

Running head: FAKING & EXTREME RESPONSES

**Applicants' Strategic Use of Extreme or Midpoint Responses
When Faking Personality Tests^{1,2}**

Cornelius J. König, Manuela Mura, and Johanna Schmidt

Universität des Saarlandes

Current status: In press in *Psychological Reports*

¹ Address correspondence to Cornelius J. König, Fachrichtung Psychologie, Universität des Saarlandes, Campus A1 3, D-66123 Saarbrücken, Germany or ckoenig@mx.uni-saarland.de

² The order of the second and third author was alphabetically determined.

Summary. --- Faking, the intentional distortion of answers to personality tests, is likely a complex process. In particular, participants in previous research have mentioned that they used different kind of strategies to appear more hireable, including systematically more extreme or more midpoint responses. However, quantitative evidence is still lacking. Thus, we conducted an experiment in which 327 students (173 women, 153 men; mean age = 22.1 years, $SD = 2.8$) were randomly assigned to two groups. Hypothetical job advertisements primed participants into believing that the hiring company preferred a person with either a “strong” (Strong Character group) or a “well-balanced” character (Well-balanced Character group). Participants filled out 40 items that were chosen from four established questionnaires because of being neither socially desirable nor undesirable. The responses to these items were used to calculate two extreme response measures and one midpoint response measure. T-tests revealed that the Strong Character group used extreme scores more often than the Well-balanced Character group (and the midpoint scores less often), independently of mean differences. This suggests that fakers use more sophisticated strategies than is often assumed.

Keywords: faking; extreme response style; midpoint response style

Applicants faking responses on questionnaires is a common concern for human resource professionals, posing a threat to the validity of the answers (e.g., Robie, Tuzinski, & Bly, 2006). This worry is substantiated by considerable evidence suggesting that applicants often do fake responses (e.g., Boss, König, & Melchers, 2015; Donovan, Dwight, & Schneider, 2014; Griffith, Chmielowski, & Yoshita, 2007).

Research on faking answers suggests that faking could be a complex strategic process for applicants. A study by König, Merz, and Trauffer (2012) showed that at least some applicants use “sophisticated faking strategies” (p. 450). In particular, some respondents mentioned that they preferred to endorse extreme response categories and avoided the middle category of a Likert-scaled personality test (extreme responding, see, e.g., Borgatta & Glass, 1961, and Van Vaerenbergh & Thomas, 2013), whereas others reported the opposite: preferring the middle category and avoiding extreme response categories (midpoint responding, see, e.g., Van Vaerenbergh & Thomas, 2013). They explained their strategic use of these response styles by voicing corresponding opinions regarding the appropriateness of extreme versus midpoint responding: Extreme responders considered midpoint responding to be evaluated badly by hiring organizations because it could be a sign of mediocrity, whereas midpoint responders considered extreme responding as negatively evaluated because it could indicate a polarizing and obstinate person.

What makes such strategic use of extreme or midpoint responses particularly interesting is that it can result in faked responses without elevating means, whereas mean changes are used as an indicator of faking in many studies (e.g., Birkeland, Manson, Kisamore, Brannick, & Smith, 2006; Burns, Fillipowski, Morris, & Shoda, 2015; Viswesvaran & Ones, 1999). If applicants change items towards the midpoint or the extremes independently of agreement or

nonagreement, their response style might average out over several items, so that using mean changes as the indicator for faking would not detect this phenomenon.

However, König et al.'s (2012) findings must be interpreted with caution as they were from only a small qualitative interview study in which respondents talked about their use of response strategies. Although qualitative studies have strengths, e.g. allowing in-depth exploration of particular issues and generating new hypotheses (Lee, Mitchell, & Sablinski, 1999), it is unclear whether the results can be generalized. Therefore, the current study tests experimentally whether applicants indeed use extreme or midpoint responding in a strategic way. To do so, we constructed one hypothetical job advertisement in a way that indicated that the hiring company preferred a person with a “strong character” (a weak cue that could be construed as implying a preference for more extreme responses) and second job advertisement where the company preferred a person with a “well-balanced character” (a cue that could be construed as implying a preference for more midpoint responses). We hypothesized that this manipulation should lead to more extreme responding and less midpoint responding without affecting means.

Method

Sample

Participants were 327 students attending several business lectures at a German university and who participated voluntarily (173 women, 153 men; 1 person did not indicate her/his sex). The mean age was 22.1 years ($SD = 2.8$). Nearly half of the students (48.3%) worked part-time and 31.8% worked full-time (the rest had no work experience or did not answer this question).

Procedure and manipulation

Participants were asked to imagine that they were looking for a trainee position and found a job advertisement by a hypothetical company. The advertisement was presented to them and

included general information about the company, a description of an attractive trainee program, and a list of the expected qualifications and skills. The latter part included the manipulation: In the Strong Character group ($n = 162$), it was mentioned that applicants should have a “starke Persönlichkeit mit ausgeprägten Charaktereigenschaften” (German original, which can be translated as a “strong character with distinctive traits”), whereas the Well-balanced Character group ($n = 155$) read about a “harmonische Persönlichkeit mit ausgeglichenen Charaktereigenschaften” (German original, which can be translated as “a harmonious character with well-balanced traits”). Participants were then asked to fill out the personality test (the items described below) as applicants.

Response style measures

Standard measures for extreme responses (Borgatta & Glass, 1961; Hamilton, 1968) are to count how often a participant endorsed the endpoints of scales (here called the Extreme Response 1 measure) or the two lowest and the two highest response categories, respectively (here the Extreme Response 2 measure). A standard measure for midpoint responding (Stening & Everett, 1984) is to count how often a participant endorses the midpoint response category (here called the Midpoint Response measure). Although these measures can be used with all items having rating scales, they can be confounded with social desirability. If items are socially desirable, participants may endorse positive extreme response categories without being motivated to endorse extreme response categories in general (and similarly, participants may endorse the negative extreme response category for undesirable items). To find suitable items, the three authors went through several personality questionnaires, choosing 40 items using the criterion that they were neither socially desirable nor undesirable. Ten were chosen from the Thrill and Adventure Seeking subscale of the Sensation Seeking Scale (Form V, German

version, Beauducel, Strobel, & Brocke, 2003) and then altered: In its original form, this scale employs a forced-choice format in which participants must choose between two statements, but we only used the first statements (e.g., “I often wish I could be a mountain climber”). Eight additional items were taken from the NEO-PI-R subscale Excitement-Seeking (German version, Ostendorf & Angleitner, 2004), with a sample item being “I like to be where the action is.” Twelve additional items were taken from a German interest inventory (“Differentieller Interessen-Test”, Todt, 1967), with sample items being “I would like to train a sports team for a competition” and “I like to draw cartoons.” A further ten items came from the Inventory of Polychronicity Values in its German translation (König, Bühner, & Mürling, 2005), with polychronicity being the attitude towards multitasking. A sample item is “I like to juggle several activities at the same time.” All items had to be answered on a five-point rating scale (ranging from 1 = *strong disagreement* to 5 = *strong agreement*). To test whether the two groups differed in their means of all personality items, we conducted a MANOVA across all items; it was not significant, Wilk’s $\Lambda = .86$, $F(40,276) = 1.15$, $p = .26$. This shows that the manipulation did not result in a general score increase that could have confounded the effect of the manipulation on extreme and midpoint responding, thus supporting the selection of items.

All three response style measures were internally consistent: $\alpha_{\text{Extreme Response 1}} = .87$, $\alpha_{\text{Extreme Response 2}} = .84$, and $\alpha_{\text{Midpoint Response}} = .71$. Given their high intercorrelations ($r_{\text{Extreme Response 1, Extreme Response 2}} = .91$, $r_{\text{Extreme Response 1, Midpoint Response}} = -.51$, $r_{\text{Extreme Response 2, Midpoint Response}} = -.82$), the measures can be considered as different operationalizations of the same phenomenon.

Manipulation Check. We asked participants at the very end whether the advertisement contained the expressions “distinctive traits”, “strong character”, “well-balanced traits”, and “harmonious character” (plus “good English skills” and “family-run business” as filler items).

The Strong Character group crossed “yes” to the first two expressions significantly more than the Well-balanced Character group, whereas the Well-balanced Character group crossed “yes” to the second two expressions significantly more than the Strong Character group, all four $\chi^2 > 72$, all $ps < .001$.

Analyses

To test our hypothesis, we compared the means of the two groups by using t-tests (one for each of the three operationalizations).

Results

The Strong Character group used extreme scores more often than the Well-balanced Character group (see Table 1). This effect was significant for the Extreme Response 1 measure, $t(315) = 2.15$, $p_{\text{one-sided}} = .016$, Cohen’s $d = 0.24$, and the Extreme Response 2 measure, $t(315) = 2.31$, $p_{\text{one-sided}} = .011$, Cohen’s $d = 0.26$. Furthermore, the Strong Character group used the midpoint response category significantly less often than the Well-balanced Character group (see Table 1), $t(315) = 1.82$, $p_{\text{one-sided}} = .035$, Cohen’s $d = -0.21$. These results support our hypothesis.

Discussion

Kuncel, Goldberg, and Kiger’s (2011, p. 373) called “for focusing our efforts more deeply on the processes by which people make responses to personality test items.” This study tests a specific hypothesis about such responses: whether applicants strategically use extreme or midpoint categories of rating scale questionnaires, regardless of the items’ content. More precisely, if the job advertisement hinted that the organization was searching for somebody with “a strong character” instead of somebody with “a well-balanced character”, we found more extreme responding and less midpoint responding in the data, using three different operationalizations for these response strategies.

The results thus indicate that applicants use particular strategies when they fill out a personality test under circumstances that imply different personal characteristics are needed, and that the framing of the advertisement affects the strategy used. Psychologists typically look at the means of scales to assess faking responses (cf. Burns & Christiansen, 2011), but applicants' cognitions can be much more complex, including a restrictive use of the midpoint response category or the extreme response categories. This was also mentioned by some of the participants in the qualitative interview study by König et al. (2012) - a finding that has now been quantitatively supported by this experimental study for the first time.

This study also contributes to the literature on extreme and midpoint responding. So far, research has mainly treated extreme and midpoint responding as a bias and searched for correlates on the individual, cultural, or measurement level (e.g., Naemi, Beal, & Payne, 2009; for a review see Van Vaerenbergh & Thomas, 2013). Van Vaerenbergh and Thomas called for more research on stimulus-related antecedents of response styles. The present study suggests another aspect: that people might strategically use either extreme or midpoint responding if they believe that this is advantageous for them.

One limitation is that the present study used a hypothetical setting. However, given the high stakes involved when applicants apply for real jobs, we considered an experimental field study to be unfeasible. Hooper and Sackett (2008) have meta-analytically shown that faking effects in the field and in the lab are similar. Furthermore, the effect sizes were modest in this study. However, it should be noted that the employed manipulation was minimal because only two words were different in the advertisement, and the effects obtained should be seen in the light of this minimal manipulation (cf. Prentice & Miller, 1992). In fact, it might be considered surprising that already such a small manipulation works. At the same time, real-world job

advertisement may contain similarly weak cues for response styles, which implies that effects in the field might also be small.

Future research should explore whether the results generalize to more common personality tests (although they likely include more socially desirable items). In addition, researchers should test the effects of triggering response styles on the criterion- and construct-validity of questionnaires and test possible interactions with individual differences and item types (e.g., Ferrando & Anguiano-Carrasco, 2013). Furthermore, research should test whether applicants' extreme or midpoint response styles are related to other response styles such as acquiescence (e.g., Jackson & Messick, 1961) or random responding (as captured, e.g., by the Minnesota Multiphasic Personality Inventory's Variable Response Inconsistency scale, Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989; see also Handel, Ben-Porath, Tellegen, & Archer, 2010, and Nichols, Greene, & Schmolck, 1989).

For practitioners, this study can be seen as yet another warning not to ignore faking issues. Even if they can compare applicants' personality scores to either scores reported in a test manual or scores obtained from job incumbents and do not observe mean changes, this does not preclude that applicants might still have faked: They might have tried to use midpoint or extreme responses to appear more hireable. Furthermore, practitioners should think carefully about their job advertisements because including or excluding certain attributes can trigger response styles.

References

- Beauducel, A., Strobel, A., & Brocke, B. (2003) Psychometrische Eigenschaften und Normen einer deutschsprachigen Fassung der Sensation Seeking-Skalen, Form V [Psychometric properties and norms of a German version of the Sensation Seeking Scales, Form V]. *Diagnostica*, 49, 61-72.
- Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006) A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, 14, 317-335.
- Borgatta, E. F., & Glass, D. C. (1961) Personality concomitants of extreme response set (ERS). *Journal of Social Psychology*, 55, 213-221.
- Boss, P., König, C. J., & Melchers, K. G. (2015) Faking good and faking bad among military conscripts. *Human Performance*, 28, 26-39.
- Burns, G. N., & Christiansen, N. D. (2011) Methods of measuring faking behavior. *Human Performance*, 24, 358-372.
- Burns, G. N., Phillipowski, J. N., Morris, M. B., & Shoda, E. A. (2015) Impact of electronic warnings on online personality scores and test-taker reactions in an applicant simulation. *Computers in Human Behavior*, 48, 163-172.
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *MMPI-2: Manual for administration and scoring*. Minneapolis: University of Minnesota Press.
- Donovan, J. J., Dwight, S. A., & Schneider, D. (2014) The impact of applicant faking on selection measures, hiring decisions, and employee performance. *Journal of Business and Psychology*, 29, 479-493.

- Ferrando, P. J., & Anguiano-Carrasco, C. (2013) A structural model-based optimal person-fit procedure for identifying faking. *Educational and Psychological Measurement*, 73, 173-190.
- Griffith, R. L., Chmielowski, T., & Yoshita, Y. (2007) Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review*, 36, 341-355.
- Hamilton, D. L. (1968) Personality attributes associated with extreme response style. *Psychological Bulletin*, 69, 192-203.
- Handel, R. W., Ben-Porath, Y. S., Tellegen, A., & Archer, R. P. (2010) Psychometric functioning of the MMPI-2-RF VRIN-r and TRIN-r scales with varying degrees of randomness, acquiescence, and counter-acquiescence. *Psychological Assessment*, 22, 87-95.
- Hooper, A. C., & Sackett, P. R. (2008, April). *Self-presentation on personality measures in lab and field settings: A meta-analysis*. Paper presented at the 23rd annual conference of the Society for Industrial and Organizational Psychology, San Francisco, CA.
- Jackson, D. N., & Messick, S. (1961) Acquiescence and desirability as response determinants on the MMPI. *Educational and Psychological Measurement*, 21, 771-790.
- König, C. J., Bühner, M., & Mürling, G. (2005) Working memory, fluid intelligence, and attention are predictors of multitasking performance, but polychronicity and extraversion are not. *Human Performance*, 18, 243-266.
- König, C. J., Merz, A.-S., & Trauffer, N. (2012) What is in applicants' minds when they fill out a personality test? Insights from a qualitative study. *International Journal of Selection and Assessment*, 20, 442-452.
- Kuncel, N. R., Goldberg, L. R., & Kiger, T. (2011) A plea for process in personality prevarication. *Human Performance*, 24, 373-378.

- Lee, T. W., Mitchell, T. R., & Sablinski, C. J. (1999) Qualitative research in organizational and vocational psychology, 1979-1999. *Journal of Vocational Behavior*, 55, 161-187.
- Naemi, B. D., Beal, D. J., & Payne, S. C. (2009) Personality predictors of extreme response style. *Journal of Personality*, 77, 261-286.
- Nichols, D. S., Greene, R. L., & Schmolck, P. (1989) Criteria for assessing inconsistent patterns of item endorsement on the MMPI: Rationale, development, and empirical trials. *Journal of Clinical Psychology*, 45, 239-250.
- Ostendorf, F., & Angleitner, A. (2004). *NEO-Persönlichkeitsinventar nach Costa und McCrae: Revidierte Fassung [The NEO personality inventory after Costa and McCrae: Revised version]*. Göttingen, Germany: Hogrefe.
- Prentice, D. A., & Miller, D. T. (1992) When small effects are impressive. *Psychological Bulletin*, 112, 160-164.
- Robie, C., Tuzinski, K. A., & Bly, P. R. (2006) A survey of assessor beliefs and practices related to faking. *Journal of Managerial Psychology*, 21, 669-681.
- Stening, B. W., & Everett, J. E. (1984) Response styles in a cross-cultural managerial study. *Journal of Social Psychology*, 122, 151-156.
- Todt, E. (1967). *Differentieller Interessentest (DIT) [Differential interest test (DIT)]*. Bern, Switzerland: Huber.
- Van Vaerenbergh, Y., & Thomas, T. D. (2013) Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, 25, 195-217.
- Viswesvaran, C., & Ones, D. S. (1999) Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement*, 59, 197-210.

Table 1

Mean differences between groups

	Strong Character group	Well-balanced Character group
	<i>M (SD)</i>	<i>M (SD)</i>
Extreme Response 1	10.75 (6.90)	9.15 (6.28)
Extreme Response 2	40.75 (10.40)	38.15 (9.61)
Midpoint Response	9.99 (4.96)	11.00 (4.88)

Note. Extreme Response 1 = how often participants endorsed the endpoints of the scales; Extreme Response 2 = how often participants endorsed the two lowest and two highest response categories; Midpoint Response = how often participants endorsed the midpoint response category.