

The effect of speediness on personality questionnaires: an experiment on applicants within a job recruiting procedure

LALE KHORRAMDEL¹ & KLAUS D. KUBINGER

Abstract

The authors conducted an experiment to determine how a particular design of personality questionnaires influences applicant responses on personality scales. A completely crossed 2 x 2 x 2 design was carried out with real-world applicants and individuals in a job application training program in which speed (with or without a time limit), response format (dichotomous or analogue), and instructions (neutral standard instruction or a repeated warning that people who fake can be detected) were manipulated. Two hundred eight participants completed the Myers-Briggs Type Inventory and a German Interpersonal Circumplex (IPC)-based questionnaire. Although providing a warning showed no influence, response format and the interaction between speed and response format showed a significant effect for some scales.

Key words: personality questionnaire, faking good, social desirability, personnel selection, psychological assessment, response format, instruction, speed

¹ Lale Khorramdel and Klaus D. Kubinger, Center of Testing and Consulting, Division of Psychological Assessment and Applied Psychometrics, Faculty of Psychology, University of Vienna. Correspondence concerning this article may be addressed to Lale Khorramdel, University of Vienna, Liebiggasse 5, 1010 Vienna, Austria, Email : lale.khorramdel@univie.ac.at, or Klaus Kubinger, Email: klaus.kubinger@univie.ac.at

Personality questionnaires are the best known and the most popular tools used to measure personality. However, personality questionnaires often show a high transparency; that is, it is often evident to the test-taker what constructs the test measures. Because test-takers can infer what constructs items may measure, they may distort their responses in order to present themselves favourably. This may be particularly problematic in the context of personnel selection, where applicants may “fake good” in an attempt to secure a job offer (cp. Kanning & Holling, 2001; Karner, 1999, 2002).

Considerable research has shown that even voluntary participants are able to intentionally fake good when instructed to empathize with a selection candidate (Kubinger, 1996; 2002) or to adapt to a given job profile (Hoeth, Büttel, & Feyerabend, 1967; Lammers & Frankenfeld, 1999). Krahe and Hermann (2003) found similar results when analysing the susceptibility of the NEO-Five Factor Inventory (NEO-FFI) to systematic response tendencies. Because of these potential faking effects, data from self-descriptions should always be regarded carefully (Deller & Kuehn, 2003).

Faking tendencies in real-world selection situations, however, are actually fewer than in simulated situations. Some studies show that adjusting personality scores based on social desirability scores does not decrease the validity of a test (Hough et al., 1990; Moorman & Podsakoff, 1992; Ones, Viswesvaran & Reiss, 1996; Ones, Viswesvaran & Schmidt, 1993), and there is even an established opinion that personality questionnaires are valid methods for personnel selection despite their high transparency (Schmidt & Hunter, 1998; cf. also Marcus, 2003). However, the extent to which validity is decreased by the influence of social desirability bias is unknown (Kanning, 2003). Furthermore, because candidates who fake are more likely to be selected than those who answer honestly, faking may make selection systems unfair (Ellingson, Sackett & Hough, 1999; Hough, 1998). Therefore, test-users should take precautions to prevent or reduce applicant faking on personality questionnaires (Hough & Ones, 2002; McFarland, 2003).

Past research has explored whether it is possible to detect individuals who may be faking. Two means of detection have primarily been used: measuring/analysing response latencies (i.e., the time between item responses; Esser & Schneider, 1998; Holden & Hibbs, 1995; Holden, Kroner, Fekken & Popham, 1992; Hsu, Santelli & Hsu, 1989; Kuntz, 1974; Robie et al., 2000; Schneider & Hübner, 1980) and imbedding social desirability scales (a.k.a., lie scales) within personality measures (Crowne & Marlowe, 1960; Edwards, 1957; Hoeth, Büttel & Feyerabend, 1967; Paulhus, 1991; Schneider-Düker & Schneider, 1977). In the detection literature using response latencies, the general assumption is that response latencies indicate the fidelity of the response. Response latencies may indicate whether a participant's response reflects their self-concept (i.e., an honest response) or a response style (i.e., a faked response). In addition, response latencies may indicate that a test-taker has responded at random (which would affect the reliability and validity of a score; cp. Wagner-Menghin, 2002). Holden and colleagues (Holden & Hibbs, 1995; Holden, Kroner, Fekken, & Popham, 1992) proposed a model of personality test item response dissimulation. In their model, a respondent attempts to compare test item content to either a relevant cognitive self-schema or to an adopted schema (for example, faking good represents an adopted schema). They found that responses congruent with self-schemas are faster than when answers are not congruent with self-schemas. These results suggest that adopting the schema to fake good may produce longer response latencies, which can be used to identify fakers. However, the authors acknowledge a number of limitations in their research. First, their research has fo-

cused on dimensions of maladjustment rather than personality scales. Second, longer response latencies are associated with items that have relatively extreme social desirability levels, have extreme endorsement proportions, and are predominately positively-keyed (rather than reverse-keyed). Third, their design compared the response latencies of volunteer participants instructed to fake versus to volunteer participants instructed to answer honestly. Indeed, the results of other response latency studies have produced discrepant results. Kuntz (1974) found significantly longer latencies under both “fake bad” and “fake good” instructions than under standard conditions. In contrast, Hsu, Santelli and Hsu (1989) found shorter latencies under both faking conditions.

The benefits of social desirability (lie-) scales or repeated items to check for consistency (control items) are debatable, mostly, because of their high transparency (Seiwald, 2003). Even control items, if recognized, decrease the participant’s motivation to answer honestly. Moreover, the validity of lie scales seems doubtful, because they measure not only the participant’s tendency to fake but also a personality trait. Attempts to use social desirability scales to statistically adjust personality scores decreases rather than increases the criterion-related validity of the personality measures (Borkenau & Ostendorf, 1992; McCrae & Costa, 1983; Ones et al., 1996; Piedmont et al., 2000; cf. also Hülshager et al., 2004). Hülshager et al. (2004) argue that even the attempt to identify and exclude invalid profiles with these scales is a poor strategy because honestly-responding participants may also be erroneously eliminated. Test-takers who have high social desirability scores are not necessarily faking; indeed, high-scorers might simply have answered the questionnaire honestly but have a high degree of the trait measured by social desirability scales. Thus, correcting questionnaires for faking is particularly concerning given that such corrections may have considerable influence on who receives a job offer, yet there is an absence of empirical evidence to support the use of these corrections (Goffin & Christiansen, 2003).

Efforts to suppress faking good

Recent attempts have aimed at making personality questionnaires less fakeable by adjusting aspects of how they are administered. These attempts include adjusting the response format, method of administration, and item positioning. In a summary of this research, no general conclusions could be drawn for these adjustments as the effects appear to be influenced by many moderating variables (Kubinger, 2003a). However, some research suggests that using analogue scales (in which participants mark along a continuous line to indicate the extent of their agreement) as a response format may be less prone to faking than a dichotomous response, multiple-choice, or Q-Sort format (cf. Seiwald, 2002). Questionnaires administered with either paper and pencil or with a computer have not shown any difference in fakeability, nor have verbal as opposed to non-verbal questionnaires (cf. Amelang, Schäfer and Yousfi, 2002). With respect to the effects of the item-positions, it has been shown that test-takers are more likely to fake their answers at the beginning rather than at the end of a questionnaire.

In addition, researchers have attempted to limit faking by adjusting the instructions given to test-takers (Mummendey, 1999). Typically, personality measures are administered with the instruction to answer “as candidly and honestly as possible.” A warning instruction goes beyond this by informing test-takers that the test administrator can detect intentional re-

sponse distortion. Hülshager et al. (2004) found no effect of the warning that “untruthful” response patterns can be detected; however, their sample consisted only of student volunteers. Most research, though, has suggested that these warnings are effective at reducing the prevalence of faking, although the effects are weak (Hoeth & Köbler, 1967; Braun & La Faro, 1968; Dwight & Donovan, 2003). In particular, warning applicants that faking can be detected with an imbedded social desirability scale or by analyzing the response latencies decreases faking (Doll, 1971; Kluger & Colella, 1993; Nias, 1972; Robie et al., 2000; Wheeler, Hamill & Tippins, 1996). However, McFarland (2003) calls attention to certain practical consequences of this procedure. Warning the applicants in personality questionnaires tends to provoke negative reactions by the applicants (e.g., Rosse, Miller & Stecher, 1994; Smither et al., 1993; Steiner & Gilliland, 1996). Hence, such a warning may make applicants feel that the employer distrusts them or that applicants cannot present themselves as they would like to be seen. On one hand, this negative reaction may make the most qualified applicants self-select themselves out of the selection process (cf. Ryan, Sacco, McFarland & Kriska, 2000). On the other hand, the applicants’ test-taking motivation could be decreased. Hence, such test perceptions may affect selection decisions (Chan et al., 1997), and the validity of the questionnaire (Schmit & Ryan, 1992). Overall, McFarland did not find any negative reactions affected by warnings in her study; however, her sample was limited to voluntary participants who were instructed to imagine a job-application situation.

An indirect method to decrease faking is to exert a time pressure on a participant. Some authors assert that test-takers need more time in order to fake, so adding time pressure may decrease faking tendencies. For example, Bartley (1958) suggested that overly long reaction times are caused by a test-taker searching for substitute responses that will mask his/her initial reaction. Answering without manipulating one’s own attitudes takes less time than reflecting some prototypic attitudes in responses. Indeed, the most prior studies support the idea faking causes longer response times; but these have used inadequate samples (such as volunteers rather than job applicants). On the other hand, in an early study, Sutherland and Spilka (1964) have demonstrated that time pressure can result in responses in the direction of what is socially approved. However, they used voluntary students as participants, as well, and the decision for a response interval of 2 seconds per item is neither explained nor evident. The results of Krämer and Schneider (1987) are similar, but, instead of using real time pressure, the participants were only given the instruction to answer quickly and spontaneously. Again, only volunteers served as participants, and, furthermore, the sample was considerably small. Neubauer and Malle (1997) likewise use a speed instruction, and the results show a lower mean neuroticism score in the Eysenck-Personality-Inventory (EPI). Again, however, the study is based on volunteers, and there was no real time pressure.

Aim of the current study

Although the results of some previous studies aiming to reduce applicant faking have been encouraging (particularly the analogue scale), the majority have primarily used non-applicant volunteer samples. The present experiment uses real-world applicants to test the effects on faking of manipulating speed, response format, and instruction.

Hypotheses

The first hypothesis is that a limited response time may decrease the phenomenon of faking good in real-world selection situations. We expect this decrease in faking because the time pressure will preclude the test-taker from thinking about the best (most socially desirable) answer and force the test-taker to answer spontaneously. The second hypothesis is that an analogue scale response format might lead to a more honest self-presentation than would a dichotomous one. We expect that an analogue scale is more difficult to fake because it is more difficult to determine what intensity is socially desired but not suspect of indicating a faked response. The third hypothesis is that a warning instruction may lead to less socially desirable answers.

To detect possible fakers within the scales of the personality questionnaire used, we assume that it is more advantageous for a participant in selection situations to show high values for certain scales and low values for other scales. We expect that fakers would inflate their scores on the scales Self-Concept of Own Competences, Internality, Extroversion, and Thinking, whereas we expect that fakers would provide lower ratings for the scales Powerful Others Control, Chance Control, and Introversion (a description of the personality tests used is given below). Test-takers with such scores might try to present themselves in a socially desirable manner. These predictions are guided by what is commonly believed to be desirable in the jobs for which the applicants in our sample applied (office managers, salesmen/women, tradesmen/-women, and middle echelon managers). For the scales Feeling, Judging, Perceiving, Intuition, and Sensing, it is not clear whether high or low values would indicate faking because the scales are not clearly indicative of desirable traits for these specific jobs.

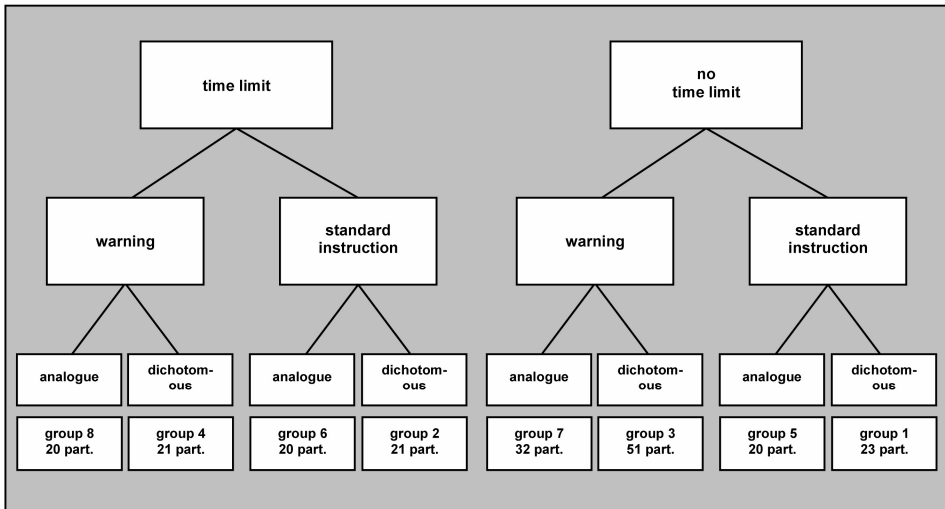
Method

Design

To test whether time limitations, scale response format, and warning instructions affect applicant faking, participants completed personality measures in a completely crossed 2 x 2 x 2 design. Thus, participants were randomly assigned to one of eight groups, representing a combination of the three factors, as shown in Figure 1. Figure 1 also provides the information about the sample size in each of the eight groups. Manipulations of each independent variable are described below.

Response Time and Speed. As is often the case with ability and achievement measures, we constructed a speeded questionnaire. The participants either received an overall limited response time for the items (per page of items) or they received no time limit. The time limitations were based on the results of a pilot test with 10 participants between the age of 18 and 56 who came from different educational backgrounds (primary and secondary education). Each of the 10 participants completed the personality questionnaires as quickly as possible. The time per page each participant needed was recorded, and the means were calculated. Thus, we chose time limits between 45 seconds and 1 minute and 40 seconds for each page of the questionnaire (depending on the number of items per page). Three filler items were presented at the end of each page, which served to guarantee that the time limit

Figure 1:
Experimental design



Note: part. = participants

did not preclude slower test-takers from completing all relevant personality items. Test-takers were instructed to start with the first item on each page, and not to leave a single item out.

Item Response Format. As in previous experiments, participants completed personality questionnaires with either a dichotomous response format or an analogue scale response format. That is, participants with a dichotomous response format decided between total agreement to the given statement of an item or total disagreement (“yes“ or “no,“ and “true” or “false”); participants with the analogue scale response format indicated their response by making a mark on a continuous line. There were at least 39 invisible points on a line in accordance with the length of the 39 mm line, a length based on the layout of the answer sheet. However, the analogue scale was scored dichotomously, so that marks on the left half of the line indicated “true/agree” and marks on the right half indicated “false/disagree.”

Instruction. All participants received a conventional, neutral instruction. Those in the warning group also received an additional warning that faking can be detected. The warning was given once at the beginning of the questionnaire and once again in the middle of the questionnaire to ensure that the participants did not forget. The standard instruction was:

There are no correct or incorrect answers in this questionnaire. Your answers merely provide information on how you see things or how you normally make a decision.

The additional warning instruction was:

Afterwards your answers will be checked by a complex computer-based evaluation programme, in order to ascertain whether your answers are given in an honest manner. Therefore, it does not pay off to fake the questions. You would then simply be asked to answer the questionnaire again.

Measures

It seemed important that the questionnaire be neither too short (to avoid giving the impression that little effort was required), nor too long (to avoid making participants fatigued or frustrated). The final questionnaire set was a battery of well-known paper-pencil personality questionnaires.

Myers-Briggs Type Indicator (MBTI) – German edition. The MBTI (Bents & Blank, 1991) is based on Jung’s personality typology and consists of 90 items. The scales are: Extroversion, Introversion; Sensing, Intuition; Thinking, Feeling; Judging, Perceiving. Descriptions of these scales are provided in Table 1. Because the fifth item of the MBTI originally has three response categories, it was necessary to remove the middle category to create a dichotomous response format for this item. Reliability coefficients were calculated for each scale and range from .007 to .639 (see Table 2).²

Interpersonal Circumplex (IPC)-based questionnaire. The German IPC-based questionnaire (FKK; Krampen, 1991) consists of 32 items that measure the “locus of control of reinforcement” concept from J.B. Rotter’s social learning theory of personality (Rotter, 1982; Rotter, Chance & Phares, 1972; Rotter, Seeman & Liverant, 1962). The scales are: Self-Concept of Own Competences, Internality, Powerful Others Control, and Chance Control. Descriptions of these scales are provided in Table 1. The items from the FKK also originally had six response categories; therefore, only the categories “right” and “wrong” were used. The split-half reliabilities of the FKK scales presented in its test-manual range from .63 to .79 (see Table 3).

As explained before, we added three additional items at the end of each page of the questionnaire – altogether 42 items – to guarantee that almost all interesting items (the items of the MBTI and the FKK) are actually answered by each participant, despite having a time limit. These filler items were not analysed. In addition, neither measure originally had an analogue scale. Therefore, we had to establish one. Altogether, the resulting questionnaire consisted of 164 items, of which 122 were actually analysed (omitting the 42 filler items). This questionnaire was presented either with a dichotomous response format or an analogue scale.

² We thank an anonymous reviewer for suggesting that we note the links between all the scales and the Big Five dimensions of personality. Because there is no empirical evidence of such links between the FKK and the Big Five dimensions, we simply give some plausible correspondence in a separate column. However, the links between the MBTI and the Big Five dimensions are based on the findings of McCrae and Costa (1989), which were supported by the findings of Furnham, Moutafi and Crump (2003).

Table 1:
Description of the scales of the personality questionnaires

Scales	Description	Correspondence to the Big Five dimensions of personality (McCrae & Costa, 1989; Furnham, Moutafi, & Crump, 2003)
MBTI:		
extroversion	external orientation, extrovert attitude	extroversion (E)
introversion	internal orientation, introvert attitude	extroversion (E)
sensing	sensual perception; perceptual processes by means of the five senses; orientation on experiences in the present (here and now)	openness to experience (O)
intuition	intuitive perception; perception of possibilities, meanings and relations which happens by insight	openness to experience (O)
thinking	analytical judgment; judgment due to logical linked imaginations	agreeableness (A)
feeling	emotional judgment; judgment due to personal and social values	agreeableness (A)
judging	judging attitude; focus on decisions and planning of action sequences	conscientiousness (C)
perceiving	perceptual attitude; focus on receipt and perception of information	conscientiousness (C)
German IPC-based questionnaire (FKK):		Plausible correspondence to the Big Five dimensions of personality
self-concept of own competences	generalized expectation to have action possibilities – at least one – at disposal in life of action situations	neuroticism (N)
internality	subjectively noticed control of own life and events of the person specific environment	openness to experience (O)
powerful others control	generalized expectation that important events in life depend on the influence of others	agreeableness (A)
chance control	generalized expectation that life and important events in it depend on destiny, fortune, bad luck and chance	neuroticism (N)

Table 2:
Cronbachs alpha and split-half reliability (Spearman-Brown) of the MBTI-scales
(from the current data; n = 208)

MBTI-scales	<i>Cronbachs Alpha</i>	<i>Spearman-Brown</i>
extroversion	.268	.244
introversion	.322	.348
sensing	.274	.387
intuition	.007	.188
thinking	.074	.040
feeling	.047	.146
judging	.631	.639
perceiving	.639	.608

Table 3:
Split-half reliability (Spearman-Brown) of the FKK-scales (Krampen, 1991)

Study	N	self-concept of own competences	internality	powerful others control	chance control
1	62	.79	.74	.72	.73
2	258	.70	.67	.75	.78
3	152	.72	.63	.65	.67
4	38	.71	.64	.70	.76
5	248	.72	.68	.70	.69
6	2028	.71	.64	.67	.70

Sample

Two hundred eight participants completed the personality questionnaire. Initially, we intended to use only test-takers who were actual job applicants being recruited. Participant data was gathered from two separate sources. First, 113 of the participants were recruited from a special job-application training course consisting of long-term unemployed individuals within a re-education programme. Near the end of this programme, participants were assessed as part of the course training to prepare for real-world job-applications. That is, for this group of participants, the questionnaire was administered as a simulated-selection process. This testing was part of an evaluation of the effects of the training programme. The participants received personal feedback regarding their individual results. Second, 95 participants were real-world job applicants whose data were taken from two personnel and management consulting companies.

Altogether, 96 women and 112 between the age of 18 and 56 with various educational backgrounds were tested by seven test instructors. These instructors had received an exact verbal instruction and written guide describing how to instruct the participants. The test-takers were randomly assigned to the eight experimental conditions; however, their sex and original institution were noted to ensure that men, women, and participants of each of the three institutions are represented adequately in each of the 8 experimental groups.

Because data collection occurred as an on-going process at different locations, it was not possible to assign exactly the same number of participants to each group. That is, we had to conform with the routine of the different institutions. As a result, 116 participants filled out the questionnaire with the dichotomous response format, and 92 participants filled out the questionnaire with the analogue scale response format. Eighty-two participants were given the questionnaire without any time limit, and 126 participants were given the questionnaire with the time limit described above. Eighty-four participants were given only the standard instruction, and 124 were additionally given the warning instruction. The participants of the special training course were tested in groups (with a maximum of 15 persons per group), whereas the job applicant sample was tested individually.

Results

The means and standard deviations for all scales in each experimental condition are given in Table 4. In addition, Appendix A shows the intercorrelations between all scales for the given data.

A multivariate analysis of variance (MANOVA) was used to compare means across conditions ($\alpha = .05$). With a sample size of $208/8=26$ and $\alpha = .05$, a MANOVA has adequate power (.80) to detect a mean difference of 2/3 standard deviations (Rasch & Kubinger, 2006). To test the homogeneity of variance across cells, a Levene's test was calculated for each scale. However, four MBTI scales (Extroversion, Introversion, Thinking, and Feeling) failed the Levene's test for homogeneity of variance ($p = .047, .006, .018, \text{ and } .013$, respectively). Hence, only the scales Judging, Perceiving, Intuition, and Sensing from the MBTI and the scales Self-Concept of Own Competences, Internality, Powerful Others Control, and Chance Control from the FKK are taken into consideration in the following. Box's M-Test for testing the homogeneity of the variance-covariance matrix was conducted on the remaining scales but was not significant ($p = .239$), indicating that the resulting F -values of multivariate analysis of variance may be interpreted. Table 5 shows the results of the MANOVA testing the main and the interaction effects of the three factors. The multivariate analysis of variance shows a significant effect of the response format (dichotomous vs. analogue), and a significant interaction effect between response format and time (limited time vs. unlimited time).

To more clearly understand these significant effects, each scale was considered individually. Tables 6 and 7 present the result of univariate factorial ANOVAs, the associated effect sizes, and their confidence intervals. Either a (significantly) higher or lower score (dependent on the meaning of the particular scales) may be interpreted as indicating faking good.

Table 6 and 7 show that only one scale (Internality) showed significantly different means across response format groups ($p = .038$). However, for the time x response format interaction, Self-Concept of Own Competences ($p < .001$), Powerful Others Control" ($p = .008$), and Chance Control ($p < .001$) showed significant interaction effects. To determine the direction of significant differences, we examined the means of the scores (see Tables 8 and 9).

Table 4:
Means and standard deviations (SD) for all scales by experimental condition
(8 experimental groups)

Scale	No Warning				Warning			
	Dichotomous		Analogue		Dichotomous		Analogue	
	No Lim.	Lim.	No Lim.	Lim.	No Lim.	Lim.	No Lim.	Lim.
extroversion	16.77 (5.84)	14.86 (6.03)	13.15 (4.40)	16.30 (5.24)	15.41 (4.55)	15.43 (7.26)	15.66 (5.15)	14.68 (6.58)
introversion	7.82 (5.71)	9.67 (5.65)	11.55 (4.15)	8.55 (4.89)	9.29 (4.36)	9.43 (6.79)	8.88 (4.97)	9.68 (6.57)
sensing	10.05 (3.51)	10.05 (4.82)	12.50 (4.03)	11.40 (3.17)	11.65 (4.25)	9.95 (4.91)	10.88 (4.08)	10.68 (5.94)
intuition	7.14 (2.98)	7.57 (3.63)	6.50 (2.96)	6.35 (2.56)	6.33 (3.06)	7.67 (3.92)	7.25 (3.37)	7.58 (4.43)
thinking	10.45 (2.92)	9.57 (4.37)	10.00 (2.25)	11.05 (2.50)	9.43 (3.53)	9.43 (2.68)	8.53 (3.11)	9.47 (3.01)
feeling	6.95 (2.57)	7.52 (3.50)	7.25 (1.92)	6.10 (2.36)	7.61 (3.02)	7.86 (2.13)	8.47 (2.86)	8.11 (2.47)
judging	10.73 (2.99)	11.19 (3.46)	10.55 (2.82)	10.70 (2.56)	10.63 (3.30)	10.19 (3.68)	10.00 (3.21)	10.89 (3.43)
perceiving	6.14 (3.96)	6.24 (3.83)	7.10 (3.67)	6.15 (3.57)	6.75 (4.02)	7.10 (4.33)	7.22 (4.29)	6.32 (4.22)
self-concept of own comp.	1.18 (1.87)	2.76 (2.43)	2.65 (1.81)	1.40 (1.27)	2.25 (2.07)	3.05 (3.13)	2.75 (1.97)	1.68 (1.70)

Note. Within each row, means are presented above and standard deviations presented below in parentheses.

The additional warning instruction not to fake did not influence the level of the score (see Table 5). However, the kind of response format does have an influence (see Table 6), although only for a single scale, Internality. Table 8 shows that if the dichotomous response format is used, a slightly higher tendency towards Internality is exhibited than when an analogue scale response format.

Table 5:

Multivariate Analysis of Variance – including the scales “judging”, “perceiving”, “intuition”, and “sensing” from the MBTI and the scales “self-concept of own competences”, “internality”, “powerful others control”, and “chance control” from the IPC-like questionnaire (FKK)

Effect		Value	<i>F</i>	Hypothesis <i>df</i>	Error <i>df</i>	Sig.	Partial Eta Squared
format	Pillai's Trace	.088	2.307	8	191	.022	.088
instruction	Pillai's Trace	.022	.526	8	191	.836	
time	Pillai's Trace	.019	.465	8	191	.880	
format * instruction	Pillai's Trace	.060	1.528	8	191	.150	
format * time	Pillai's Trace	.181	5.277	8	191	.000	.181
instruction * time	Pillai's Trace	.044	1.095	8	191	.368	
format * instruction * time	Pillai's Trace	.071	1.826	8	191	.074	

Table 6:

Scale-wise F-values with respect to the factor Format

Dependent Variable	<i>df</i>	<i>F</i>	Sig.	Effect Size	Confidence-Interval	
					<i>lower bound</i>	<i>upper bound</i>
judging	1	.098	.755			
perceiving	1	.058	.809			
intuition	1	.274	.601			
sensing	1	2.171	.142			
self-concept of own competences	1	.439	.508			
internality	1	4.343	.038	.3444	.0674	.6214
powerful others control	1	2.845	.093			
chance control	1	.600	.440			

Table 7:

Scale-wise F-values with respect to the factor Format x Time

Dependent Variable	<i>df</i>	<i>F</i>	Sig.	Effect Size	Confidence-Interval	
					<i>lower bound</i>	<i>upper bound</i>
judging	1	.291	.590			
perceiving	1	.956	.329			
intuition	1	.654	.420			
sensing	1	.025	.875			
self-concept of own competences	1	16.614	.000	.2191	-.1707	.6089
internality	1	.195	.660			
powerful others control	1	7.120	.008	.3402	-.0693	.7497
chance control	1	33.696	.000	.0955	-.3114	.5023

Table 8:
Means of scores at the different levels of the significant factor "Format"

Dependent Variable	Format	Mean
internality	dichotomous	2.297
	analogue scale	1.853

Table 9:
Means of scores at the different levels of the significant interactions of factor "Format"
and factor "Time"

Dependent Variable	Format	Time	Mean
self-concept of own competences	dichotomous	no time limit	1.718
		time limit	2.905
	analogue scale	no time limit	2.700
		time limit	1.542
powerful others control	dichotomous	no time limit	5.901
		time limit	5.333
	analogue scale	no time limit	4.659
		time limit	5.613
chance control	dichotomous	no time limit	5.921
		time limit	4.333
	analogue scale	no time limit	4.009
		time limit	5.795

Interpretation

Furthermore, three of the eight scales show a significant interaction effect between response format and time pressure (see Table 7). For the scales Powerful Others Control and Chance Control, the dichotomous response format produced a higher mean when paired with no time limit, whereas the analogue scale response format produced a higher mean when paired with a time limit. On the contrary, the scale Self-Concept of Own Competences, showed the opposite pattern (higher means were observed when the dichotomous format was paired with a time limit and when the analogue format was paired with no time limit). Supposing the participants think of a high degree of Self-Concept of Own Competences as being highly socially desirable and being low on convincement of external control, these time x response format effects point in the same direction: the answers to personality questionnaires are faked good if there is either a limited administration time and a dichotomous response format or if there is no limited administration time and an analogue scale response format. That is, participants were more likely to give a socially desirable answer if the time was limited on a dichotomous scale, and they were also able to answer in a more socially desirable manner if they had plenty of time to respond to an analogue scale.

Again, there is additionally an occasional trend of answering in a socially desirable manner if a dichotomous response format is used (see Table 8); this is true in so far as high In-

ternality is socially desirable. However, there is no main effect for time pressure (see Table 5).

Discussion

The current experiment used applicants from a job recruiting procedure to examine the effects of time pressure, response format, and warnings on faking. Forty-six percent of the sample consists of such job applicants, and 54% of the participants came from a special job-application training course. Of past research examining the effects of such administrative adjustments on personality questionnaire faking, very few studies have been designed like our experiment. That is to say, the current study does not apply a faking good instruction and test the respective effect only on volunteers. Nor is it an experiment that considers different administration conditions using only volunteers. Instead, the current study examines how these adjustments affect personality scale responses by actual job applicants. Furthermore, two factors of potential inflationary influence on item responses (good) are considered here: the influence of a warning instruction on the one hand, and the influence of a speeded administration on the other hand. Finally, the benefit of using an analogue scale response format was investigated once again, because evaluations so far have not disclosed any unequivocal results.

Although the observed results do not provide a clear and consistent method of administering personality measures that prevents applicant faking, the results do suggest several conclusions. There is some evidence that an analogue scale response format tends to be superior to a dichotomous response format if the psychologist is aware that the faking good phenomenon might occur. However, using an analogue scale response format is not in any way a guarantee for preventing faking good; this response format probably works in only a few of the conceivable personality questionnaires' scales. Furthermore, the response format effect of the analogue scale might be enhanced by imposing a time limit for answering the questionnaire items, whereas the dichotomous response format shows the same effect without imposing a time limit. Indeed, three of the eight scales tested produced such an interaction effect (cp. Table 7).

Again, no means have been discovered that prevent faking good in any case of personality scales. Unfortunately, warning applicants that faking can be detected did not work at all. Future research should investigate the effect of imposing a time limit for each individual item rather than for a set of items. This manipulation may make future results even more pronounced than those observed here. In addition, future research should identify a specific class of personality dimensions and personality questionnaires for which the analogue scale response format works to prevent faking good.

The current study has several limitations that future research should address. First, our sample was not a homogeneous set of job applicants; rather it contained both real job applicants and unemployed individuals participating in a job application training programme that simulated applying for a job. However, the latter group did actually carry out a job-application and had the possibility of experiencing how they would have performed in a real-world selection situation. Furthermore, these individuals had recently been through a similar selection procedure, making a real applicant setting salient. Thus, these factors suggest that these individuals completed the personality measure in a setting quite similar to real-world

job candidates. Nonetheless, there is still the possibility that they may not have had the motivation to distort their responses, or their distortions may be based on generic social desirability of traits rather than traits relevant for a particular job (as would be the case for actual applicants).

Second, it is unclear whether limiting response time (speed testing) changes the constructs being measured by a personality questionnaire. For example, such time limitations may mean the questionnaire also measures the ability to work under pressure, the ability to cope with stress, the motivation to deliver exceptional performance (i.e. trying to be as fast as possible), or the ability to understand the contents of a question quickly. Hence, certain participants may have been handicapped--namely, those who would need more time to understand the meaning of a question and who therefore may have not been able to answer in a manner representing their real behaviour or attitudes. In this case, individuals may have misunderstood items or answered at random. Thus, future studies should also conduct pre-tests of verbal comprehension and examine the baseline reaction times.

References

- Allport, F.H. (1920). The influence of the group upon association and thought. *Journal of Experimental Psychology*, 3, 159-182.
- Amelang, M., Schäfer, A., & Yousfi, S. (2002). Comparing verbal and non-verbal personality scales: Investigating the reliability and validity, the influence of social desirability, and the effects of fake good instructions. *Psychologische Beiträge [latter: Psychology Science]*, 44, 24-41.
- Bartley, S.H. (1958). *Principles of perception*. New York: Harper & Brothers.
- Bents, R., & Blank, R. (1991). *Der Myers-Briggs Typenindikator (MBTI) [Myers- Briggs Type Indicator]*. Weinheim: Beltz.
- Borkenau, P., & Ostendorf, F. (1992). Social desirability scales as moderator and suppressor variables. *European Journal of Personality*, 6, 199-214.
- Braun, J.R., & La Faro, D. (1968). Effects of salesman faking instructions on the Contact Personality Factor Test. *Psychological Reports*, 22, 1245-1248.
- Brehm, J.W. (1966). *A theory of psychological reactance*. New York: Academic Press.
- Chan, D., Schmitt, N., DeShon, R.P., Clause, C.S., & Delbridge, K. (1997). Reactions to cognitive ability tests: The relationships between race, test performance, face validity perceptions, and test-taking motivation. *Journal of Applied Psychology*, 82, 300-310.
- Crowne, D.P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24, 349-354.
- Deller, J., & Kuehn, M. (2003). Occupational Personality Questionnaire (opq32). In E. Fay (Hg.), *Tests unter der Lupe 4: Aktuelle psychologische Testverfahren – kritisch betrachtet* (pp. 76-104). Göttingen: Vandenhoeck & Ruprecht.
- Doll, R.E. (1971). Item susceptibility to attempted faking as related to item characteristics and adopted fake set. *Journal of Psychology*, 77, 9-16.
- Dwight, S.A., & Donovan, J.J. (2003). Do warnings not to fake reduce faking? *Human Performance*, 16, 1-23.
- Edwards, A.L. (1957). *The social desirability variable in personality assessment and research*. New York, NY: Dryden Press.

- 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.
- Esser, C., & Schneider, J.F. (1998). Differentielle Reaktionslatenzzeiten beim Bearbeiten von Persönlichkeitsfragebogen als möglicher Indikator für Verfälschungstendenzen [Differential response latencies by answering personality questionnaires as possible indicator for faking-tendencies]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 19, 246-257.
- Flavell, J.H., Juris, D., Feinberg, L.D., & Budin, W. (1958). A microgenetic approach to word association. *Journal of Abnormal and Social Psychology*, 57, 1-7.
- Goffin, R.D., & Christiansen, N.D. (2003). Correcting Personality Tests for Faking: A Review of Popular Tests and an Initial Survey of Researchers. *International Journal of Selection and Assessment*, 11, 340-344.
- Hoeth, F., Büttel, R., & Feyerabend, H. (1967). Experimentelle Untersuchungen zur Validität von Persönlichkeitsfragebogen [Experimental investigation on the validity of personality questionnaires]. *Psychologische Rundschau*, 18, 169-184.
- Hoeth, F., & Köbler, V. (1967). Zusatzinstruktion gegen Verfälschungstendenzen bei der Beantwortung von Persönlichkeitsfragebogen [Additional instruction against faking-tendencies on answering personality questionnaires]. *Diagnostica*, 13, 117-130.
- Holden, R., & Hibbs, N. (1995). Increment validity of response latencies for detecting fakers on a personality test. *Journal of Research in Personality*, 3, 362-372.
- 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, 272-279.
- Hough, L.M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance*, 11, 209-244.
- 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, 581-595.
- Hough, L.M., & Ones, D.S. (2002). The structure, measurement, validity, and use of personality variable in industrial, work, and organizational psychology. In N. Anderson, D.S. Ones, H.K. Sinangil & C. Viswesvaran (eds), *International handbook of work and organizational psychology*. Sage Publications.
- Hsu, L.M., Santelli, J., & Hsu, J.R. (1989). Faking detection validity and incremental validity of response latencies to MMPI Subtle and Obvious item. *Journal of Personality assessment*, 53(2), 278-295.
- Hülshager, U.R., Spinath, F.M., Küppers, A., & Etzel, S. (2004). Experimentelle Untersuchung zweier Methoden zur Reduzierung Sozialer Erwünschtheit in einem computergestützten eignungsdiagnostischen Testverfahren [Experimental investigation of two methods for reduction of social desirability within a computer based diagnostic test of qualification]. *Zeitschrift für Personalpsychologie*, 3(1), 24-33.
- Kanning, U.P. (2003). Sieben Anmerkungen zum Problem der Selbstdarstellung in der Personalauswahl [Seven notes on the Problem of self-projection in personnel selection]. *Zeitschrift für Personalpsychologie*, 2(4), 193-197.
- Kanning, U.P., & Holling, H. (2001). Struktur, Reliabilität und Validität des NEO-FFI in einer Personalauswahlsituation [Structure, reliability and validity of the NEO-FFI in situations of personnel selection]. *Zeitschrift für Differentielle und Diagnostische Psychologie*. 22, 239-247.
- Karner, T. (1993). Eine empirische Anwendung des Modells von Müller für kontinuierliche Antwortskalen (mittels des computerisierten Myers-Briggs Typenindikators) [An empirical

- use of the model of Müller for continuous response scales (by means of the computer based Myers-Briggs Type Indicator)]. Unpublished degree dissertation, University of Vienna, Vienna.
- Karner, T. (1999). Eine systematische Untersuchung der Auswirkungen verschiedener Antwortmodi auf die Qualität Psychologischer Fragebögen [A systematical investigation of the effect of different response modes on the quality of psychological questionnaires]. Unpublished doctoral dissertation, University of Vienna, Vienna.
- Karner, T. (2002). The volunteer effect of answering personality questionnaires. *Psychologische Beiträge* [later: *Psychology Science*], 44, 42-49.
- Kluger, A.N., & Colella, A. (1993). Beyond the mean bias: The effect of warning against faking on biodata item variances. *Personnel Psychology*, 46, 763-780.
- Krämer, H.J., & Schneider, J.F. (1987). Validität von Fragebogendaten in Abhängigkeit von Antwortzeit-Instruktion und der intraindividuellen Variabilität der Probanden [Validity of questionnaires dependent on response-instruction and intraindividual variability of participants]. *Psychologische Beiträge* [later: *Psychology Science*], 29, 458-468.
- Krahé, B., & Herrmann, J. (2003). Verfälschungstendenzen im NEO-FFI: Eine experimentelle Überprüfung [Fake-tendencies in the NEO-FFI: an experimental investigation]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 24, 105-117.
- Krampen, G. (1991). Fragebogen zu Kompetenz- und Kontrollüberzeugungen (FKK) [Questionnaire of competence and control persuasion]. Göttingen (Germany): Hogrefe.
- Kubinger, K.D. (1996). Zur Leichtgläubigkeit der Psychologen: Die unselige Anwendung von Persönlichkeitsfragebögen [On the credulity of psychologists: The unfortunate use of personality questionnaires]. In M. Jirasko, J. Glück & B. Rollett (Eds.) (1996). *Perspektiven psychologischer Forschung in Österreich* [Perspectives of psychological research in Austria] (pp. 87-91). Vienna: WUV.
- Kubinger, K.D. (2002). On faking personality inventories. *Psychologische Beiträge* [later: *Psychology Science*], 44, 10-16.
- Kubinger, K.D. (2003a). (Un-) Verfälschbarkeit. In K.D. Kubinger, & R.S. Jäger (Eds.), *Schlüsselbegriffe der Psychologischen Diagnostik* [Keynotes to Psychological Assessment] (pp. 429-432). Weinheim (Germany): Beltz.
- Kubinger, K.D. (2003b). Objektivier Persönlichkeitstest. In K.D. Kubinger, & R.S. Jäger (Eds.), *Schlüsselbegriffe der Psychologischen Diagnostik* [Keynotes to Psychological Assessment] (pp. 304-309). Weinheim (Germany): Beltz.
- Kuntz, D. (1974). Effects of faking instructions on the word-association test. *Psychological Reports*, 35, 1183-1192.
- Lammers, F., & Frankenfeld, V. (1999). Effekte gezielter Antwortstrategien bei einem Persönlichkeitsfragebogen mit „forced-choice“-Format [Effects of aimed response strategies at a personality questionnaire with forced-choice-format.] *Diagnostica*, 45, 65-68.
- Marcus, B. (2003). Das Wunder sozialer Erwünschtheit in der Personalauswahl. [The miracle of social desirability in personnel selection settings] *Zeitschrift für Personalpsychologie*, 2(3), 129-132.
- McCrae, R.R., & Costa, P.T.Jr. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology*, 51, 882-888.
- McFarland, L.A. (2003). Warning Against Faking on a Personality Test: Effects on Applicant Reactions and Personality Test Scores. *International Journal of Selection and Assessment*, 11(4), 265-276.

- Moorman, R.H., & Podsakoff, P.M. (1992). A meta-analytic review and empirical test of the potential confounding effects of social desirability response sets in organizational behaviour research. *Journal of Occupational and Organizational Psychology*, 65, 131-149.
- Mummendey, H.D. (1999). *Die Fragebogen-Methode [The questionnaire method]*. (3rd. Ed.). Göttingen: Hogrefe.
- Neubauer, A.C., & Malle, B.F. (1997). Questionnaire Response Latencies: Implications for Personality Assessment and Self-Schema Theory. *European Journal of Psychological Assessment*, 13, 109-117.
- Nias, D.K.B. (1972). The effects of providing a warning about the lie scale in a personality inventory. *British Journal of Educational Psychology*, 42, 308-312.
- 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, 660-679.
- Ones, D.S., Viswesvaran, C. & Schmidt, F.L. (1993). Comprehensive meta-analysis of integrity test validities: findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology*, 78, 679-703.
- Paulhus, D.L. (1991). Measurement and control of response bias. In J.P. Robinson, P.R. Shaver, & L.S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). San Diego, CA: Academic Press.
- Piedmont, R.L., McCrae, R.R., Riemann, R., & Angleitner, A. (2000). On the invalidity of validity scales: Evidence from self-reports and observer ratings in volunteer samples. *Journal of Personality and Social Psychology*, 78, 582-593.
- Ponocny, I., & Klauer, K.C. (2002). Towards identification of unscalable personality questionnaire respondents: The use of person fit indices. *Psychologische Beiträge [latter: Psychology Science]*, 44, 94-107.
- Rasch, D. & Kubinger, K.D. (2006). *Statistik für das Psychologiestudium – Mit Softwareunterstützung zur Planung und Auswertung von Untersuchungen sowie zu sequentiellen Verfahren [Statistics for the study of psychology – software application of planning and sequential procedures]*. München: Spektrum.
- Robie, C., Curtin, P.J., Foster, T.C., Phillips, H.L. IV, Zbylut, M., & Tetrick, L.E. (2000). The effect of coaching on utility of response latencies in detecting fakers on a personality measure. *Canadian Journal of Behavioral Science*, 32(4), 226-233.
- Rosse, J.G., Miller, J.L., & Stecher, M.D. (1994). A field study of job applicants' reactions to personality and cognitive ability testing. *Journal of Applied Psychology*, 79, 987-992.
- Rotter, J.B. (1982). *The development and application of social learning theory*. New York, NY: Praeger.
- Rotter, J.B., Chance, J.E., & Phares, E.J. (Eds.) (1972). *Applications of a social learning theory of personality*. New York, NY: Holt, Rinehart & Winston.
- Rotter, J.B., Seeman, M., & Liverant, S. (1962). Internal versus external control of reinforcement: A major variable in behaviour theory. In N.F. Washburn (Ed.), *Decisions, values, and groups* (Vol. 2, pp. 473-516). London: Pergamon.
- Ryan, A.M., Sacco, J.M., McFarland, L.A. & Kriska, S.D. (2000). Applicant self-selection: Correlates of withdrawal from a multiple hurdle process. *Journal of Applied Psychology*, 85, 163-179.
- Sanders, G.S., & Baron, R.S. (1975). The motivating effects of distraction on task performance. *Journal of Personality and Social Psychology*, 32, 956-963.
- Schmidt, F.L., & Hunter, J.E. (1998). The validity and utility of personnel selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.

- Schmit, M.J., & Ryan, A.M. (1992). Test-taking dispositions: A missing link? *Journal of Applied Psychology*, *77*, 629-637.
- Schneider, J.F., & Hübner, R. (1980). Einfluss von Verfälschungsinstruktionen auf die Bearbeitungszeit von Persönlichkeitsfragebögen. [The influence of instruction to fake on time needed to respond to personality questionnaires] *Zeitschrift für Experimentelle und Angewandte Psychologie*, *27*, 565-597.
- Scheider-Düker, M., & Schneider, J.F. (1977). Untersuchungen zu Beantwortungsprozess bei psychodiagnostischen Fragebögen. [Analyses of the process of responding to personality questionnaires] *Zeitschrift für Experimentelle und Angewandte Psychologie*, *24*, 282-302.
- Seiwald, B.B. (2002). Replicability and generalizability of Kubinger's results: Some more studies on faking personality inventories. *Psychologische Beiträge* [latter: *Psychology Science*], *44*, 17-23.
- Seiwald, B.B. (2003). Antworttendenzen (response set). In K.D. Kubinger, & R.S. Jäger (Eds.), *Schlüsselbegriffe der Psychologischen Diagnostik* [Keynotes to Psychological Assessment] (pp. 29-32). Weinheim (Germany): Beltz.
- Smither, J.W., Reilly, R.R., Millsap, R.E., Pearlman, K., & Stoffey, R.W. (1993). Applicant reactions to selection procedures. *Personnel Psychology*, *46*, 49-76.
- Steiner, D.D., & Gilliland, S.W. (1996). Fairness reactions to personnel selection techniques in France and the United States. *Journal of Applied Psychology*, *81*, 134-141.
- Sutherland, B.V., & Spilka, B. (1964). Social desirability, item-response time and item significance. *Journal of Consulting Psychology*, *28*, 447-451.
- Usdansky, G., & Chapman L.J. (1960). Schizophrenic-like responses in normal subjects under time pressure. *Journal of Abnormal and Social Psychology*, *60*, 143-146.
- Wagner-Menghin, M.M. (2002). Towards the identification of non-scalable personality questionnaire respondents: Taking response time into account. *Psychologische Beiträge* [latter: *Psychology Science*], *44*, 62-77.
- Wheeler, J.K., Hamill, L.S., & Tippins, N.T. (1996). Warning against candidate misrepresentation: Do they work? Paper presented at the 11th annual conference of the Society of Industrial and Organizational Psychology, San Diego.

Appendix see next page.

Appendix

Intercorrelations (Pearson) between all scales (from the current data; n = 208)

	judging	perceiving	feeling	thinking	intuition	sensing	extro- version	intro- version	self-concept of own competences	internality	powerful others control	chance control
judging	1											
perceiving	-.951	1										
feeling	-.178	.143	1									
thinking	.235	-.202	-.943	1								
intuition	-.490	.483	.342	-.339	1							
sensing	.480	-.470	.342	-.339	.307	1						
extroversion	-.085	.083	-.014	.006	.097	-.950	1					
introversion	.085	-.085	.012	-.006	-.097	.118	-.983	1				
self-concept of own competences	-.085	.107	.225	-.227	.061	-.028	-.399	.424	1			
internality	-.234	.264	.143	-.194	.124	-.155	-.089	.107	.300	1		
powerful others control	-.090	.073	-.099	.120	.082	-.131	.341	-.341	-.442	-.088	1	
chance control	.172	-.189	-.206	.221	-.102	.048	.153	-.166	-.410	-.192	-.394	1