of answers to items. It is reasonable to expect that similar pattern could emerge in other questionnaires, which try to control answering strategies with positive-negative balance in items (Byravan & Ramanaiah, 1995). One of such questionnaires is a Big Five Questionnaire (BFQ; Caprara, Barbaranelli & Borgogni, 1993), which roots are in the five-factor model of personality (Costa & McCrae, 1992).

The aim of the present study was to examine concurrently the effect of ROE on answering the two personality questionnaires, 16PF and BFQ. Namely, both questionnaires, trying to control different answering strategies, consist of half of positively stated and half of negatively stated items for each dimension or subdimension.

METHOD

Participants

Three hundred eighty six participants (63% female, M_{age} = 21.8, SD_{age} = 1.9) of different levels of education and professional profiles completed the instruments as a part of job or scholarship selection procedure for various government administration positions in Slovenia.

Instruments

The Slovenian translations and adaptations of two well known self-report phrase personality questionnaires were used in the study.

The Cattell's 16 Personality Questionnaire (16PF, version C+D; Cattell, 1943a, 1943b; Cattell et al., 1970; Slovene adaptation described in Lamovec, 1980) consists of 105 items in balanced design regarding positively or negatively scored items for each factor. The respondent gives answers to items in a way that he/she chooses one of three alternatives, *a*, *b* or *c*, which represent behaviours typical for that item. The final scores represent the 16-factor profile of the respondent if interpreted at the first-order level, and 4-factor profile if interpreted at the second-order level. The first-order factors are as follows: A – Affectothymia, C – Ego strength, E – Dominance, F – Surgency, G – Superego strength, H – Threctia/Parmia, I – Harria/Premsia, L – Alaxia/Protension, M – Praxernia/Autia, N – Shrewdness, O – Guilt proneness, Q1 – Conservatism/Radicalism, Q2 – Social dependency/Self-sufficiency, Q3 – Integration and Q4 – Ergic tension. B scale ("intelligence") was excluded from the analysis because of different method of scoring. The second-order factors are: I – Low anxiety, II – Introversion, III – Emotionality and IV – Subduedness.

The Big Five Questionnaire (BFQ; Bucik, 1998; Caprara et al., 1993; Caprara, Barbaranelli, Borgogni, Bucik & Boben, 1997; Caprara, Barbaranelli, Borgogni & Perugini, 1993), which is a self-report factorial questionnaire which consists of 132 items measuring five dimensions of personality according to the five-factor model of personality (see Table 1 for the names of dimensions and subdimensions in BFQ and for the comparison of the names to those in one of the most established questionnaires following the five-factor model, NEO-PI-R and NEO-FFI; Costa & McCrae, 1992). The respondent rates on the 5-point scale how typical for him/her is the situation or behaviour de-

Table 1

The names of BFQ dimensions and subdimensions (abbreviations used in statistical analyses are given in parentheses; Caprara, Barbaranelli, & Borgogni, 1993) compared to the names of five dimensions in the NEO-FFI and NEO-PI-R (Costa & McCrae, 1992)

dimensions BFQ	subdimensions BFQ		dimensions NEO-FFI, NEO-PI-R
Energy (EN)	Dynamism (DYN)	Dominance (DOM)	Extraverison
Friendliness (FR)	Cooperativeness (COO)	Politeness (POL)	Agreeableness
Conscientiousness (CO)	Scrupulousness (SCR)	Perseverance (PER)	Conscientiousness
Emotional stability (ES)	Emotion control (ECO)	Impulse control (ICO)	Neuroticism
Openness (OP)	Openness to culture (OCU)	Openness to experience (OEX)	Openness
Lie scale (L)			

scribed in certain item in the questionnaire. Each subdimension is represented by 12 items, 6 of them are scored positively and 6 negatively. Two subdimensions (e.g. 12 positively and 12 negatively scored items) are contained in each dimension. Twelve items represent the *Lie* or *Social desirability* scale.

Procedures

Both questionnaires were applied in several group sessions. The order in which the participants were completing the instruments was changed randomly. The usual scaling procedures according to instructions in the manuals were performed on both instruments with some major changes described in the following paragraph.

For the purpose of the analysis we split each 16PF first- and second-order scale into two halves, one half consisting of items, where the answer *a* gets two, answer *b* one and answer *c* zero points (e.g. A+, C+, D+, ... for the first-order factors and I+, II+, ... for the second-order factors), and the other half consisting of the reverse-scored items (A-, C-, D-, ... and I-, II-, ... respectively).

In BFQ we also split each subdimension and dimension according to positive and reversed items (+Dynamism, +Dominance, +Cooperativeness, etc. and -Dynamism, -Dominance, -Cooperativeness for subdimensions, and +Energy, +Friendliness etc. and -Energy, -Friendliness etc. for dimensions). In both questionnaires, 16PF and BFQ, social desirability (Lie) scales were excluded from the analysis.

Confirmatory factor analyses were performed on the covariance matrices of these split-half scales, using LISREL 8.51 *maximum likelihood* algorithm (Du Toit & Du Toit, 2001; Jöreskog & Sörbom, 1994). Analyses were conducted on four sets of data:

- 16PF first-order factor structure (with 15 positive and 15 negative scales or factors),
- 16PF second-order factor structure (with 4 positive and 4 negative scales),
- BFQ subdimension structure (with 10 positive and 10 negative scales or subdimensions),
- BFQ dimensions (5 positive and 5 negative scales or dimensions).

RESULTS AND DISCUSSION

Different concurrent models for both instruments were analysed in the study, though not the whole set of results is presented here. Especially we didn't include the specific obtained parameter values of different empirical models (e.g. loadings and errors of measurement [i.e. parameters λ_X and θ_δ in LISREL terminology] and the factor variances). The reason is, as shall be seen in the following table 2, that the overall fit of the proposed models – even the best ones – can hardly be called satisfactory. For this reason we limited our presentation only to improvement of the overall fit of the models to the empirical data, just to see how the fit of the theoretical models progress from bad (classical, widely used models of personality, represented by frequently used questionnaires 16PF and BFQ) to better (but still not acceptable – at least in the statistical sense), including the method factor, which we termed "response-order effect" (ROE). As the fit of the proposed models doesn't meet the statistical criteria of at least borderline goodness of fit (but it clearly shows the trend or the pattern), the particular estimated parameters are not further investigated and they are not presented here.

Generally, different models were evaluated in the analysis: in Model 1 (the example for BFQ is presented in Figure 1) only the postulated trait factors were introduced as latent variables for all four sets of data. In Model 2 (Figure 2 for BFQ) the postulated trait factors plus the tested method factor, ROE, were introduced as latent variables. In Model 3 (Figure 3 for BFQ) the method factor was proposed as two distinct possible influences, response-order effect (ROE) for positive and the ROE for negatively stated items. In all models covariance matrix of trait factors was unconstrained. Covariance of the method factors with trait factors was constrained to zero. Measurement model parameters (e.g. λ_{XA+} and λ_{XA-}) were constrained to yield equal reliability of both halves of each trait's scale. Model factor loadings in Model 2 (e.g. λ_{XMA+} and λ_{XMA-}), as well as in Model 3, were constrained to be equal in absolute terms for all the half-scales. Loadings were positive for positively scored halves and negative for reverse scored halves. Theoretical basis for equality constraints was the notion that since the format of all items is equal, the method factor makes the same contribution to each item's variance, according to congenerity and ô-equivalence presumptions of the classical test theory (Bucik, 1997; Cronbach, 1990), which represented the methodological origin when creating both questionnaires.

Only the summary statistics for all three models in all four sets of data are presented (Table 2).

The Null-model (also called the "zero" or the "independence" model) presumes no relations between the units involved in the model (i.e. between constructs, between observed variables and between constructs and observed variables). This model is important as some goodness-of-fit indices rely and depend on it. In the confirmatory factor analysis goodness-of-fit indices should meet certain statis-

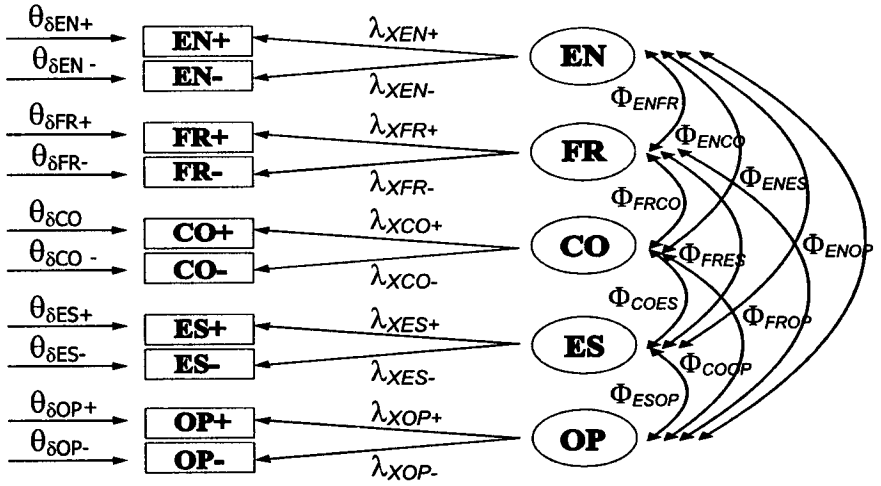

Figure 1. Model 1 for BFQ dimensions: A model without method factor.

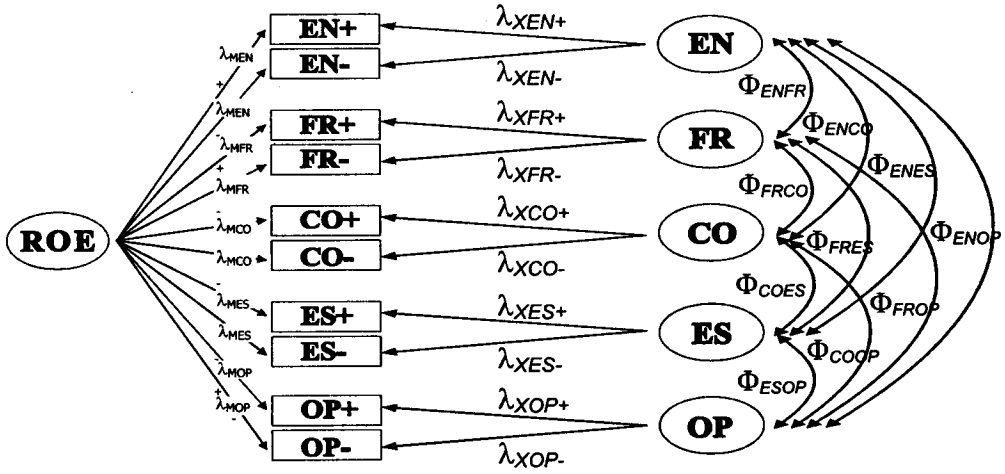

Figure 2. Model 2 for BFQ dimensions: A model with method factor.

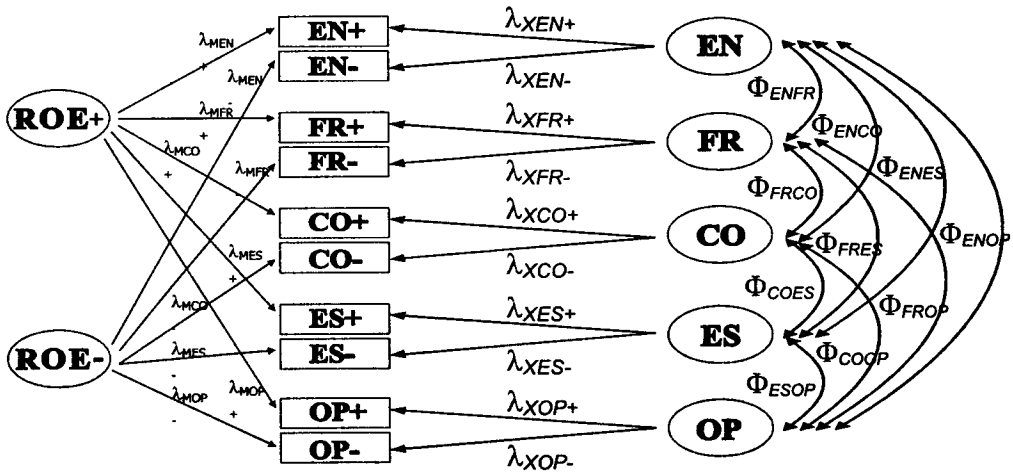

Figure 3. Model 3 for BFQ dimensions: A model with two method factors (positive ROE and negative ROE).

Table 2

Summary statistics of the model fit for Model 1, Model 2 and Model 3 in four data sets: the goodness-of-fit statistics (χ^2, χ^2/df ratio, *AGFI* and *NNFI*) for the estimated models regarding the Null-model (or the "independence" model) and differences in the indices between models

Data set	index	Null model	Model 1	Model 2	Model 3	$\Delta df_{(M1-M2)}$	$\Delta\chi^2_{(M1-M2)}$	$\chi^2_t(\Delta df;=\alpha\ 0.01)$
16PF I.order	*df*	335	315	299	285	16	189.40	32.00
	χ^2	2539.53	977.18	787.78	882.76			
	χ^2/df	7.58	3.10	2.63	3.00			
	AGFI	-	0.81	0.82	0.81			
	NNFI	-	0.63	0.66	0.64			
16PF II.order	*df*	28	18	13	9	5	15.29	15.09
	χ^2	711.78	65.10	49.81	52.22			
	χ^2/df	25.42	3.62	3.83	5.80			
	AGFI	-	0.92	0.91	0.90			
	NNFI	-	0.89	0.88	0.88			
BFQ–subdimensions	*df*	190	135	124	119	11	405.57	24.73
	χ^2	3262.44	888.61	483.04	698.57			
	χ^2/df	17.17	6.58	3.89	5.87			
	AGFI	-	0.65	0.82	0.70			
	NNFI	-	0.65	0.82	0.70			
BFQ–dimensions	*df*	45	30	24	19	6	253.03	16.81
	χ^2	1423.02	361.09	108.06	212.90			
	χ^2/df	31.62	12.09	5.15	11.20			
	AGFI	-	0.67	0.88	0.69			
	NNFI	-	0.64	0.88	0.67			

Note: *AGFI* - Adjusted Goodness-of-Fit Index; *NNFI* - Non-Normed Fit Index; Δdf – the difference in degrees of freedom between the two compared models; $\Delta\chi^2$ – the difference in χ^2 goodnes-of-fit indices for the two compared models; χ^2_t (Δdf; α=0.01) – theoretical borderline χ^2 at the 1% error level and at particular Δdf.

tical criteria to be of satisfactory value. Calculated χ^2-index should be lower than the theoretical χ^2-index with determined degrees of freedom, the χ^2/df ratio should not exceed the value of 2, and Adjusted Goodness-of-fit index (*AGFI*) and Non-normed fit index (*NNFI*) need to be as close to 1 as possible (Bentler & Bonett, 1980; Bollen, 1989; Brown & Cudeck, 1989, 1992; Tucker & Lewis, 1973). A substantial drop in the χ^2 test statistics associated with the loss of certain degrees of freedom was observed when the model with the ROE factor included (Model 2) was compared to the model without it (Model 1). This was the case in each of the four structures (16PF first-order factors, 16PF second-order factors, BFQ subdimensions and BFQ dimensions). Also other goodness-of-fit indices showed an improvement when the "method" factor (ROE) was included in the model. Further outcome points out that proposing two distinct method factors, the ROE for positively and ROE for negatively expressed items in both questionnaires isn't the appropriate solution. The relative decrease in χ^2-index and increase in other indices of goodness-of-fit in Model 3 from Model 1 is smaller than in Model 2. Therefore only the drops in χ^2 from Model 1 to Model 2 were pointed out. Positive and negative items may not function as parallel measurements, but it seems more likely that one method factor (it can be called "response order effect" or given some other similar name), rather than two, burdens the measurement. In addition, it was shown that the effect of ROE was stronger in BFQ than in 16PF on the level of dimensions and subdimensions, which can be seen when comparing the drops in χ^2-indices from the Null-model to Model 1, Model 2 and Model 3 for all four sets of data. It should be admitted, however, that the fit of the models, expressed by χ^2-index, χ^2/df ratio, *AGFI* and *NNFI*, is not very good and that the calculated solutions must be interpreted with caution. It is true, however, that the size of the drop of χ^2-index, and especially the drop in χ^2/df ratio is not the direct proof of relative size of ROE. It shows indirectly, but clearly, that the important amount of overall variance in the model was being explained with a newly introduced factor which deserves further investigation and explanation.

As BFQ is of particular interest to psychologists as a "modern questionnaire following the contemporary model of personality" (Caprara et al., 1997), we also performed the comparison of goodness-of-fit statistics for two identical models of BFQ, one including only positive halves of

subdimensions and one including only negative halves of the same subdimensions. The subdimensions were Dynamism, Dominance, Cooperativeness, Politeness, Scrupulousness, Perseverance, Emotion control, Impulse control, Openness to culture, Openness to experience. The results show that the model with positive items fit the theoretical model better than the model with negative items. Namely, the goodness-of-fit indices are as follows: for the negative items χ^2 (df = 25) = 144.50; $AGFI$ = 0.86; $NNFI$ = 0.79; for the positive items χ^2 (df = 25) = 113.23; $AGFI$ = 0.89; $NNFI$ = 0.82 (the lower the χ^2 with certain degrees of freedom and the higher the $AGFI$ and $NNFI$, the better). The same analysis was performed with 16PF, but the results are not reported here, as the tendency was almost exactly the same as with BFQ.

CONCLUSION

Psychologists investigating the factor structure of the self-report personality questionnaires, which include positive-negative balanced items, should consider the role of the response-order effect among other factors that possibly influence the result in the personality questionnaires. It is an important method factor, which distorts results in the structure of personality and drags the observed value off the true value. The effect has been detected in BFQ as well as in 16PF. This is interesting, because the two questionnaires belong to different models or theories of the structure of human personality, and might just as well mean that ROE is free of the particular model or theory of personality or strategy used in building personality questionnaires.

Notably, BFQ is more ROE dependent than 16PF. It cannot be said unequivocally from the results of this study that differences in the ROE effect in 16PF and BFQ are due to the differences in the rating scales used in both instruments, although one could speculate about that possibility, too. The difference might be merely the consequence of the fact that 16PF is classical, more established and metrically more refined instrument. Or it might be the case that the 3-point scale is simply more resistant to the influence of such factors as ROE than the 5-point scale when using one or the other in the self-report personality questionnaire.

Positive items seem to be slightly less problematic than negative ones. This was expressed in both questionnaires although only the results in BFQ were presented here. Does it mean that it would be better to abandon the negative or reversed items and start to use some other strategy of balancing items to control answering strategies or response sets? Probably not, as the negatively expressed items have their specific purpose in psychometric tests, questionnaires and scales. Test constructors just ought to take special care about the nature of such items. Maybe it is not essential that the balance between the positive and negative items follows the "fifty-fifty" rule. The cost-benefit trade-off in gathering information about personality structure is obviously not in favour of such a strategy. One question remains unanswered, namely, why do the reversed items not serve their purpose as they are expected? Do respondents have more serious problems comprehending the items that are not positively oriented? Let us not forget that positive items in the instruments analysed in our study are merely the reflection of the expression of the dimension. It simply means that when choosing "yes" or giving high marks to certain items, which represent particular dimension, the respondents express that this typical behaviour reflects their trait, which is actually the positive side of this dimension. It seems that items through which the respondents "collect" points in certain dimension by answering "no" or giving low marks just do not have the same efficacy than the positive items.

The results of our study point to the conclusion that, if we control the answering strategies solely with positive-negative balance in items, we may not attain our goal completely. Comparative examination of the effect of this systematic method factor on answering the two personality questionnaires, the Slovenian adaptations of 16PF and the Big Five Questionnaire, has indicated that the method factor (ROE) has an important influence on the answers in both questionnaires, and that positively scored items are more reliable representatives of the constructs. The negatively scored items seem to contribute more to the error of measurement. It is also possible that ROE is not an one-dimensional source of error as we presupposed as it may just have different impact on positive- than on negative-scored items as well. It would be interesting to examine the correlation between the ROE and characteristics of different respondents' personality and to analyse how some other similar questionnaires, which use the same positive-negative balance in items, reflect sensitivity to the response-order effect. These questions need to be resolved in the near future.

REFERENCES

BARRETT, P., & KLINE, P. (1982). An item and radial parcel analysis of the 16PF questionnaire. *Personality and Individual Differences, 3,* 259-270.

BEN-SIMON, A., BUDESCU, D.V., & NEVO, B. (1997). A comparative study of measures of partial knowledge in multiple-choice tests. *Applied Psychological Measurement, 21,* 65-88.

BENTLER, P.M., & BONETT, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588-600.

BOLLEN, K.A. (1989). *Structural equation models with latent variables.* New York: Wiley.

BRAND, C.R. (1994a). How many dimensions of personality? - The "Big 5", the "Gigantic 3" or the "Comprehensive 7"? *Psychologica Belgica, 34,* 257-273.

BRAND, C.R. (1994b). Open to experience - closed to intelligence: Why the 'Big Five' are really the 'Comprehensive Six'. *European Journal of Personality, 8,* 299-310.

BRENK, K., & BUCIK, V. (1994). Guessing of answers in objective tests, general mental ability and personality traits according to 16PF questionnaire. *Review of Psychology, 1,* 11-20.

BROWN, M.W., & CUDECK, R. (1989). Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research, 24,* 445-455.

BROWN, M.W., & CUDECK, R. (1992). Alternative ways of assessing model fit. In K.A. Bollen, & J.S. Long (Eds.), *Evaluating structural models* (pp. 63-75). Beverly Hills, CA: Sage.

BUCIK, V. (1997). *Osnove psihološkega testiranja [Essentials of psychological testing].* Ljubljana: Univerza v Ljubljani, Filozofska fakulteta, Oddelek za psihologijo.

BUCIK, V. (1998). Konvergentna in diskriminativna veljavnost modela "velikih pet" faktorjev osebnosti [Convergent and discriminant validity of the Big Five model od personality]. *Psihološka obzorja / Horizons of Psychology, 7* (2), 5-24.

BUCIK, V. (1999). Response-order effect in the self-reported personality questionnaires: The case of 16PF and BFQ. *Paper presented at the XIV. Psychological meeting "Days of Ramiro Bujas" - Zagreb, Croatia, December 16-18* (book of abstracts, pg. 23).

BUCIK, V. (2001). Ugibanje odgovora v testu dosežka, osebnostna struktura ter inteligentnost [Speculating about the right answer in achievement tests, personality and intelligence]. *Psihološka obzorja / Horizons of Psychology, 10* (4), 75-88.

BUCIK, V., & BRENK, K. (1992). Strategija odgovarjanja v objektivnih testih in nekatere osebnostne lastnosti - hierarhični pristop [The response strategy in objective tests and personality: Hierarchical approach]. *Psihološka obzorja / Horizons of Psychology, 1* (1), 15-19.

BUCIK, V., & BRENK, K. (1997). Response-order effect and its influence on the Five-factor model. *Paper presented at the IV. European Conference on Psychological Assessment - Lisbon, Portugal, September 7-10 (book of abstracts, pg. 79).*

BUYST, V., DE FRUYT, F., & MERVIELDE, I. (1994). Parental description of children's personality: A five-factor model classification. *Psychologica Belgica, 34,* 231-255.

BYRAVAN, A., & RAMANAIAH, N.V. (1995). Structure of the 16PF fifth edition from the perspective of the five-factor model. *Psychological Reports, 76,* 555-560.

CAPRARA, G.V., BARBARANELLI, C., & BORGOGNI, L. (1993). *BFQ - Big Five Questionnaire. Manuale.* Firenze: Organizzazioni Speziali.

CAPRARA, G.V., BARBARANELLI, C., BORGOGNI, L., BUCIK, V., & BOBEN, D. (1997). *Model "velikih pet": Pripomočki za merjenje strukture osebnosti. Priročnik [The Big Five Model: Instruments for measuring the structure of personality. Manual].* Ljubljana: Center za psihodiagnostična sredstva.

CAPRARA, G.V., BARBARANELLI, C., BORGOGNI, L., & PERUGINI, M. (1993). The "Big Five Questionnaire": A new questionnaire to assess the five factor model. *Personality and Individual Differences, 15,* 281-288.

CATTELL, R.B. (1943a). The description of personality: I. Foundations of trait measurement. *Psychological Review, 50,* 559-594.

CATTELL, R.B. (1943b). The description of personality. II. Basic traits resolved into clusters. *Journal of Abnormal and Social Psychology, 38,* 476-507.

CATTELL, R.B., EBER, H.W., & TATSUOKA, M.M. (1970). *The handbook for the Sixteen Personality Factor Questionnaire.* Champaign, IL: Institute for Personality and Ability Testing.

COHEN, A.S., KIM, S.H., & WOLLACK, J.A. (1996). An investigation of the likelihood ratio test for detection of differential item functioning. *Applied Psychological Measurement, 20,* 15-26.

COHEN, R.J., MONTAGUE, P., NATHANSON, L.S., & SWERDLIK, M.E. (1988). *Psychological testing: An introduction to tests & measurement.* Mountain View, CA: Mayfield.

COSTA, P.T., JR., & MCCRAE, R.R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual.* Odessa, FL: Psychological Assessment Resources.

CRONBACH, L.J. (1990). *Essentials of psychological testing.* New York: Harper Collins.

DAHLBACH, O. (1990). Personality and risk-taking. *Personality and Individual Differences, 11,* 1235-1242.

DAHLBACH, O. (1991). Accident proneness and risk-taking. *Personality and Individual Differences, 12,* 79-85.

DEINZER, R., STEYER, R., EID, M., NOTZ, P. SCHWENK-METZGER, P., OSTENDORF, F., & NEUBAUER, A. (1995). Situational effects in trait assessment: the FPI, NEOFFI, and EPI questionnaires. *European Journal of Personality, 9,* 1-23.

DRAYCOTT, S.G., & KLINE, P. (1995). The Big Three or the Big Five - the EPQ-R vs the NEO-PI: a research note, replication and elaboration. *Personality and Individual Differences, 18,* 801-804.

DU TOIT, M., & DU TOIT, S. (2001). *Interactive LISREL: User's guide.* Lincolnwood, IL: Scientific Software International.

FURNHAM, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences,* 7, 385-400.

FURNHAM, A. (1990a). Can people accurately estimate their own personality test scores? *European Journal of Personality,* 4, 319-327.

FURNHAM, A. (1990b). Faking personality questionnaires: Fabricating different profiles for different purposes. *Current Psychology, Research & Reviews,* 9, 46-55.

FURNHAM, A. (1990c). The fakebility of 16PF, Myers-Briggs and Firo-B personality measures. *Personality and Individual Differences, 11,* 711-716.

GERBING, D.W., & TULEY, M.R. (1991). The 16PF related to the five-factor model of personality: Multiple-indicator mesurement versus the a priori scales. *Multivariate Behavioral Research, 26,* 271-289.

GORDON, E.D., & HOLDEN, R.R. (1996). Use of item ratings to examine personality test item cognitive response processes. *Personality and Individual Differences, 21,* 897-905.

GORDON, E.D., & HOLDEN, R.R. (1998). Personality test item validity: Insights from "self" and "other" research and theory. *Personality and Individual Differences, 25,* 103-117.

GREENE, R.L. (1996). Mirror effect in order and associative information: Role of response strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 687-695.

HARVIL, L.M., & DAVIS, G. (1997). Medical students' reasons for changing answers on multiple choice tests. *Academic Medicine, 72 (10, Suppl 1),* S97-S99.

JÖRESKOG, K. G., & SÖRBOM, D. (1994). *LISREL 8 - User's reference guide.* Chicago, IL: Scientific Software International.

KLINE, P. (1993). *The handbook of psychological testing.* London: Routledge.

KRUG, S.E., & JOHNS, E.F. (1986). A large sample validation of second-order personality structure defined by the 16PF. *Psychological Reports, 59,* 683-693.

LAMMERS, F., & FRANKENFELD, V. (1999). Effekte gezielter Antwortstrategien bei einem Persönlichkeitsfragebogen mit "forced choice" Format. / Effects of selective response strategies in a personality questionnaire with forced choice format. *Diagnostica, 45,* 65-68.

LAMOVEC, T. (1980). *Eksperimentalni priročnik iz psihologije motivacije, emocij, osebnosti in mišljenja [Experimental textbook from psychology of motivation, emotions, personality and cognition].* Ljubljana: Univerza v Ljubljani, Filozofska fakulteta, Oddelek za psihologijo.

LIENERT, G.A., & RAATZ, U. (1994). *Testaufbau und Testanalyse.* Weinheim: Beltz.

LIM, J., & BUTCHER, J.N. (1996). Detection of faking on the MMPI 2: Differentiation among faking bad, denial, and claiming extreme virtue. *Journal of Personality Assessment, 67,* 1-25.

MARSH, H.W., & BAILEY, M. (1991). Confirmatory factor analyses of multitrait-multimethod data: A comparison of alternative models. *Applied Psychological Measurement, 15,* 47-70.

MATTHEWS, G. (1989). The factor structure of the 16PF: Twelve primary and three secondary factors. *Personality and Individual Differences, 10,* 931-940.

MCCRAE, R.R. (1989). Why I advocate the five factor model: Joint analysis of the NEO-PI and other instruments. In D.M. Buss & N. Cantor (Eds.), *Personality psychology: Recent trends and emerging directions* (pp. 237-245). New York: Springer Verlag.

MERVIELDE, I. (1994). The five-factor model: Quo vadis? *Psychologica Belgica, 34,* 275-281.

MURPHY, K.R., & DAVIDSHOFER, C.O. (1988). *Psychological testing: Principles and applications.* Englewood Cliffs, NJ: Prentice-Hall.

NAZOR, M. (1983). Zavisnost uratka u testu inteligencije od osobina ličnosti ispitanika [Personality character-

istics' impact on the results in intelligence test]. *Primijenjena psihologija, 4,* 49-53.

NEWTON, K., STEIN, S.M., & LUCEY, C. (1998). Influence of mailing strategies on response to questionnaires. *Psychiatric Bulletin, 22,* 692-694.

OSTENDORF, F., & ANGLEITNER, A. (1994). The five-factor taxonomy: Robust dimensions of personality description. *Psychologica Belgica, 34,* 175-194.

PATZ, R., & JUNKER, B.W. (1999). Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses. *Journal of Educational and Behavioral Statistics, 24,* 342-366.

PETZ, B., & ŽUŽUL, M. (1987). "Riskiranje" u testovima znanja i neke karakteristike ličnosti ["Risk-taking" in achievement tests and some personality characteristics]. *Primijenjena psihologija, 8,* 37-42.

PRIETO, G., & DELGADO, A.R. (1999a). The effect of instructions on multiple-choice test scores. *European Journal of Psychological Assessment, 15* (2), 143--150.

PRIETO, G,. & DELGADO, A.R. (1999b). The role of instructions in the variability of sex-related differences in multiple-choice tests. *Personality and Individual Differences, 27,* 1067-1077.

ROGERS, W.T., & HARLEY, D. (1999). An empirical comparison of three and four choice items and tests: Susceptibility to testwiseness and internal consistency reliability. *Educational and Psychological Measurement, 59,* 234-247.

RORER, L.G. (1965). The great response-style myth. *Psychological Bulletin, 63,* 129-156.

SHANKAR, V. (1999). New product introduction and incumbent response strategies: Their interrelationship and the role of multimarket contact. *Journal of Marketing Research, 36,* 327-344.

TUCKER, L.R., & LEWIS, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38,* 1-10.

VODOPIVEC, B., BRENK, K., & BUCIK, V. (1992). Response style (method) factor in Cattell's 16PF: Confirmatory Analysis. *International Journal of Psychology, 27,* 383.

WANG, J., & CALHOUN, G. (1997). A useful function for assessing the effect of guessing on true-false and multiple-choice tests. *Educational and Psychological Measurement, 57,* 179-185.

WEITZMAN, R.A. (1996). The Rasch model plus guessing. *Educational and Psychological Measurement, 56,* 779-790.

WILDE, G.J.S. (1977). Trait description and measurement by personality questionnaires. In R.B. Cattell & R.M. Dregere (Eds.), *Handbook of modern personality theory* (pp. 213-252). New York: Halsted.

ZALESKI, Z. (1980). Personality traits in high- and low-risk takers according to R.B. Cattell's 16-factor questionnaire. *Polish Psychological Bulletin, 11,* 191-201.

ZAREVSKI, P., & RIJAVEC, M. (1990). Personality correlates of guessing proneness in objective tests. *Revija za psihologiju, 20,* 27-33.

Received: August, 2003.
Accepted: November, 2003.