Eötvös Loránd University, Faculty of Education and Psychology

Olivér Nagybányai Nagy

The Effect of Response Style Characteristics on the Measuring Efficiency of Self-administered Testing Methods

DOCTORAL (PH.D.) THESIS BOOKLET

Doctoral School of Psychology:
Head of Doctoral School: Prof. Attila Oláh

Personality and Health Psychology Program
Head of Doctoral Program: Prof. Attila Oláh

Supervisor: Prof. Attila Oláh

2013

Content

I. Introduction (1): Establishing a trait taxonomy on the basis of factor analyses

One of the most accepted taxonomy and model in classifying and systemizing individual differences in personality is the Big Five approach and the Five Factor Model itself, which is statistically and methodologically based on factor analysis. By using this method, the numerous concepts describing personality (i.e., the heterogeneity of personality dispositions) can be captured in larger units (aggregated to factors), which could also shed light on the structure of these relevant concepts. Data underlying the factor analytic approach was obtained from studies concerning the basic dimensions of earlier theories (see Eysenck, 1967), and from the analysis of the whole vocabulary related to personality (e.g., Cattell, 1943). Establishing the taxonomy of phenotypic characteristics based on natural language looks back to a relatively long history: the definition of the five factors (used in our recent understanding) dates back to the 1940s (Fiske, 1949), although their breakthrough and real acceptance goes back to the 1980s (Goldberg, 1981).

According to the main results of Hungarian studies of trait taxonomy (Szirmák & De Raad, 1994), the first four dimensions were found to be relatively unambiguous, however, the presence of the fifth factor and the name of its relating dimension was questionable. When comparing the four and five-factor models it became clear that the fifth factor evolved by decoupling the Agreeableness dimension. One of the two separable dimensions was characterized by the attributes of “stubborn-conformable”, while the other was described by the “sacrificing-intrigue” word pair. Statistically and in content both were related to the Agreeableness dimension, but yet they formed distinct dimensions.

In regard to test-methodology, by investigating the structure of the widely-used Hungarian Big Five personality questionnaires we could observe that all of these instruments measure the aspects of Openness as the content of the fifth factor. They do not comprise most of the attributes related to the fifth dimension relevant in Hungarian respects separately, but distribute to other factors. Therefore it would be important to make more questionnaires available which fit well the Hungarian Big Five structure in personality measurement.

II. Introduction (2): Online-testing and the detection of response style in testing research

Lately, the methods of online and computerized evaluation and interpretation are in essence predominant compared to the classical measurements of personality (i.e., by paper-pencil tests). Through the analysis of indirect digital “footprints” which are generated during online activities, we could study the indices of “response pattern” and “response latency” as secondary features of responding (i.e., completing the questionnaire) besides content-based analysis.

“Response frequencies” are defined as systematic biases when participants’ responding is not only based on the content of the items, but also affected by other external variables. In this research field, instead of “response tendency”, the concept of “response style” (Jackson and Messick, 1958), and the slightly different concept of “response set” (originally introduced by Cronbach, 1946; Cheung & Rensvold, 2000; Watkins & Cheung, 1995) have become conventional (De Beuckelaer, Weijters, & Rutten, 2010; Van Herk, Poortinga, & Verhallen, 2004). Authors define response set with the form and content of the items, or with situational (external) factors. However response style is a general personal feature (internal disposition) that primilarly does not depend on the characteristics of items (e.g. form and content).

Various indicators are available for studying individual differences in response frequencies (Baumgartner & Steenkamp, 2001). The two most frequently investigated patterns are the “extreme response style” (ERS) and the “agreement response style” (ARS). ERS is considered as a consistently appearing characteristic which is stable in time (Bachman & O'Malley, 1984; Greenleaf, 1992; Naemi, Beal, & Payne, 2009), and separable from ARS or social desirability.

“Midpoint response style” (MidRS) is investigated more rarely partly because it detects only midpoint choices (in case of a 5-point Likert-scale only response “3” is considered). A midpoint choice could indicate a moderate or average level of response relevant to a certain item at the first place, but it could be interpreted as a “neutral response option” halfway between two extreme ends as well (Weijters, Geuens, & Schillewaert, 2010). If the given responses cannot be described by ERS or MidRS, a supplementary indicator labeled “moderate response style” (ModRS) should be measured.

We should distinguish two axes in the taxonomy of response biases (Griffith, Peterson, Isaacson, Quist, & Gammon, 2009); one from the aspect of motivation (active/conscious or passive/unconscious), and the other from the aspect of outcome (positive/functional or negative/dysfunctional). In this framework we could easily interpret the concept of “social desirable responding” (SDR), which is a special form of response bias. Two aspects of SDR are known (Paulhus, 1991): the “impression management” (IM) which refers to conscious deception, and the “self-deceptive enhancement” (SDE) regarded as a truthful and positive-oriented bias. According to the empirical data, these two aspects are considered as relating factors (Ones, Viswesvaran, & Reiss, 1996). As SDR is linked to specific situations, it might include a dispositional aspect, and by this reason SDR per se could be associated more with the individual response style. We could study the individual differences of social desirability within the skill and motivation of impression management bias (McFarland & Ryan, 2000, 2006).

Another relevant group of indices in studying response style besides response frequencies is related to “response latencies”. According to different models, the interpretation of response times is derived from the individual variability in responding, and from the heterogeneity of items. The former is influenced by personality dispositions, and also by the reading speed and motor execution; the latter is affected by item-related features such as linguistic complexity and length of item (Holden, Fekken, & Cotton, 1991; van der Linden, 2009). Dependent variables related to response latencies could be classified into two types: indices based on absolute temporal features and on transformed temporal features. Amongst the latter, the most common techniques are standardization (Popham & Holden, 1990) and logarithmic transformation (van der Linden, 2009).

The inverse U-shaped relation (“response time effect”) between response latency and personality measures was first introduced by Kuiper (1981), but other studies have also confirmed this specific pattern since (Amelang, Eisenhut, & Rindermann, 1991; Mueller, Thompson, & Dugan, 1986). As an interpretation of this phenomenon it was assumed that a certain aspect of the self schema could have a cognitive facilitation effect on responses given to self-relevant items (Akrami, Hedlund, & Ekehammar, 2007). Response speed has been explained as an indicator of “schematicness” which has yielded schema-driven decision making. In another theoretical framework this decision strategy could be conceptualized as a stereotypic response style. Further studies on social desirability effect (e.g., Holden, Kroner, Fekken, & Popham, 1992) showed that response times to socially more desirable items were shorter in case of dishonest (positively biased) completion of questionnaires. Authors state if respondents are to form a positive impression about themselves, it takes less time to answer socially desirable items (items which are in accordance with the scheme).

Indices of response style can be obtained on diverse cases, e.g., from different item pools and response scales. 25 to 30 percent of the respondents are characterized by

extreme response style according to the results of previous researches (Austin, Deary, & Egan, 2006; Eid & Rauber, 2000), hence this particular response style seems to be rather frequent in various testing situations using various questionnaires. Longitudinal studies showed a high temporal stability of ARS and ERS (Bachman & O'Malley, 1984), even if several years have passed between the two time-points of testing (Billiet & Davidov, 2008). It is important to note that the magnitude of test-retest correlations (r = 0.8) of ERS are similar to those measured in case of personality dispositions; additionally, the same value of ARS is moderately high (r = 0.5), as well (Littvay, 2010). Considering the above-mentioned results, it is presumable that latent psychological attributes could underlie the response style characteristics (Austin, et al., 2006).

Regarding cultural and language differences, it is outstanding that ERS is more frequent in Western countries, while ARS is preferred in Eastern-Asian countries (Gilman et al., 2008; Hamamura, Heine, & Paulhus, 2008). These facts draw attention to a possible methodological confound: significant results obtained from cross-cultural studies or from others aiming to measure accultural strategies might contain artifacts as a consequence of not controlling response tendency. Simulation studies and others using common datasets showed that statistically significant but low positive correlations among scales could change or could turn into the opposite (negative) direction as accounting for by individual response biases in the model besides the content-based scores of the scales (Johnson, 2003). Psychometric studies underpin that response style could have a remarkable impact on the reliability and validity of the test results. Preferring ERS could enhance their reliability then again reduce their validity (Clarke, 2000). Virtually, any kind of consistent shift or accumulation in the distribution of responses yields some “content-independent” shared variance which could occasionally threaten the inner consistency of scales.

In sum, response tendency as an individual attribute prevailing independent of response content could considerably affect all answers given in a questionnaire. If we regard this “content-independent bias effect” as a general phenomenon, we should expect test results manifesting unreal profiles (type I error) or hidden latent traits (type II errors).

III. Research questions and the structure of empirical studies

1. The Hungarian adaptation of Facet5 personality questionnaire enables to use a test which fits well the results of the Hungarian taxonomical research on traits. By using this tool we could measure response characteristics that is infrequent in the Hungarian testing practice. If the previously described patterns of response tendency and response latency are detectable in Hungarian testology, their specific relations to personality traits could be unfolded.

2. Based on simpler techniques, the creation of new indices related to response latency would be useful. If these are able to effectively emphasize individual differences, it will be easier to clarify their relation to personality traits.

3. By means of studying the variations in response tendency and response latency among different situations, I have made an attempt to detect an “effort to manipulate test results” which could have a considerable effect on personality profiles.

IV. Empirical study (1): Online testing of personality traits according to the Hungarian Big Five structure: the adaptation of Facet5

By adapting the online Facet5 questionnaire (Buckley & Williams, 2002), which uses easy-to-understand, everyday concepts, we would have a personality test effective in non-clinical settings such as work psychology environment (Mirnics, 2006, p. 159). In this questionnaire, respondents give their answers on a 5-point response scale (from 1 to 5); they have to choose between two items of opposite content (i.e., ipsative measure) through 106 questions (e.g. "*Sloppy workmanship is a sign of an irresponsible attitude*" vs. "*There is no need to get too concerned about details*").
By summing up the responses Sten-scores can be obtained for the five scales (colors used in figures throughout are given in brackets): *Will* scale / Agreeableness (green); *Energy* scale / Extraversion (yellow); *Affection* scale / Integrity (red); *Control* scale / Conscientiousness (blue); *Emotionality* scale / Neuroticism (purple)

The data analyzed here was obtained solely by online testing. Gender distribution of the 1026 respondents was balanced (N_{male} = 543; N_{female} = 492), while their age was rather heterogeneous (age range: 18-65 years; M = 32.2, SD = 9.2).

The inner consistency of the scales (Cronbach alpha values) in the Hungarian sample was comparable to those obtained by the test developers on their original sample (Rawling & Reid, 2006). Reliability coefficients felt into a narrow interval (CI95 = 0.66 – 0.81), and according to the Feldt tests (Feldt, 1969), alpha values of the scales did not differ significantly from the original scale reliability indices.

By performing confirmatory factor analysis (CFA), the five-factor model yielded acceptable fit to the Hungarian data, however, some of the fit indices indicated poor fit. When evaluating the Hungarian model (Model HU) we should consider that the original English model (Model UK) exhibited essentially the same level of fit in that sample.

	N	χ2/df	TLI	CFI	SRMR	RMSEA [CI_{95}]
Model UK	1055	3,67	0,517	0,502	0,084	0,050 [0,049-0,051]
Model HU	1035	2,87	0,529	0,542	0,071	0,042 [0,041-0,043]

χ2/df = Ratio of chi square to its degrees of freedom; CFI = Comparative Fit Index; TLI = Tucker-Lewis index; SRMR = Standardized Root Mean Square Residual; RMSEA = Root Mean Squared Error of Approximation

Table 1. Fit index of CFA

The background of this result might be the particular 5-point response scale of Facet5, which is different from the traditional Likert-scale. It results in different response frequencies and spreading of response categories which could have an impact on the average correlation amongst variables, and ultimately, it might influence the CFI in a crucial extent. However, similar values of fit have also revealed in case of other Big Five questionnaires operating with orthogonal factors (e.g., the NEO-PI-R; TLI = 0,52; CFI = 0,55; see McCrae, Zonderman, Costa, Bond, & Paunonen, 1996, p. 556). Further studies are needed in order to achieve better model fit shown by these auxiliary fit indices.

The measurement of temporal stability of the Facet5 scales was based on a restricted sample of undergraduate students (n = 85; matched on gender). There was one month between the two testing sessions; test-retest correlations were satisfactory (r = 0.68 – 0.71; p < 0.01).

128 participants from another undergraduate sample filled in 3 questionnaires, one each week: Facet5, NEO-PI-3 (McCrae, Costa Jr, & Martin, 2005), and BFI (Big Five Inventory; John & Srivastava, 1999). This dataset served as a basis for studying

convergent validity of the Facet5 questionnaire. The special structure of personality traits obtained in Hungarian lexical studies was confirmed in Facet5 by measuring the classical five dimensions with all the questionnaires. Earlier studies showed that Agreeableness was decomposed to two distinct aspects, and the dimension of Openness was de-emphasized. Correlations among Facet5 and NEO-PI-3 were in line with these former results.

Pearson correlation	**NEO Extrav.**	**NEO Agreeabl.**	**NEO Consci.**	**NEO Neurotic.**	**NEO Openness**
F5 Will	0,285**	-0,495**	0,111	-0,137	0,071
F5 Energy	0,732**	-0,194*	-0,104	-0,088	0,277**
F5 Affection	-0,054	0,670**	0,087	-0,119	0,023
F5 Control	-0,091	0,175*	0,641**	-0,035	-0,350**
F5 Emot.	-0,309**	0,017	-0,146	0,755**	-0,007

***p < 0,01; * p < 0,05*

Table 2. Correlation between Facet5 and NEO-PI-3 scales

Common factor analysis performed on the three questionnaires yielded essentially the same results: scales related to *Agreeableness* of BFI and NEO-PI-3 loaded on the same factor as *Affection* and *Will* (with negative loading) from Facet5. When omitting the dimensions of Openness, the five-factor structure showed similar results to those of Szirmák: the *Agreeableness* of NEO and the *Agreeableness* of BFI loaded on the same factor as the *Affection* of Facet5, while *Will* was manifested as a distinct fifth factor, showing minor overlap with other factors. Is sum, it seems that the scales of Facet5 fit well the Hungarian structure of trait taxonomy.

Scales	**F a c t o r s**				
	1	**2**	**3**	**4**	**5**
F5_Energy	*0,908*		*-0,181*		
BFI_Extrav.	*0,886*	*-0,147*		*-0,126*	
NEO_Extrav.	*0,885*	*-0,178*			
BFI_Neurot.		*0,904*			*0,249*
NEO_ Neurot.	*-0,141*	*0,900*	*-0,222*		
F5_Emotion.	*-0,202*	*0,880*			*-0,136*
NEO_ Conscient.		*-0,199*	*0,916*	*0,103*	
BFI_Conscient.	*0,146*	*-0,232*	*0,856*		
F5_Control	*-0,149*	*0,175*	*0,832*		
F5_Affection	*-0,115*			*0,927*	
NEO_Agreeabl.	*-0,133*		*0,144*	*0,839*	*-0,303*
BFI_Agreeabl.	*0,282*	*-0,256*		*0,637*	*-0,489*
F5_Will	*0,282*		*0,135*	*-0,303*	*0,843*

Table 3. Rotated component matrix of common factor analysis

As investigating the normative (standard) scores in international comparison it became clear that two personality dimensions could be treated as “outliers” in the Hungarian version of Facet5. The mean of *Emotionality* scale had one of the highest values among other countries, while *Affection* had far the lowest value (significantly lower that the mean of other countries; Cohen’s d = 0.51-0.87).

Similar diversities regarding the Hungarian population were shown in relation to other questionnaires (pl. BFQ, lásd Rózsa, Kő, & Oláh, 2006); and the especially low average score of the *Affection* dimension was also in line with several Hungarian studies on attitudes. According to the results of these latter studies, Hungarian people are moderately willing to help others at the social level (e.g., Tóth, 2009).

V. Empirical Study (2): Pattern of response characteristics, response extrimity clasters, and the role of response-latency change in personality measurement

In order to identify personality traits behind response characteristics it is necessary to reveal the relationship between the single variabilities of response-latency indicators and the Big Five personality factors. For these analysises we applied the same dataset that was used for the valaidity and reliability analysises of the Facet5 as a part of the questionnaire's adaptational study (N = 1026).

As examining peculiarities of the pattern of response characteristics, the first and most prominent findings were the uneven distribution of response grades and the similar frequency of response options in forms of a moderate (with the values 2 and 4) and an extreme response pair (with the values 1 and 5). This pattern could be observed both in the case of separated personality scales and the total questionnaire (when it was proportioned in blocks consisted of 10 items). These response paires not only hold similar frequencies but correlate strongly by pairs in personal responses ($r = 0,72 – 0,84$; $p < 0,01$) while they are in a complementer relation with each other ($r = -0,85$; $p < 0,01$).

Two new kind of measures were designed as a common indicator of response tendency. The first kind is the so called Response Extremity Index which is the difference between extreme (1 and 5) and moderate (2 and 4) response frequency. The second kind is the so called Remnant Response Extremity Indicator calculated by items unrelated to the given personality dimension. Response Extremity Indexes related to particular scales are not independent from each other, all of them hold a strong, significant linear relationship with the others ($r = 0,69 – 0,74$; $p < 0,01$). Two strongly devided respondent clusters were identified based on the personal levels of response extremity by the Two Step Cluster algorithm in a cluster analysis procedure.

Based on our result it seems that response tendency, which seems to be a consistent personal characteristic, has a global additional effect on Facet5 questionnaire results. It was our aim to eliminate this effect by creating response cluster standards. This procedure can balance the score shifts caused by personal response tendency by creating so called corrected Sten-scores that are cleaned of this response distortion effect. The range of score corrections (y-axis) changes between -1 and +1 in function of the original Sten-scores (x-axis) by all five of the personality dimensions.

These score shifts don't seem to be enormous in terms of pure numbers, but it is important to emphasise that these corrections holds the greatest extent in the more extreme score ranges of the Sten-scale, the ranges most sensitive to minor differences. (It can happen that the very same person will be put to the 8% population frequency zone instead of the 1% zone because of the correctional procedure.)

Facet5 score corrections occure with a strikingly huge probability within the sample. In case of ERS, approx. 50% of the respondents was concerned in a minimal, half Sten-point

shift, while one-third of the respondents with ModRS could be characterized by similar or even larger score shift. Larger than one point shift (summed Sten-corrections) was present in 40% of the sample.

	Groups		
Scale correction	**Extreme-cluster**	**Moderate-cluster**	**Total sample**
Will (>0,5 Sten)	18%	20%	19%
Energy (>0,5 Sten)	44%	18%	31%
Affection (>0,5 Sten)	24%	27%	26%
Control (>0,5 Sten)	6%	35%	21%
Emotion. (>0,5 Sten)	25%	15%	20%
Total Scale-average (>1 Sten)	45%	34%	40%

Table 4. The percentage rate of the sample on each personlity dimension after Sten-point corrections

Midpoint choices are less informative choice options, and their frequency is rather low, hence they do not influence scale scores remarkably. Correlational analyses showed that most personality traits were not related to the frequency of midpoint choices, only the extremely high values on the dimension of *Control* were associated with low preference of midpoint choice.

In line with previous studies (Austin, et al., 2006), relations among response tendency and Big Five personality dimensions can be established in Facet5. 7 to 8 percent of squared scale scores of four out of five dimensions in Facet5 (*Control, Affection, Energy, Will*) was accounted for by the residual part of response extremity index. *Emotionality* was again an exception among Facet5 scales. When performing discriminant analysis a cumulative classification result of 69 % was achieved (Area Under ROC Curve: AUC = 70%). This ratio was higher in the moderate cluster (approx. 80%), but explained variance, eigenvalue (0.173), and canonical correlation (r_{cc} = 0.38) were low.

One third of the whole sample manifested rather different patterns is both response clusters. Respondents in the moderate cluster were described by values of scale scores above average, while others in the extreme cluster had scores below average. This “tierce” of the sample seemed to be a specific group showing “inverse” patterns. Hence, extremity of responses are not in accordance with the scores of personality scales; style and content of responses are separable attributes.

If response extremity was a stable, consistent, and hardly conscious behavioral characteristic, choices based on content would be those not fitting well and resulting in different scale scores. This hypothetically means that respondents may not answer honestly to the items. I assume that manipulated test results would be associated with enhanced response extremity.

Four of the five scales shifted to positive values, while *Emotionality* shifted to lower values. Social desirability effect would be a hypothetical reason of this phenomena. It is presumable that a stereotypic, schematic response style would underlie the attempt to social desirability and the more extreme responding. Both of the above-mentioned hypotheses were investigated in a “test manipulation situation” in the third empirical study.

Response latencies can be explored not only by using absolute temporal properties, but also by ranks of the items. This technique could eliminate confounds arising from the complexity of the text (e.g., the length of sentence), and from individual differences

concerning information processing (e.g., reading speed). As an additional advantage, familiarity effect has a lower influence on these rank indices, hence data obtained from repeated sessions can be easily compared to one another.

Likert-scales of Facet5 showed the expected inverse U-shaped relation; by using rank scores of the items, this pattern is even more emphasized.

The expected inverse U-shaped pattern (Akrami, et al., 2007) is demonstrable in case of scale scores, but only if rank scores are used. As extremity of scale scores increases, mean rank scores of response latency increase as well. Compared to the average response speed of the whole sample, latency of responses given to all personality dimensions becomes shorter, respectively.

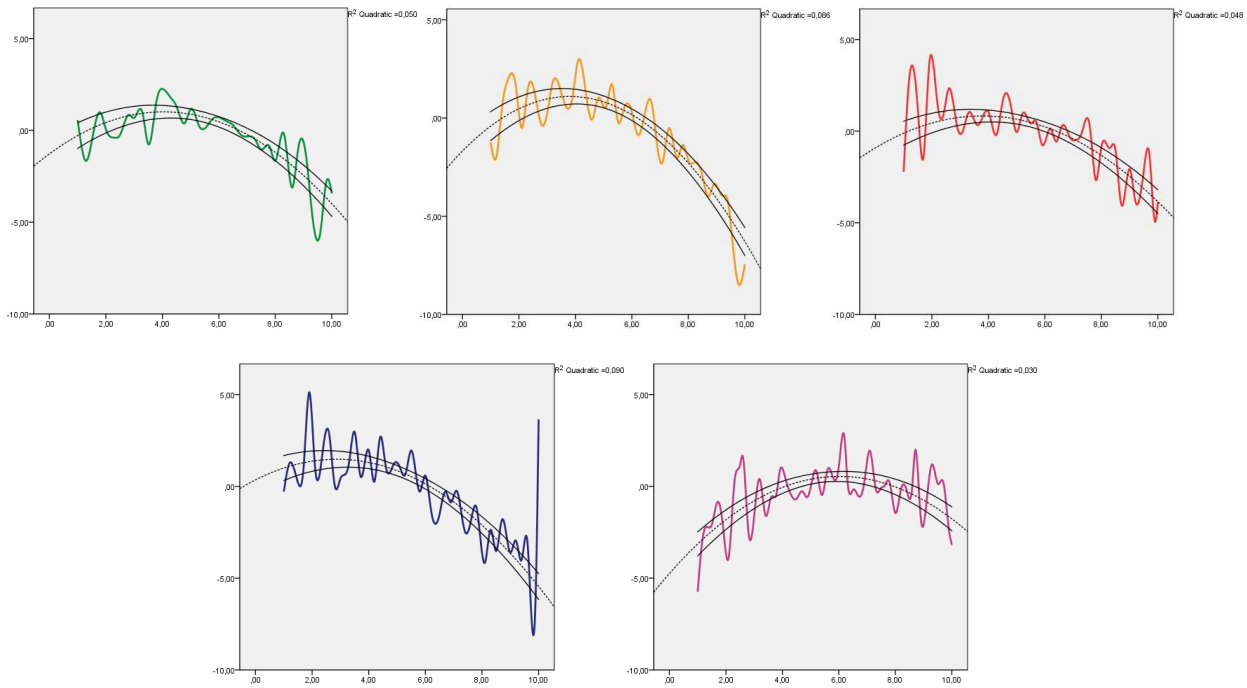

Figure 2. Relation between scale-level temporal response ranks (Axis Y) and the scale scores of Will, Energy, Affection, Control and Emotionality (Axis X)

This inverse U-shaped trend only shows up in the average of a given a scale, but not in the items of other scales. The more a respondent has an average score on a personality dimension, the slower are his/her responses in the given area (i.e., choosing between the two contrasting item). Reduction in response time is related to content, while response extremity is rather a global factor.

In the third empirical study it is assumed that in case of a “manipulated testing session” a more schematic responding yields reduced response latencies.

VI. Empirical study (3): Characteristics of response style in test manipulated situations

Two questions were investigated in this study: the asynchrony between response style and scale scores, and the score shift as a result of social desirability. First, participants were instructed to complete the Facet5 questionnaire in a usual manner (Normal Completion), then they were asked to manipulate the test results. Their task was

to complete the questionnaire according to an ideal personality profile (Ideal Profile), then to “create” a predefined, moderately difficult profile (False Profile). Participants were undergraduate psychology students (N = 128), among them the ratio of females was higher (70%), but the sample was homogeneous in age (M = 20.5 years; SD = 2.2 years). Within the three sequential testing sessions there was one week each time.

According to the results, the score of *Emotionality* was significantly lower in Ideal Profile compared to Normal Completion. However, other scale scores enhanced, sometimes even in the False Profile.

Participants showed more extreme responding in the Ideal Profile and in the False Profile compared to their usual individual response tendency. Flexibility in the variations of response extremity was associated with the success of manipulating test results. Respondents with the least increase in response extremity were those who were the least characterized by effective impression management. Participants who manipulated test results in the most successful manner could be described with the most emphasized response extremity; not in Normal Completion, but in False Profile. Correlations of response extremity within conditions (i.e, the type of completion) was the lowest in the latter group of participants.

Mean ranks of response latency in case of personality scales also changed compared to Normal Completion. This alteration varied in direction and in magnitude as a function of scales, but at most occasions they shifted from the baseline jointly.

Regression models based on response characteristics could account for 20-30 % of variance of the effective manipulation of a certain personality dimension.

In summary, it seems that the stereotypic, schematic response style of attempting to manipulate test results could serve as the reason for the more extreme response tendency, for the shift in scale scores towards social desirability, and for the joint variation within conditions of mean ranks of response latency. At the same time, further studies on larger samples are needed in order to reveal the consistent attempts of manipulation based on response style.

VII. Summary

1. Response-extremity is a stable, consistent personal characteristic that can also be revealed in the case of Hungarian personality measurement.

2. Individual differences in response-extremity can be mapped through response-extremity clusters.

3. Response extremity may distort scores on latent personality dimensions. It is necessary to create standard norms based on response clusters to eliminate this content-independent source of error.

4. Response-extremity and response latency can not be derived from one single personality trait but they can be related to a pattern of Big Five dimensions.

5. Intentional response manipulation is related to schematic and stereotipical response pattern that can be observed in two domains - response tendency becomes more extreme, and scale scores shift to the direction of social desirability.

References

Akrami, N., Hedlund, L. E., & Ekehammar, B. (2007). Personality scale response latencies as self-schema indicators: The inverted-U effect revisited. *Personality and individual differences, 43*(3), 611-618.

Amelang, M., Eisenhut, K., & Rindermann, H. (1991). Responding to adjective check list items: a reaction time analysis. *Personality and individual differences, 12*(6), 523-533.

Austin, E. J., Deary, I. J., & Egan, V. (2006). Individual differences in response scale use: Mixed Rasch modelling of responses to NEO-FFI items. *Personality and individual differences, 40*(6), 1235-1245.

Bachman, J. G., & O'Malley, P. M. (1984). Yea-saying, nay-saying, and going to extremes: Black-white differences in response styles. *Public Opinion Quarterly, 48*(2), 491-509.

Baumgartner, H., & Steenkamp, J. B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research, 38*(2), 143-156.

Billiet, J. B., & Davidov, E. (2008). Testing the stability of an acquiescence style factor behind two interrelated substantive variables in a panel design. *Sociological Methods & Research, 36*(4), 542-562.

Buckley, N., & Williams, R. (2002). Testing on the web - Response patterns and image management. *Selection & Development Review, 18*, 3-8.

Cattell, R. B. (1943). The description of personality: basic traits resolved into clusters. *The Journal of Abnormal and Social Psychology, 38*(4), 476.

Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology, 31*(2), 187-212.

Clarke, I. (2000). Extreme Response Style in Cross-Cultural Research: An Empirical Investigation. *Journal of Social Behavior & Personality, 15*(1), 137-152.

Cronbach, L. J. (1946). Response sets and test validity. *Educational and psychological measurement, 6*(4), 475-494.

De Beuckelaer, A., Weijters, B., & Rutten, A. (2010). Using ad hoc measures for response styles: a cautionary note. *Quality and Quantity, 44*(4), 761-775.

Eid, M., & Rauber, M. (2000). Detecting measurement invariance in organizational surveys. *European Journal of Psychological Assessment, 16*(1), 20.

Eysenck, H. J. (1967). *The Biological Basis of Personality*. New Brunswick, New Jersey: Transaction Publishers.

Feldt, L. S. (1969). A test of the hypothesis that cronbach's alpha or kuder-richardson coefficent twenty is the same for two tests. *Psychometrika, 34*(3), 363-373.

Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sources. *The Journal of Abnormal and Social Psychology, 44*(3), 329.

Gilman, R., Huebner, E. S., Tian, L., Park, N., O'Byrne, J., Schiff, M., Langknecht, H. (2008). Cross-National Adolescent Multidimensional Life Satisfaction Reports: Analyses of Mean Scores and Response Style Differences. *Journal of Youth and Adolescence, 37*(2), 142-154.

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), *Review of personality and social psychology* (Vol. 2, pp. 141-165). Beverly Hills, CA: SAGE.

Greenleaf, E. A. (1992). Measuring extreme response style. *Public Opinion Quarterly, 56*(3), 328-351.

Griffith, R. L., Peterson, M. H., Isaacson, J., Quist, J., & Gammon, A. (2009). *One of these things is not like the other: A person specific taxonomy of the determinants of applicant responding.* Paper presented at the 24th Annual Conference for the Society for Industrial and Organizational Psychology, New Orleans, LA, USA. .

Hamamura, T., Heine, S. J., & Paulhus, D. L. (2008). Cultural differences in response styles: The role of dialectical thinking. *Personality and individual differences, 44*(4), 932-942.

Holden, R. R., Fekken, G. C., & Cotton, D. H. (1991). Assessing psychopathology using structured test-item response latencies. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3*(1), 111.

Holden, R. R., Kroner, D. G., Fekken, G. C., & Popham, S. M. (1992). A model of personality test item response dissimulation. *Journal of Personality and Social Psychology, 63*(2), 272.

Jackson, D. N., & Messick, S. (1958). Content and style in personality assessment. *Psychological Bulletin, 55*(4), 243.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research, 2*, 102-138.

Johnson, T. R. (2003). On the use of heterogeneous thresholds ordinal regression models to account for individual differences in response style. *Psychometrika, 68*(4), 563-583.

Kuiper, N. A. (1981). Convergent Evidence for the Self as a Prototype The" Inverted-U RT Effect" for Self and Other Judgments. *Personality and Social Psychology Bulletin, 7*(3), 438-443.

Littvay, L. (2010). The Genetic Heritability of Survey Response Styles.
McCrae, R. R., Costa Jr, P. T., & Martin, T. A. (2005). The NEO–PI–3: A more readable revised NEO personality inventory. *Journal of personality assessment, 84*(3), 261-270.
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.
McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *The Journal Of Applied Psychology, 85*(5), 812-821.
McFarland, L. A., & Ryan, A. M. (2006). Toward an Integrated Model of Applicant Faking Behavior. *Journal of Applied Social Psychology, 36*(4), 979-1016.
Mirnics, Z. (2006). *A személyiség építőkövei. Típus-, vonás- és biológiai elméletek*. Budapest: Bölcsész Konzorcium.
Mueller, J. H., Thompson, W. B., & Dugan, K. (1986). Trait distinctiveness and accessibility in the self-schema. *Personality and Social Psychology Bulletin, 12*(1), 81-89.
Naemi, B. D., Beal, D. J., & Payne, S. C. (2009). Personality Predictors of Extreme Response Style. *Journal of personality, 77*(1), 261-286.
Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*(6), 660-679.
Paulhus, D. L. (1991). Measurement and control of response bias.
Popham, S. M., & Holden, R. R. (1990). Assessing MMPI constructs through the measurement of response latencies. *Journal of personality assessment*.
Rawling, K., & Reid, J. (2006). *British Psychological Society - Psychological Testing Centre Test Reviews: Facet5 (www.psychtesting.org.uk)*. www.envisiatools.com/products/Personality-test/Facet5/Facet5-Resources/BPSReview.pdf (Letöltés: 2012.okt.)
Rózsa, S., Kő, N., & Oláh, A. (2006). Rekonstruálható-e a Big Five a hazai mintán? A Carparaféle „Big Five Kérdőív" (BFQ) felnőtt változatának hazai adaptációja és nemzetközi összehasonlító elemzése. *Pszichológia, 26*(1), 57–76.
Shoss, M. K., & Strube, M. J. (2011). How do you fake a personality test? An investigation of cognitive models of impression-managed responding. *Organizational Behavior and Human Decision Processes, 116*(1), 163-171.
Szirmák, Z., & De Raad, B. (1994). Személyiségtaxonómia a magyar nyelv személyleíró szókincse. *Magyar Pszichológiai Szemle, 50*(1-2), 39-65.
Tóth, I. G. (2009). Bizalomhiány, normazavarok, igazságtalanságérzet és paternalizmus a magyar társadalom értékszerkezetében. A gazdasági felemelkedés társadalmi-kulturális feltételei című kutatás zárójelentése. Budapest: TÁRKI Társadalomkutatási Intézet Zrt.
van der Linden, W. J. (2009). Conceptual Issues in Response Time Modeling. *Journal of Educational Measurement, 46*(3), 247-272.
Van Herk, H., Poortinga, Y. H., & Verhallen, T. M. M. (2004). Response styles in rating scales. *Journal of Cross-Cultural Psychology, 35*(3), 346.
Watkins, D., & Cheung, S. (1995). Culture, Gender, and Response Bias: An Analysis of Responses to the Self-Description Questionnaire. *Journal of Cross-Cultural Psychology, 26*(5), 490-504.
Weijters, B., Geuens, M., & Schillewaert, N. (2010). The stability of individual response styles. *Psychological Methods, 15*(1), 96-110.