

Are observer ratings of applicants' personality also faked? Yes, but less than self-reports*

Cornelius J. König¹ | Larissa A. Steiner Thommen² | Anne-Marie Wittwer² |
Martin Kleinmann²

¹Fachrichtung Psychologie, Universität des Saarlandes, Saarbrücken, Germany

²Psychologisches Institut, Universität Zürich, Zürich, Switzerland

Correspondence

Cornelius J. König, Universität des Saarlandes, AE Arbeits- und Organisationspsychologie, Campus A1 3, D-66123 Saarbrücken, Germany.
Email: ckoenig@mx.uni-saarland.de

Abstract

Although self-report personality tests are a comparatively cheap and easy-to-administer personnel selection tool, researchers have criticized them for not predicting enough criterion-related variance. Researchers have suggested using observer-ratings of personality (e.g., as part of a reference check from a supervisor) because observer-ratings have been reported to be more predictive. However, it is theoretically and empirically unclear whether supervisors also engage in faking (the intentional distortion of responses). Study 1 explored faking among managers who were first asked to imagine that a subordinate had to leave his/her job for private reasons and then to rate the personality of the subordinate. A week later, managers rated their subordinates honestly. A repeated-measures MANOVA indicated that managers did fake. Study 2 (among supervisors of working students) replicated the above findings but also showed that there is less faking in supervisor-ratings than in self-ratings. Furthermore, we found no evidence that the validity of personality scales for predicting academic performance depends on self- versus observer-ratings or on an applicant versus an honest condition. These two studies thus show that practitioners should not equate personality ratings obtained from observers in a selection context with honest personality ratings.

1 | INTRODUCTION

Personality tests are an attractive personnel selection tool for organizations because they are comparatively cheap, easy to administer, and have at least a modest amount of criterion validity (e.g., Ones, Dilchert, Viswesvaran, & Judge, 2007). Furthermore, practitioners who believe that performance is predicted by can-do factors such as general mental ability as well as will-do factors such as conscientiousness (cf. Rynes, Colbert, & Brown, 2002) likely consider personality tests a convenient tool to measure such will-do factors. Thus, it does not come as a surprise that surveys in several countries show that a considerable number of organizations use personality tests for personnel selection (e.g., Diekmann & König, 2015; Di Milia, 2004; Fakir & Laher, 2015; Furnham, 2008).

Conversely, personality tests have been criticized for not predicting enough criterion-related variance (e.g., Campion and Schmitt in Morgeson et al., 2007). In comparison with many other selection tools (in particular with tests of general mental ability), personality tests explain only a small percentage of validity. One way to increase the predictive validity of personality tests is to rely on observer-ratings in addition to self-ratings. In fact, meta-analytic research has shown that observer-ratings of personality have higher validities than self-ratings of personality (Connelly & Ones, 2010; Poropat, 2014; Oh, Wang, & Mount, 2011). There are several reasons for the advantage of observer-ratings over self-ratings. In particular, observer-ratings of personality are more reliable than self-ratings (i.e., higher Cronbach's alphas, Balsis, Cooper, & Oltmanns, 2015), and reliability is a necessary condition for validity (Nunnally & Bernstein, 1994). Furthermore, there is evidence suggesting that self-ratings are more error-prone than observer-ratings, for example because self-ratings seem to suffer more from selective attention effects and from overly positive self-views (Connelly & Hülsheger, 2012). The reduced error-proneness associated with others ratings also fits well to the finding that there is considerable agreement between observer-ratings (Connelly & Ones, 2010),

*The order of the second and the third author was alphabetically determined. Cornelius J. König worked previously at the Universität Zürich in Switzerland. Larissa A. Steiner Thommen works now for Talent & Leadership Solutions AG, Zürich, Switzerland, and Anne-Marie Wittwer works now for Swisscom, Zürich, Switzerland.

indicating that observer-ratings are fairly systematic. Other might even have a more complete picture than people themselves (see Dietl, Meurs, & Blickle, 2017).

Due to the advantages of observer-ratings of personality, researchers have suggested to use observer-ratings in selection contexts (Connelly & Hülshager, 2012; Oh et al., 2011; see also Murphy and Campion in Morgeson et al., 2007). In particular, Oh et al. proposed that personality judgments could be obtained from previous supervisors as they are often asked to provide a reference for job applicants. As such, a reference check could be extended to include personality items (see also Zimmerman, Triana, & Barrick, 2010).

Although including personality tests in reference check could be considered a feasible way to take advantage of higher validities of observer-ratings, researchers have wondered whether supervisor-rated personality measures can be faked (Connelly & Hülshager, 2012). We formally define faking as intentional distorting of personality test scores. This definition is a slightly simplified version of typical faking definitions. For example, Fell and König (2016) define faking as "applicants' conscious distortion of responses in order to achieve better scores (e.g., on a personality test) and to increase their chances of being hired" (p. 672; for similar definitions see, e.g., McFarland & Ryan, 2000; Roulin & Krings, 2016). Such a simplification is necessary because typical faking definitions only capture faking in self-reports and not faking in other-reports. Thus, our definition focuses on whether or not personality scores are distorted rather than whose personality scores are distorted. Importantly, our definition is consistent with many authors that faking requires intent (see e.g., also Ziegler, MacCann, & Roberts, 2011; Griffith & McDaniel, 2006).

Applicants' faking of personality tests seems to be fairly common (e.g., Birkeland, Manson, Kisamore, Brannick, & Smith, 2006; Salgado, 2016) and thus worries many practitioners (e.g., Robie, Tuzinski, & Bly, 2006) because they and several researchers fear that it affects the predictive validity of personality tests (e.g., Donovan, Dwight, & Schneider, 2014; Levin & Zickar, 2002). In particular, if some applicants fake more than others do, it can change who gets a job offer (or at least who jumps of the personality test hurdle in a selection process). However, other researchers do not consider applicants' faking as troublesome (e.g., Hogan, Barrett, & Hogan, 2007). In fact, there is an ongoing debate on the effect of applicants' faking or the noneffect on criterion-related validity (e.g., Morgeson et al., 2007; Tett & Christiansen, 2007).

If those who fear the effect of applicants' faking on validity are correct, then they should also fear that supervisor fake if they rate the personality of their subordinates as part of a reference check. This fear has likely also stimulated the call for research on the fakability of observer-ratings of personality in selection context (Connelly & Hülshager, 2012). We answer this call by examining the fakability of observer-ratings for the first time.

As previous faking definitions focused on applicants' faking, so do theoretical models of faking as none of the faking models so far (e.g., Marcus, 2009; McFarland & Ryan, 2000; Roulin, Krings, & Binggeli, 2016) have included other-rated personality measures. Nevertheless, basic tenets of these models are that both a motivational and an ability

component is needed to explain faking (e.g., Goffin & Boyd, 2009; Marcus, 2009; McFarland & Ryan, 2000; Snell, Sydell, & Lueke, 1999), and these basic tenets can be used to generate arguments as to why supervisors might fake or respond honestly when rating the personality of their subordinates.

Regarding supervisors' motivation to fake, it is possible to develop arguments that supervisors are motivated and unmotivated to fake if they rate the personality of their subordinates in a selection context. On the one hand, it is the applicant who gets the job, not the supervisor, and self-interest is the crucial motivational factor behind faking according to several faking models (e.g., Marcus, 2009; McFarland & Ryan, 2000). Current supervisors might even sympathize with future supervisors and feel obliged to act in accordance with their interest, which is to obtain honest ratings of applicants (Bangerter, Roulin, & König, 2012). Furthermore, supervisors might be interested in maintaining their own reputations and thus do not want to give unrealistically high ratings because they might jeopardize their reputations (i.e., supervisors may look bad if applicants are hired without having the qualities suggested by the supervisors' ratings).

On the other hand, supervisors might indeed be motivated to fake because they might feel an emotional connection with their subordinates. They likely feel proud of a former subordinate. Especially if they are climbing the ranks, supervisors might think that their former subordinate deserves a promotion. Thus, supervisors might want to help their subordinates and this can motivate them to fake (cf. Marcus, 2009; McFarland & Ryan, 2000). In addition, supervisors might also be motivated to fake because they might perceive social or even legal pressure to provide a positive reference, believing that it is not appropriate to criticize subordinates and potentially fearing legal troubles (cf. Compton & Albinsson, 2013).

Regarding the ability to fake, it can be argued that supervisors might have a rather high ability to fake. They might be more knowledgeable about personnel selection processes and personality tests because they have worked for longer than their subordinates have and are likely to have been on the other side of the selection process (e.g., as a hiring committee member) as well. Serving on a hiring committee would afford supervisors unique insight as to what is typically expected of applicants (cf. Jansen et al., 2013). Furthermore, their experience with selection situations might make items particularly transparent to them and thus reduce the likelihood of sticking to wrong assumptions about personality tests (König, Merz, & Trauffer, 2012; König, Mura, & Schmidt, 2015). In the language of theoretical faking models (e.g., Marcus, 2009; McFarland & Ryan, 2000), this means that supervisors might have a particularly high ability to fake and thus would find it easy to do so (in case they are motivated to fake).

Given the above competing arguments, we designed two studies to explore faking among observer-ratings of applicant personality. Study 1 is a within-subject experiment in which supervisor rated the personality of subordinates, once in an applicant situation (in which they imagined that one of their subordinates was leaving his/her job for private reasons and that the potentially recruiting organizations asked for an observer-rating of his/her personality) and once honestly.

With Study 2, we aim at replicating Study 1 with supervisors of (working) students as participants and at extending it by additionally testing whether the amount of faking is similar in observer-rating and in self-ratings.

2 | STUDY 1

2.1 | Method

2.1.1 | Sample

Participants were 56 leaders from two Swiss organizations (out of 76 who we invited by email). They were on average 48.3 years old ($SD = 6.90$). The organizations did not permit us to collect any further demographic data to ensure anonymity.

Of the observed subordinates (34 males, 22 females), 5 were between 21 and 30 years old, 17 between 31 and 40 years, 22 between 41 and 50 years, and 11 were older than 51 years (1 missing). Supervisors had known them on average for 6.0 years ($SD = 5.2$). Thirty-two saw their member daily, 17 several times a week, four on a weekly basis, and three saw the subordinates once or twice a month. When asked whether they “know the person to be rated well enough to rate him/her accurately,” participants overwhelmingly agreed ($M = 4.5$, $SD = 0.7$, on a 5-point Likert scale from 1 = *strongly disagree* to 5 = *strongly agree*). Furthermore, 30 participants reported to have had experience with personality tests (24 had not; 2 missing).

2.1.2 | Procedure

For this within-subject experiment, we contacted participants via email and sent them a link to an online questionnaire. To ensure that participants' choice who to rate was not based on liking, participants were asked to rate a member of their team by choosing the subordinate whose family name was either first or last alphabetically.

Participants were asked to imagine that one of their subordinates had to leave his/her job for private reasons, that s/he had found an interesting job, that s/he had mentioned the supervisor's name as a potential reference person, and that the potential employer has now contacted the supervisor to fill out a personality test regarding the personality of the subordinate.

About two weeks later, we contacted participants again and asked them to fill out the personality test a second time but this time, they were asked to respond as honestly as possible. We were able to link the questionnaires because of a code that each participant generated for him-/herself. The order (the application condition before the honest condition) was based on the meta-analytical results (Hooper & Sackett, 2008) that lab studies with such an order produce results that are more consistent with field research than the alternative order.

2.1.3 | Measures

To assess personality, we measured three Conscientiousness facets (Achievement-Striving, Self-Discipline, and Cautiousness) and the Extraversion facet Assertiveness, using scales from the International Personality Item Pool (Goldberg et al., 2006). These scales are widely used as alternate versions to the proprietary NEO PI-R

inventory (Costa & McCrae, 1992) and correlate highly with them (Goldberg, 1999). Extraversion and Conscientiousness were chosen because they are the two traits with the greatest interrater reliabilities and self-other correlations (Connelly & Ones, 2010). A sample item for Achievement-Striving is “Turns plans into actions” (in total nine items); a sample item for Self-Discipline is “Is always prepared” (in total eight items); a sample item for Cautiousness is the reverse-coded item “Rushes into things” (in total eight items); and a sample item for Assertiveness is the reverse-coded item “Waits for others to lead the way” (in total nine items). Items were either written for a male or a female person, depending on the gender of the person to be rated, and they can be found in the Appendix. The online survey presented all personality items in a randomized order and employed a Likert response format ranging from 1 = *strongly disagree* to 5 = *strongly agree*.

3 | RESULTS AND DISCUSSION

As a manipulation check, at the end of the first round of data collection we asked participants whether they found it easy to imagine providing a reference for the person in the described selection situation to which they agreed ($M = 4.00$, $SD = 0.91$, on a 5-point Likert scale from 1 = *strongly disagree* to 5 = *strongly agree*). Furthermore, at the end of the second round of data collection, we asked them whether they had responded to the items outside the context of a personnel selection situations to which, on average, they strongly agreed ($M = 4.55$, $SD = 0.87$, on the same scale).

Table 1 reports the means, correlations, and reliabilities for all variables. A repeated-measures MANOVA over all four personality dimensions revealed a significant effect, Wilks $\Lambda = .77$, $F(4, 52) = 3.88$, $p < .01$, $\eta^2 = .23$. Additional follow-up univariate t -tests (see Table 1) showed significant differences for all four personality dimensions.

Study 1 shows that supervisors do fake personality ratings of their subordinates. When supervisors were asked to imagine that one of their subordinates was leaving his/her job for private reasons and the potentially recruiting organizations was asking for a personality assessment, supervisors then described their subordinates more positively than when they were asked to rate them honestly.

Although Study 1 reports preliminary evidence that observers fake as well, it remains unclear whether the extent of faking in observer-ratings is similar to self-reports. We have already outlined above the competing arguments for and against observer faking, and these arguments led us to hypothesize that self-report ratings are faked more than observer-ratings. Study 2 tests this hypothesis using a new sample of (working) students and their supervisors. In addition, Study 2 should replicate the finding of Study 1 that observer-ratings are also faked.

Furthermore, Study 2 will allow us to analyse whether personality scores in the fake and honest conditions and self-rated and observer-rated personality scores have criterion-related validity. Faking's effect or noneffect on criterion-related validity is a highly debated topic (e.g.,

TABLE 1 Descriptive statistics (Study 1)

Variable	M	SD	1	2	3	4	5	6	7	8	d_{within}	$t_{df = 55}$
"Applicant" condition												
1. Achievement-striving	4.04	0.61	.85									
2. Self-discipline	4.02	0.61	.69**	.84								
3. Cautiousness	4.10	0.53	.35**	.60**	.77							
4. Assertiveness	3.34	0.76	.46**	.22	-.16	.86						
"Honest" condition												
5. Achievement-striving	3.86	0.62	.80**	.64**	.37**	.34*	.86				0.46	3.32 ^{††}
6. Self-discipline	3.90	0.71	.59**	.83**	.47**	.23	.69**	.90			0.30	2.22 [†]
7. Cautiousness	3.96	0.61	.46**	.69**	.77**	-.11	.57**	.66**	.84		0.36	2.72 [†]
8. Assertiveness	3.22	0.71	.53**	.33*	-.04	.81**	.55**	.31*	.01	.85	0.26	1.99 [†]

Note. Cronbach's alphas displayed in the diagonal (in italics). $N = 56$. d_{within} is Cohen's within-subject effect size calculated as follows:

$$d_{\text{within}} = M_{\text{applicant}} - M_{\text{honest}} / \text{SQRT}(SD_{\text{applicant}}^2 + SD_{\text{honest}}^2 - 2r * SD_{\text{applicant}} * SD_{\text{honest}}).$$

* $p < .05$, two-tailed; ** $p < .01$, two-tailed; † $p < .05$, one-tailed; †† $p < .01$, one-tailed.

Morgeson et al., 2007; Tett & Christiansen, 2007). If faking does in fact affect criterion-related validity, does it do so in a positive or in a negative way (for different positions see, e.g., Donovan et al., 2014; Hogan et al., 2007; Levin & Zickar, 2002; Ones, Viswesvaran, & Reiss, 1996)?

4 | STUDY 2

4.1 | Method

4.1.1 | Sample

Participants were 84 pairs of (working) students and their supervisors. Participants were recruited through diverse channels (posters, flyers, emails, and advertisements at universities). As an incentive, we offered job application training for student participants.

Supervisors were on average 39.8 years-old ($SD = 10.5$), with one person not answering this item. Fifty-four participants were male and 30 were female. Forty-five percent were supervising students in service jobs (e.g., waitressing), 17% were supervising students in internships, 24% were supervising students conducting their Master theses, and 14% were supervision students in other jobs. Supervisor worked at universities (27%), in the financial industry (13%), in the building/engineering industry (11%), or in various other industries (ranging from health to retail and catering). When we asked supervisors whether they "know the person to be rated well enough to rate him/her accurately," participants clearly agreed ($M = 4.0$, $SD = 0.9$, on a 5-point Likert scale from 1 = *strongly disagree* to 5 = *strongly agree*). In nearly three quarter of the cases (73%), the supervision relationship had not yet ended. On average, the pairs have/had worked together for two years ($M = 2.0$, $SD = 3.0$). Furthermore, 43 supervisors indicated that they had experience with personality tests (vs. 39 who had not and 2 who did not answer this item).

Students were primarily from the University of Zurich (60%) and the Swiss Federal Institute of Technology in Zurich (18%) and studied various subjects (48% arts/humanities, 18% natural sciences/mathematics, 14% business studies, 12% engineering, 5% law, 4% architect-

ture). Forty-five were female and 39 male. Sixteen of the students met their supervisors daily, 25 several times a week, 23 on a weekly basis, 10 once or twice a month, and 7 less than one per month (three did not report the frequency of their meetings). They were on average 27.0 years old ($SD = 5.0$).

4.1.2 | Procedure

This study employed a 2×2 design, with source (self- vs. supervisor-ratings) being a within-subject factor and with "honest vs. applicant" being a within-subject factor. We generated a code to link all questionnaires, which were all online.

Working students were asked to imagine that they had applied to a job at their dream company and had already jumped through the first phase of the selection process and that the company had asked them to fill out a personality test as the next phase. After filling out the 34 personality items, students responded to a few additional questions (mainly demographics). Roughly one week later, we contacted the working students again and asked them to fill out the personality test again, but this time to respond as honestly as possible.

Supervisors were asked to imagine that their student had mentioned the supervisor's name as a potential reference person and that the potential employer had now contacted her/him to fill out a personality test regarding the student's personality. Afterward, participants answered a few additional questions (mainly demographics). Roughly one week later, we emailed the supervisors again and asked them to fill out the personality test again, but this time to respond as honestly as possible.

4.1.3 | Measures

To measure personality, we used the same scales as in Study 1. To measure academic performance, students reported their Bachelor grade point average (or if, not applicable, their latest grade point average). In Switzerland, the best grade is 6.0 and everything below 4.0 is a fail.

5 | RESULTS AND DISCUSSION

As a manipulation check, we asked supervisors whether they found it easy to imagine providing a reference in the applicant condition to which they agreed, $M = 3.68$, $SD = 1.25$ (supervisors responded on a 5-point Likert scale ranging from 1 = *strongly disagree* to 5 = *strongly agree*). Furthermore, we asked them whether they had filled out the “honest” items outside the context of the personnel selection situation, and participants agreed on average, $M = 3.28$, $SD = 1.66$ (participants responded on the same scale).

Table 2 reports the means, correlations, and reliabilities for all variables. We first repeated the analysis conducted in Study 1 using only the supervisor data. A repeated-measures MANOVA over all four personality dimensions revealed a significant effect, Wilks $\Lambda = .86$, $F(4, 80) = 3.36$, $p < .05$, $\eta^2 = .14$, which replicates the findings from Study 1. Follow-up univariate t -tests (see Table 2), however, showed significant differences only for Cautiousness (and only small nonsignificant effect sizes for the other three personality dimensions).

Second, we analyzed the full data set (i.e., the supervisor- and the self-rated data). A two-way repeated-measures MANOVA over all four personality dimensions revealed a significant effect for the applicant versus honest factor, Wilks $\Lambda = .62$, $F(4, 80) = 12.09$, $p < .01$, partial $\eta^2 = .38$, for the self- versus observer-rating factor, Wilks $\Lambda = .64$, $F(4, 80) = 11.27$, $p < .01$, partial $\eta^2 = .36$, and for the interaction, Wilks $\Lambda = .84$, $F(4, 80) = 3.91$, $p < .01$, partial $\eta^2 = .16$. The significant interactions show that there was more faking in the self- than the supervisor-rated data. The main effect of the applicant versus the honest factor indicates that overall, personality scores were higher in the applicant condition, whereas the main effect of the self-vs. observer-rating factor demonstrates that the observer-ratings were on average higher than self-ratings of personality.

Table 2 reports the correlations between academic performance and the personality facets rated by the working students themselves and by their supervisors. In particular, the last row shows that all correlations with the Conscientiousness facets (i.e., Achievement-Striving, Self-Discipline, and Cautiousness) are around .18 no matter whether the facets were self- or supervisor-rated or if they were obtained in the as applicant or the honest condition. The correlations sometimes reached significance. In contrast, the Extraversion facet Assertiveness did not correlate significantly with academic performance (in neither condition).

In addition, Table 3 shows the results of follow-up univariate two-way repeated-measures ANOVAs for each of the four personality facets. As can be seen there, we found significant interaction effects only for Achievement-Striving and Self-Discipline, and no significant effects for either Cautiousness or Assertiveness.

To summarize, Study 2 shows again that observer fake personality ratings of the people they supervise. In addition, Study 2 also demonstrates that there is less faking in observer-rating than in self-ratings—in fact, sometimes supervisors fake to such a small extent that the effects are only visible if analyzed in a multivariate way. In addition, we found no evidence that the validity of personality

scales is affected by self- or observer-ratings or by an applicant or an honest condition.

6 | GENERAL DISCUSSION

The aim of this research project was to investigate whether supervisors engage in faking when they provide personality ratings for their subordinates. We found that yes, supervisors do fake but less than the subordinates self-report ratings of personality.

This research was conducted as the literature includes competing arguments with regard to supervisors' motivation and ability to fake. The results support the “pro” side: In particular, it seems likely that supervisors are motivated to fake because they feel emotionally connected to their subordinates and want to support them and their career by describing them in a slightly more positive light. In addition, they might perceive social or even legal pressure not to write something negative if providing a reference (cf. Compton & Albinsson, 2013). Furthermore, if supervisors are motivated to fake, they find it comparatively easy (i.e., have a high ability to fake, Marcus, 2009; McFarland & Ryan, 2000) because their experience with personnel selection in general might have them privy to what recruiting managers are likely looking for (cf. Jansen et al., 2013) and how personality tests work (cf. König et al., 2012, 2015). Future research is, however, needed to test the validity of these assumed mediating mechanisms.

Although we found evidence that supervisors did fake in two studies, it should be kept in mind that Study 2 showed that subordinates faked more than supervisors. If faking models are correct in assuming that faking is caused by a motivational and an ability component plus situational factors (e.g., Marcus, 2009; McFarland & Ryan, 2000; Roulin et al., 2016), the difference between supervisors and subordinates could be due to subordinates' greater motivation to fake, because supervisors might have a higher ability to fake than subordinates (as argued above) and because situational factors were kept controlled (e.g., no use of warnings).

In our data, personality scores obtained in the applicant condition correlated equally with academic performance in comparison with personality scores obtained in the honest condition. Although this is consistent with arguments that faking has no effect on criterion-related validity (e.g., Ones et al., 1996), it is inconsistent with other studies and practitioners' fear of faking. In particular, Peterson, Griffith, Isaacson, O'Connell, and Mangos (2011) found that faking correlated positively with counterproductive work behavior (see also Donovan et al., 2014), which speaks to practitioners' concern over the effect of faking on validity (e.g., Robie et al., 2006). In contrast, Blickle, Momm, Schneider, Gansen, and Kramer (2009) reported evidence that distorted personality scores predict performance more than honest personality scores (for a similar result see Ingold, Kleinmann, König, & Melchers, 2015). Clearly, more empirical data are needed to understand these inconsistencies.

Moreover, the evidence for the relationship between observer-rated personality scores obtained in a selection context and academic performance suggest that asking supervisors for personality ratings of

TABLE 2 Descriptive statistics (Study 2)

Variable	M	SD	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	d_{within}	$t_{df=83}$
"Applicant" condition																					
1. Achievement-striving (self)	4.11	0.47	0.74																		
2. Achievement-striving (obs.)	4.13	0.67	0.16	.88																	
3. Self-discipline (self)	4.02	0.58	0.65**	.16	.84																
4. Self-discipline (obs.)	4.17	0.65	0.10	.82**	.15	.81															
5. Cautiousness (self)	4.08	0.43	0.46**	.25*	.43**	.24*	.74														
6. Cautiousness (obs.)	4.19	0.51	0.11	.56**	.10	.60**	.33**	.76													
7. Assertiveness (self)	3.51	0.57	0.21*	.04	.29**	.07	.07	-.01	.84												
8. Assertiveness (obs.)	3.27	0.67	0.17	.40**	.14	.29**	-.03	.12	.38**	.83											
"Honest" condition																					
9. Achievement-striving (self)	3.88	0.43	.55**	.21*	.49**	.06	.28**	.03	.24*	.15	.70										
10. Achievement-striving (obs.)	4.08	0.61	.12	.86**	.09	.67**	.17	.51**	.08	.42**	.15	.85									
11. Self-discipline (self)	3.71	0.55	.42**	.05	.67**	-.03	.13	-.01	.23*	-.01	.58**	-.05	.83								
12. Self-discipline (obs.)	4.10	0.62	.17	.72**	.31**	.72**	.13	.55**	.10	.33**	.22*	.76**	.16	.86							
13. Cautiousness (self)	3.95	0.46	.34**	.21*	.36**	.28**	.64**	.36**	.09	-.14	.48**	.09	.33**	.20	.77						
14. Cautiousness (obs.)	4.08	0.54	.18	.57**	.21*	.62**	.34**	.79**	.02	.21*	.14	.57**	.07	.64**	.45**	.83					
15. Assertiveness (self)	3.47	0.60	.19	.00	.35**	.00	.02	-.13	.87**	.41**	.26*	.02	.26*	.06	-.01	-.12	.86				
16. Assertiveness (obs.)	3.22	0.61	.11	.41**	.13	.39**	.04	.19	.43**	.85**	.08	.43**	-.10	.32**	-.08	.23*	.40**	.84			
17. Academic performance	4.93	0.41	.16	.24*	.19	.27*	.24*	.11	.03	-.05	.08	.20	.09	.22*	.22*	.19	-.04	.01	-		

Note. Cronbach's alphas is displayed in the diagonal (in italics). self, self-rating; obs., observer-rating. $N = 84$ apart from academic performance where $N = 76$. d_{within} is Cohen's within-subject effect size calculated as follows: $d_{within} = M_{applicant} - M_{honest} / \sqrt{SD_{applicant}^2 + SD_{honest}^2 - 2r * SD_{applicant} * SD_{honest}}$.

* $p < .05$, two-tailed; ** $p < .01$, two-tailed; † $p < .01$, one-tailed.

TABLE 3 Results of the univariate two-way repeated ANOVAs (Study 2)

Personality facet	Applicant vs. honest condition			Self- vs. observer-rating condition			Interaction		
	Wilks Λ	$F(1, 83)$	Partial η^2	Wilks Λ	$F(1, 83)$	Partial η^2	Wilks Λ	$F(1, 83)$	Partial η^2
Achievement-striving	.79	21.50**	.21	.97	2.28 ^{n.s.}	.03	.91	8.69**	.09
Self-discipline	.75	26.96**	.25	.87	12.13**	.13	.88	11.79**	.12
Cautiousness	.85	15.20**	.15	.94	5.06*	.06	1.00	0.09 ^{n.s.}	.00
Assertiveness	.95	4.03*	.05	.87	12.50**	.13	1.00	0.03 ^{n.s.}	.00

Note. $N = 84$.

* $p < .05$, two-tailed; ** $p < .01$, two-tailed.

their subordinates might be a worthwhile selection tool worthwhile among practitioners (see also Zimmerman et al., 2010). In particular, supervisor-rated personality scores could be added to reference letters (cf. Kuncel, Kochevar, & Ones, 2014).

A surprising finding in Study 2 was that personality scores rated by supervisors were higher than the ones rated by the working student themselves. This is surprising because supervisors are known to give their subordinates lower performance ratings than they do themselves (Heidemeier & Moser, 2009), and self-ratings are generally considered to suffer more from overly positive perceptions than observer-ratings (Connelly & Hülshager, 2012). One could thus have believed that the supervisors in this study were less lenient when ratings others personality. This might have been due to the fact that data was collected in Switzerland, a country where modesty is highly valued trait (cf. König, Hafsteinsson, Jansen, & Stadelmann, 2011), and although very modest individuals might rate themselves lower than less modest individuals, modesty likely does not affect supervisor ratings. This unexpected finding might also be a reason why we could not replicate the general finding in the literature which shows that observer-ratings have more predictive power than self-reports (Connelly & Ones, 2010; Oh et al., 2011; Poropat, 2014), whereas personality only moderately predicted academic performance in our data, regardless of whether it was supervisor- or self-rated. In the previous studies, the tendency to give overly positive ratings might have distorted self-ratings but not observer-ratings (thus reducing the correlation between performance and self-rated personality), whereas this advantage of observer-ratings was likely missing in our study, reducing the correlation between performance and observer-ratings of personality. Furthermore, the Extraversion facet, Assertiveness, did not correlate with our academic performance measure at all, which might have been due to the fact that Swiss grade point average is mostly based on written exams, in which differences in Assertiveness might not play a major role.

As in all research, the present study is not free of limitations. First, we used hypothetical scenarios to study faking in observer-ratings, and thus we do not know for sure how well our results generalize to the field. Furthermore, using a scenario approach has the additional disadvantage that we could only assess personality traits that can be assumed to be of relevance for jobs in general (vs. specific traits that are known to be relevant for a specific job according to a job analysis;

cf. Judge & Zapata, 2015). However, it should be kept in mind that previous research has suggested that lab and field approaches converge substantially: this has been found in faking research (Hooper & Sackett, 2008) and in research on the relation between lab and field research (Mitchell, 2012; Vanhove & Harms, 2015). Second, we used scenarios in which supervisors had no motivation to fake bad (cf. Boss, König, & Melchers, 2015). Other scenarios, however, are possible in which, for instance, supervisors do not want their subordinate to leave their teams and thus might be tempted to give low personality ratings such that the new employer would no longer be interested in hiring them. (Another scenario is that supervisors try to “kick their subordinates upstairs,” i.e., to get rid of them by promoting them—a situation in which supervisors likely fake good even more.)

Future research could shed light on ways to mitigate faking of observer-ratings of personality. Previous research has already sought ways to reduce faking of self-reported personality scales (e.g., Fan et al., 2012), and two strategies seem to be particularly promising: the use of warnings (e.g., Burns, Fillipowski, Morris, & Shoda, 2015) and the use of a forced-choice instead of normative answering format (e.g., Christiansen, Burns, & Montgomery, 2005; O'Neill et al., in press). These faking-reducing strategies could also be tested with observer-ratings of personality.

Future research could also focus on predictors of faking of other-reports of personality in selection contexts. Researchers have developed several theoretical models that contain predictors of faking of self-reports (e.g., Goffin & Boyd, 2009; Marcus, 2009; McFarland & Ryan, 2000; Roulin et al., 2016), and future research could test whether these models also apply to faking of observer-ratings of personality. For example, supervisors might have more experience with personnel selection situations than applicants and thus find it easier to detect what the hiring organizations is looking for in an applicant (Jansen et al., 2013; König, Melchers, Kleinmann, Richter, & Klehe, 2006). Furthermore, if we know more about the generalizability of these models, it could elucidate models of faking of self-reports.

For practitioners, this research project has important implications because it cautions them that ratings obtained from observers in a selection context should not be considered as honest personality ratings. Although observer-ratings might be more accurate ratings of the applicant's personality than self-reports, they are still prone to distortions.

REFERENCES

- Balsis, S., Cooper, L. D., & Oltmanns, T. F. (2015). Are informant reports of personality more internally consistent than self reports of personality? *Assessment*, 22, 399–404. doi:10.1177/1073191114556100
- Bangerter, A., Roulin, N., & König, C. J. (2012). Personnel selection as a signaling game. *Journal of Applied Psychology*, 97, 719–738. doi:10.1037/a0026078
- 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. doi:10.1111/j.1468-2389.2006.00354.x
- Blickle, G., Momm, T., Schneider, P. B., Gansen, D., & Kramer, J. (2009). Does acquisitive self-presentation in personality self-ratings enhance validity? Evidence from two experimental field studies. *International Journal of Selection and Assessment*, 17, 142–153. doi:10.1111/j.1468-2389.2009.00458.x
- Boss, P., König, C. J., & Melchers, K. G. (2015). Faking good and faking bad among military conscripts. *Human Performance*, 28, 26–39. doi:10.1080/08959285.2014.974758
- Burns, G. N., Fillipowski, 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. doi:10.1016/j.chb.2015.01.051
- Christiansen, N. D., Burns, G. N., & Montgomery, G. E. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Human Performance*, 18, 267–307. doi:10.1207/s15327043hup1803_4
- Compton, N., & Albinsson, P. A. (2013). Reference letters and the uninformed business educator: A U.S. legal perspective. *Administrative Issues Journal: Connecting Education, Practice, and Research*, 3(1), Article 4. doi:10.5929/2013.3.1.5
- Connelly, B. S., & Hülshager, U. R. (2012). A narrower scope or a clearer lens for personality? Examining sources of observers' advantages over self-reports for predicting performance. *Journal of Personality*, 80, 603–631. doi:10.1111/j.1467-6494.2011.00744.x
- Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, 136, 1092–1122. doi:10.1037/a0021212
- Costa, P. T. Jr., & McCrae, R. R. (1992). *Professional manual of the revised NEO personality inventory and NEO five-factor inventory*. Odessa, FL: Psychological Assessment Resources.
- Di Milia, L. (2004). Australian management selection practices: Closing the gap between research findings and practice. *Asia Pacific Journal of Human Resources*, 42, 214–228. doi:10.1177/1038411104045364
- Diekmann, J., & König, C. J. (2015). Personality testing in personnel selection: Love it? Leave it? Understand it! In I. Nikolaou & J. Oostrom (Eds.), *Employee recruitment, selection, and assessment: Contemporary issues for theory and practice* (pp. 117–135). Hove, UK: Psychology Press.
- Dietl, E., Meurs, J. A., & Blickle, G. (2017). Do they know how hard I work? Investigating how implicit/explicit achievement orientation, reputation, and political skill affect occupational status. *European Journal of Work and Organizational Psychology*, 26, 120–132. doi:10.1080/1359432x.2016.1225040
- 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. doi:10.1007/s10869-013-9318-5
- Fakir, S., & Laher, S. (2015). Perceptions of the utility of personality assessment for personnel selection in the South African context: An exploratory study. *Journal of Psychology in Africa*, 25, 482–485. doi:10.1080/14330237.2015.1101285
- Fan, J., Gao, D., Carroll, S. A., Lopez, F. J., Tian, T. S., & Meng, H. (2012). Testing the efficacy of a new procedure for reducing faking on personality tests within selection contexts. *Journal of Applied Psychology*, 97, 866–880. doi:10.1037/a0026655
- Fell, C. B., & König, C. J. (2016). Cross-cultural differences in applicant faking on personality tests: A 43-nation study. *Applied Psychology: An International Review*, 65, 671–717. doi:10.1111/apps.12078
- Furnham, A. (2008). HR professionals' beliefs about, and knowledge of, assessment techniques and psychometric tests. *International Journal of Selection and Assessment*, 16, 300–305. doi:10.1111/j.1468-2389.2008.00436.x
- Goffin, R. D., & Boyd, A. C. (2009). Faking and personality assessment in personnel selection: Advancing models of faking. *Canadian Psychology*, 50, 151–160. doi:10.1037/a0015946
- Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. D. Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 7–28). Tilburg, The Netherlands: Tilburg University Press.
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40, 84–96. doi:10.1016/j.jrp.2005.08.007
- Griffith, R. L., & McDaniel, M. (2006). The nature of deception and applicant faking behavior. In R. L. Griffith & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 1–19). Greenwich, CT: Information Age Publishing.
- Heidemeier, H., & Moser, K. (2009). Self-other agreement in job performance ratings: A meta-analytic test of a process model. *Journal of Applied Psychology*, 94, 353–370. doi:10.1037/0021-9010.94.2.353
- Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology*, 92, 1270–1285. doi:10.1037/0021-9010.92.5.1270
- 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.
- Ingold, P. V., Kleinmann, M., König, C. J., & Melchers, K. G. (2015). Shall we continue or stop disapproving of self-presentation? Evidence on impression management and faking in a selection context and their relation to job performance. *European Journal of Work and Organizational Psychology*, 24, 420–432. doi:10.1080/1359432x.2014.915215
- Jansen, A., Melchers, K. G., Kleinmann, M., Lievens, F., Brändli, M., Fraefel, L., & König, C. J. (2013). Situation assessment as an ignored factor in the behavioral consistency paradigm underlying the validity of personnel selection procedures. *Journal of Applied Psychology: An International Review*, 98, 326–341. doi:10.1037/a0031257
- Judge, T. A., & Zapata, C. P. (2015). The person–situation debate revisited: Effect of situation strength and trait activation on the validity of the Big Five personality traits in predicting job performance. *Academy of Management Journal*, 58, 1149–1179. doi:10.5465/amj.2010.0837
- König, C. J., Hafsteinsson, L. G., Jansen, A., & Stadelmann, E. H. (2011). Applicants' self-presentational behavior across cultures: Less self-presentation in Switzerland and Iceland than in the US. *International*

- Journal of Selection and Assessment*, 19, 331–339. doi:10.1111/j.1468-2389.2011.00562.x
- König, C. J., Melchers, K. G., Kleinmann, M., Richter, G. M., & Klehe, U.-C. (2006). The relationship between the ability to identify evaluation criteria and integrity test scores. *Psychology Science*, 48, 369–377.
- 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. doi:10.1111/ijsa.12007
- König, C. J., Mura, M., & Schmidt, J. (2015). Applicants' strategic use of extreme or midpoint responses when faking personality tests. *Psychological Reports*, 117, 429–436. doi:10.2466/03.02.PRO.117c21z2
- Kuncel, N. R., Kochevar, R. J., & Ones, D. S. (2014). A meta-analysis of letters of recommendation in college and graduate admissions: Reasons for hope. *International Journal of Selection and Assessment*, 22, 101–107. doi:10.1111/ijsa.12060
- Levin, R. A., & Zickar, M. J. (2002). Investigating self-presentation, lies and bullshit: Understanding faking and its effects on hiring decisions using theory, field research, and simulation. In J. M. Brett & F. Drasgow (Eds.), *The psychology of work: Theoretically based empirical research* (pp. 253–276). Mahway, NJ: Erlbaum.
- Marcus, B. (2009). "Faking" from the applicant's perspective: A theory of self-presentation in personnel selection settings. *International Journal of Selection and Assessment*, 17, 417–430. doi:10.1111/j.1468-2389.2009.00483.x
- McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology*, 85, 812–821. doi:10.1037/0021-9010.85.5.812
- Mitchell, G. (2012). Revisiting truth or triviality: The external validity of research in the psychological laboratory. *Perspectives on Psychological Science*, 7, 109–117. doi:10.1177/1745691611432343
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, 60, 683–729. doi:10.1111/j.1744-6570.2007.00089.x
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York, NY: McGraw-Hill.
- O'Neill, T. A., Lewis, R. J., Law, S. J., Larson, N., Hancock, S., Radan, J., ... Carswell, J. J. (in press). Forced-choice pre-employment personality assessment: Construct validity and resistance to faking. *Personality and Individual Differences*, doi:10.1016/j.paid.2016.03.075
- Oh, I.-S., Wang, G., & Mount, M. K. (2011). Validity of observer ratings of the five-factor model of personality traits: A meta-analysis. *Journal of Applied Psychology*, 96, 762–773. doi:10.1037/a0021832
- Ones, D. S., Dilchert, S., Viswesvaran, C., & Judge, T. A. (2007). In support of personality assessment in organizational settings. *Personnel Psychology*, 60, 995–1027. doi:10.1111/j.1744-6570.2007.00099.x
- 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. doi:10.1037/0021-9010.81.6.660
- Peterson, M. H., Griffith, R. L., Isaacson, J. A., O'Connell, M. S., & Mangos, P. M. (2011). Applicant faking, social desirability, and the prediction of counterproductive work behaviors. *Human Performance*, 24, 270–290. doi:10.1080/08959285.2011.580808
- Poropat, A. E. (2014). Other-rated personality and academic performance: Evidence and implications. *Learning and Individual Differences*, 34, 24–32. doi:10.1016/j.lindif.2014.05.013
- 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. doi:10.1108/02683940610690204
- Roulin, N., & Krings, F. (2016). When winning is everything: The relationship between competitive worldviews and job applicant faking. *Applied Psychology: An International Review*, 65, 643–670. doi:10.1111/apps.12072
- Roulin, N., Krings, F., & Binggeli, S. (2016). A dynamic model of applicant faking. *Organizational Psychology Review*, 6, 145–170. doi:10.1177/2041386615580875
- Rynes, S. L., Colbert, A. E., & Brown, K. G. (2002). HR professionals' beliefs about effective human resource practices: Correspondence between research and practice. *Human Resource Management*, 41, 149–174. doi:10.1002/hrm.10029
- Salgado, J. F. (2016). A theoretical model of psychometric effects of faking on assessment procedures: Empirical findings and implications for personality at work. *International Journal of Selection and Assessment*, 24, 209–228. doi:10.1111/ijsa.12142
- Snell, A. F., Sydel, E. J., & Lueke, S. B. (1999). Towards a theory of applicant faking: Integrating studies of deception. *Human Resource Management Review*, 9, 219–242. doi:10.1016/S1053-4822(99)00019-4
- Tett, R. P., & Christiansen, N. D. (2007). Personality tests at the crossroads: A response to Morgeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt (2007). *Personnel Psychology*, 60, 967–993. doi:10.1111/j.1744-6570.2007.00098.x
- Vanhove, A. J., & Harms, P. D. (2015). Reconciling the two disciplines of organisational science: A comparison of findings from lab and field research. *Applied Psychology: An International Review*, 64, 637–673. doi:10.1111/apps.12046
- Ziegler, M., MacCann, C., & Roberts, R. D. (2011). Faking: Knowns, unknowns, and points of contention. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 3–16). New York, NY: Oxford University Press.
- Zimmerman, R. D., Triana, M. d. C., & Barrick, M. R. (2010). Predictive criterion-related validity of observer ratings of personality and job-related competencies using multiple raters and multiple performance criteria. *Human Performance*, 23, 361–378. doi:10.1080/08959285.2010.501049

How to cite this article: König CJ, Steiner Thommen LA, Wittwer A-M, Kleinmann M. Are observer - ratings of applicants' personality also faked? Yes, but less than self-reports. *Int J Select Assess*. 2017;25:183–192. <https://doi.org/10.1111/ijsa.12171>

APPENDIX

This Appendix displays the German items used in this study (in the male observer-report form) plus the English versions (see Goldberg et al., 2006).

Scale achievement striving	
Er verfolgt stets seine Ziele.	He goes straight for the goal.
Er arbeitet hart.	He works hard.
Pläne setzt er in Taten um.	He turns plans into actions.
In Aufgaben stürzt er sich mit Herzblut.	He plunges into tasks with all his heart.
Er tut mehr, als von ihm erwartet wird.	He does more than what is expected of him.
Er setzt sich hohe Ziele für sich selbst und andere.	He sets high standards for himself and others.
Qualität ist ihm wichtig.	He demands quality.
Er ist nicht sehr erfolgsmotiviert.	He is not highly motivated to succeed. (reverse coded)
Er leistet gerade genug um durchzukommen.	He does just enough work to get by. (reverse coded)
Er investiert wenig Zeit und Aufwand in seine Arbeit.	He puts little time and effort into his work. (reverse coded)
-	Get chores done right away. (not used because not work-related)
-	He finds it difficult to get down to work. (item not used because it captures an inner process that cannot be rated by an observer)
Self-discipline	
Er ist immer gut vorbereitet.	He is always prepared.
Er geht Aufgaben sofort an.	He starts tasks right away.
Mit seiner Arbeit beginnt er umgehend.	He gets to work at once.
Seine Vorhaben führt er aus.	He carries out his plans.
Er vergeudet seine Zeit.	He wastes his time. (reverse coded)
Er braucht einen Anstoss, um eine Aufgabe zu beginnen.	He needs a push to get started. (reverse coded)
Er hat Schwierigkeiten, Aufgaben anzugehen.	He has difficulty starting tasks. (reverse coded)

(Continues)

(Continued)

Self-discipline	
Entscheidungen zögert er hinaus.	He postpones decisions. (reverse coded)
Cautiousness	
Fehler vermeidet er.	He avoids mistakes.
Er wählt seine Worte mit Bedacht.	He chooses his words with care.
Er hält sich an gewählte Wege.	He sticks to his chosen path.
Er stürzt sich unüberlegt in Dinge.	He jumps into things without thinking.
Er trifft voreilige Entscheidungen.	He makes rash decisions.
Er handelt gerne nach Lust und Laune.	He likes to act on a whim.
Er geht Dinge überstürzt an.	He rushes into things.
Er handelt ohne zu denken.	He acts without thinking.
Pläne macht er oft in letzter Minute.	He often makes last-minute plans.
-	He does crazy things. (not used because it is not enough work-related)
Assertiveness	
Er übernimmt die Verantwortung.	He takes charge.
Er versucht andere zu führen.	He tries to lead others.
Er kann andere überreden, Dinge zu tun.	He can talk others into doing things.
Er versucht andere zu beeinflussen.	He seeks to influence others.
Er übernimmt Kontrolle über Dinge.	He takes control of things.
Er überlässt die Führung anderen.	He waits for others to lead the way. (reverse coded)
Er bleibt im Hintergrund.	He keeps in the background. (reverse coded)
Er hat wenig zu sagen.	He has little to say. (reverse coded)
Er zieht ungern Aufmerksamkeit auf sich.	He doesn't like to draw attention to himself. (reverse coded)
Er hält sich mit seiner Meinung zurück.	He holds back his opinions. (reverse coded)