What if applicants knew how personality tests are scored? A minimal intervention study

Article *in* Journal of Personnel Psychology · January 2017

CITATIONS
0

READS
30

3 authors, including:

Cornelius J. König
Universität des Saarlandes
96 PUBLICATIONS **1,253** CITATIONS
SEE PROFILE

Anne Jansen
University of Applied Sciences and Arts North…
13 PUBLICATIONS **125** CITATIONS
SEE PROFILE

What if applicants knew how personality tests are scored? A minimal intervention study

In press at the *Journal of Personnel Psychology*

Cornelius König

Universität des Saarlandes

Anne Jansen

Fachhochschule Nordwestschweiz

Peter Lüscher Mathieu

Universität Zürich

We thank Neil Christiansen for his feedback on this research project. Cornelius König and Anne Jansen worked previously at the Universität Zürich in Switzerland. Peter Lüscher Mathieu works now for Optimus AG, Thun, Switzerland. Correspondence concerning this article should be addressed to Cornelius J. König, Universität des Saarlandes, AE Arbeits- und Organisationspsychologie, Campus A1 3, D-66123 Saarbrücken, Germany, ckoenig@mx.uni-saarland.de.

Abstract

Having access to information on personality tests might make faking of personality tests easier because applicants are not hindered by incorrect assumptions about the scoring. Thus, this experiment tests whether very briefly telling applicants how personality tests are scored affects faking. Management assistants (N = 187), asked to imagine themselves as job applicants, were either informed about the scoring key or were given no information before filling out a Big Five personality test. Results revealed that this minimal manipulation increased faking. This finding supports the notion that applicants often incorrectly assume that scoring procedures are overly complex and gives practitioners additional reasons to worry about more future faking.

Keywords: faking; self-presentation; personality tests; personnel selection

What if applicants knew how personality tests are scored? A minimal intervention study

Faking of personality tests describes the phenomenon that a considerable number of applicants present themselves in a positive way when they complete a personality test. Specifically, individuals' personality scores, for example on Big Five dimensions, are higher if they fill them out as applicants than if they fill them out honestly (Ziegler, MacCann, & Roberts, 2011). Although several researchers have explored the reasons why applicants fake, less attention has been paid to why applicants do not fake more than they seem to do. In fact, faking on standard personality tests seems to be a rather simple task if you know how they are used: You only need to detect what the positive pole of each item is and then you mark the most positive answer.

Although personnel selection scholars and practitioners will surely know the scoring of standard personality tests (i.e., non-ipsative tests, no ideal-point scoring, e.g., Carter et al., 2014), this is less likely for applicants. In fact, König, Merz, and Trauffer (2012) found that applicants often do not know how personality tests work. For example, some applicants interviewed by König et al. assumed that hiring companies might positively evaluate the endorsement of the midpoint category because endorsing extreme values could be seen as a sign of obstinacy. An applicant with such an assumption will likely refrain from endorsing "strongly agree" when answering a socially desirable item, although endorsing this would have been a more effective faking strategy. Other applicants believed that tests measure whether they answer consistently, which can lead applicants to give similar ratings to similarly appearing items. Thus, if they have not faked one item they might be reluctant to fake another, similar item (cf. König, Mura, & Schmidt, 2015).

Such incorrect assumptions about personality could be corrected by information on how personality tests are scored. Accordingly, we designed a minimal intervention study to test the hypothesis that briefly informing individuals on how personality tests are scored increases faking behavior in an applicant setting. Minimal intervention studies (Prentice & Miller, 1992; cf. Goldberg, 1999) aim to show the importance of effects by using a minimal manipulation and often focus on a core psychological process (in this case, on correcting assumption regarding personality tests). In comparison to previous studies on teaching how to fake (a standard training in the study by Miller & Barrett, 2008; a video-based coaching in the study by Robie, Komar, & Brown, 2010; and self-coaching using an advice book on how to beat personality tests, Sliter & Christiansen, 2012), our intervention is much briefer: We simply changed two paragraphs in the instruction to our participants, assuming that it would be enough to find effects on faking.

Method

Sample

Participants were 187 Swiss management assistants (in German: "Betriebswirtschafterin/ Betriebswirtschafter"), 95 of whom were female. The mean age was 27.7 (*SD* = 5.4). Two additional participants were excluded because they skipped a whole page of the personality test (see below). Participants worked in various sectors, particularly in the financial (25.1%), the manufacturing (16.6%), and the service industry (10.7%). All participants were additionally enrolled at a "Höhere Fachschule für Wirtschaft" in a "General Management" diploma course, a Swiss practice-oriented tertiary educational degree that enables students to qualify for management positions after obtaining an apprenticeship. To enrol in the course, people need to have worked for at least two years after finishing their apprenticeship and they need to be

working at least part-time, although they often work full-time (mean in our sample: 91.2%, *SD* = 14.4%).

Procedure and manipulation

We offered a two-hour personnel selection preparation course for people enrolled at a "Höhere Fachschule für Wirtschaft." The course informed participants about typical selection instruments, and were given the opportunity to gain experience with some of them.

Participants were given a job advertisement of a fictitious large company seeking diploma holders from a "Höhere Fachschule für Wirtschaft." The ad contained cues for all Big Five dimensions: The company was seeking somebody who is "open for new challenges" (Openness), "works in a conscientious way" (Conscientiousness), "actively approaches others" (Extraversion), "gets along well with others" (Agreeableness), and "is emotionally even-tempered" (Emotional Stability).

After reading the ad, participants received some information about personality tests. They were all informed that applicants express their level of agreement with statements about different personality aspects by ticking an answer category, and that the 'best answer' for a socially desirable trait is thus "strongly agree." The level of agreement will then be used to derive conclusions about the traits of applicants and their likely behavior at work. We also explained that the personality is classified into different traits, depending on theories, and were shown two sample items measuring self-confidence.

The next part of the information text depended on the group (informed vs. not informed) to which participants were randomly assigned. The experimental group (n = 93) was informed (using the two self-confidence items as examples) that test administrators simply calculate the average across all items that describe a particular trait (after reverse-coding, if needed). The

control group (n = 94) was told that there are other personnel selection tests as well (e.g., integrity tests) that can be administered online and that online administration has its advantages and disadvantages.

Then, all participants were instructed to fill out the personality tests as if they were applicants for this position ("as applicant" condition). They filled out the NEO Five Factor Inventory (NEO-FFI, in its German translation, Borkenau & Ostendorf, 1993), a 60-item measure of the Big Five (Conscientiousness, Agreeableness, Extraversion, Openness for experience, and Emotional Stability) with 60 items (12 per factor). We offered 50 Swiss Francs to the best 5% as an incentive (around 52 US $).

Half an hour later, participants were asked to fill out the same personality test again but with the instruction to be honest ("honest" condition; for an empirical justification as to why it is better to have the applicant condition before the honest condition, see Hooper, 2007, who has meta-analytically shown that faking research in the lab with this order of conditions produces similar effect sizes to effect sizes in field studies).

Regression-adjusted difference scores

We operationalized faking (our dependent variable) as five difference scores between the "as applicant" and the faking conditions, one for each of the Big Five traits. More precisely, given the controversies about using ordinary difference scores for measuring change between "as applicant" and "honest" scores, we used regression-adjusted difference scores (developed by Pedhazur & Schmelkin, 1991; cf. Burns & Christiansen, 2011). Regression-adjusted difference scores partial out the autocorrelations of the "as applicant" scores with respective "honest" scores. Practically, this involves regressing the "as applicant" condition on the "honest"

condition (no matter whether participants belonged to the informed or the not-informed group), saving the residuals, and running the analyses with these residuals.

Results

Table 1 shows the results of *t*-tests with the regression-adjusted difference scores: All effect sizes were in the expected direction and significant for Conscientiousness, Extraversion, and Agreeableness, but were not significant for Emotional Stability or Openness. However, applicants in both groups rarely faked to the maximum (see Table 2).

As an additional test, we compared the "as applicant" means of the informed group with those of the control group (i.e., without adjustments for the "honest" condition scores). The results reported in Table 2 confirmed the regression-adjusted difference scores findings (the only difference being that the effect size for Emotional Stability turned additionally significant).

Discussion

The aim of this minimal intervention study was to test whether informing applicants about the way personality tests are scored affects applicants' faking. We found that a very brief manipulation led to more faking on a measure of Conscientiousness, Extraversion, and Agreeableness. The effect on Emotional Stability only became significant when the means of the "as applicant" condition were compared. The effect on Openness was in the expected direction but did not reach significance. This study shows that merely changing two paragraphs (that explain how personality tests are scored) gave participants enough knowledge to achieve significantly higher scores on a Big Five personality test. Furthermore, the minimal manipulation resulted in a medium effect in the case of Conscientiousness, which is similar to the results obtained by previous studies that have used more intense interventions (Miller & Barrett, 2008; Robie et al., 2010; Sliter & Christiansen, 2012).

The results of this experiment is also consistent with König et al.'s (2012, 2015) argument that applicants may hold assumptions about the scoring of personality tests that differ from the simple scoring procedures used by test developers and users. The standard scoring means that the test uses averages across all items that belong to a particular trait (exceptions being ipsative measures and ideal point methods, see, e.g., Carter et al., 2014). If applicants have incorrect assumptions about the scoring of personality tests, this likely hinders them from engaging in faking.

Interestingly, if applicants do know how personality tests are scored, they still only rarely achieve maximum scores (see Table 3). In particular, although the job ad contained cues that all Big Five dimensions are desirable, maximum scores were only slightly more common (and only for Conscientiousness and, to a lesser extent, for Emotional Stability). This suggests that applicants could have faked more but did not. It is possible that applicants were not able to detect the relevant pole for every item (cf. Jansen et al., 2013), but maximum scores can only be achieved if participants endorse *strongly agree* for all positively scored items and *strongly disagree* for all negatively scored items. It is also possible that cross-cultural differences play a role, since our participants were Swiss, and the Swiss do not appear to fake much (König, Hafsteinsson, Jansen, & Stadelmann, 2011). In addition, it has been suggested that applicants differ in their belief that the world is a competitive jungle in which only the fittest survive (cf. Roulin & Krings, 2016) vs. that faking is always morally wrong (cf. Goffin & Boyd, 2009), and if our participants shared the latter belief, this might have limited their motivation to fake.

If applicants indeed have complex and often incorrect assumptions about the interpretation of personality tests, future research should try to understand who holds what assumptions and why. For example, some applicants seem to marvel at the skills of psychologists

who develop personality tests (cf. König et al., 2012). Thus, those applicants might trust psychologists' abilities to develop sophisticated scoring. Future research should also explore who are the applicants who search for information regarding tests.

As for all studies, some limitations have to be noted. First, participants only imagined applying for a job, and the use of a hypothetical situation precluded that this study contributed to the ongoing debate about whether faking (a) negatively affects the criterion-related validity of personality tests (e.g., Donovan, Dwight, & Schneider, 2014), (b) does not meaningfully affect the criterion-related validity (e.g., Hogan, Barrett, & Hogan, 2007), or (c) positively affects the criterion-related validity (e.g., Ingold, Kleinmann, König, & Melchers, 2015). However, it would have been unethical for us to use an experimental manipulation in a non-hypothetical field situation, because the manipulation would have increased the chances of the informed applicants. Second, this experiment focused on only one variable (i.e., knowledge about the scoring), and explaining the phenomenon of faking undoubtedly needs to take more variables into account (see, e.g., Goffin & Boyd, 2009; Marcus, 2009; McFarland & Ryan, 2000). Third, our analyses would have benefitted from including control variables. For example, it is possible some participants were more familiar with the coaching literature on how to behave in selection situations than other participants were (i.e., had more knowledge about tests), and this familiarity might have influenced how much participants faked independently of the manipulation. Furthermore, the manipulation could have interacted with other variables such as the familiarity with the coaching literature (i.e., the manipulation has a stronger effect for unfamiliar participants than for familiar participants). Future research is needed to explore this.

To conclude, our study shows that applicants do not need to spend much time or money preparing themselves for completing a personality test. If an applicant wants to raise his or her

scores, it simply helps her or him to know one crucial piece of information: how personality tests are scored.

References

Borkenau, P., & Ostendorf, F. (1993). *NEO-Fünf-Faktoren Inventar (NEO-FFI) nach Costa & McCrae [The NEO-Five-Factor Inventory (NEO-FFI) after Costa & McCrae)*. Göttingen, Germany: Hogrefe.

Burns, G. N., & Christiansen, N. D. (2011). Methods of measuring faking behavior. *Human Performance, 24*, 358-372. doi:10.1080/08959285.2011.597473

Carter, N. T., Dalal, D. K., Boyce, A. S., O'Connell, M. S., Kung, M.-C., & Delgado, K. M. (2014). Uncovering curvilinear relationships between conscientiousness and job performance: How theoretically appropriate measurement makes an empirical difference. *Journal of Applied Psychology, 99*, 564-586. doi:10.1037/a0034688

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159. doi:10.1037/0033-2909.112.1.155

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

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.

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. (2007). *Self-presentation on personality measures in lab and field settings: A meta-analysis.* (Doctoral dissertation). ProQuest Dissertations and Theses database.

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, 98*, 326-341. doi:10.1037/a0031257

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., 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.PR0.117c21z2

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

Miller, C. E., & Barrett, G. V. (2008). The coachability and fakability of personality-based selection tests used for police selection. *Public Personnel Management, 37*, 339-351. doi:10.1177/009102600803700306

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, NJ: Erlbaum.

Prentice, D. A., & Miller, D. T. (1992). When small effects are impressive. *Psychological Bulletin, 112*, 160-164. doi:10.1037/0033-2909.112.1.160

Robie, C., Komar, S., & Brown, D. J. (2010). The effects of coaching and speeding on Big Five and impression management scale scores. *Human Performance, 23*, 446-467. doi:10.1080/08959285.2010.515278

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

Sliter, K. A., & Christiansen, N. D. (2012). Effects of targeted self-coaching on applicant distortion of personality measures. *Journal of Personnel Psychology, 11*, 169-175. doi:10.1027/1866-5888/a000069

Ziegler, M., MacCann, C., & Roberts, R. (Eds.). (2011). *New perspectives on faking in personality assessment*. New York, NY: Oxford University Press.

Table 1

Mean differences in faking (i.e., in regression-adjusted difference scores)

Dimension	Informed group (n = 93)		Not-informed group (n = 94)		t-value	Cohen's d [a]
	M	*SD*	*M*	*SD*		
Conscientiousness	0.07	0.30	-0.07	0.29	3.24**	0.48
Emotional Stability	0.03	0.38	-0.03	0.41	1.11	0.15
Extraversion	0.04	0.24	-0.04	0.27	2.18*	0.31
Openness	0.02	0.27	-0.02	0.25	0.96	0.15
Agreeableness	0.04	0.28	-0.04	0.26	1.78*	0.30

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

[a] According to Cohen (1992), effect sizes of $d \approx .50$ can be considered as medium effects and effect sizes of $d \approx .20$ as small effects.

Table 2

Additional analyses: Mean comparisons of the "as applicant" condition

Dimension	Informed group (n = 93)			Not-informed group (n = 94)			t-value	Cohen's d [b]
	M [a]	SD	%max	M [a]	SD	%max		
Conscientiousness	4.53	0.37	12.9%	4.33	0.41	2.1%	3.46**	0.51
Emotional Stability	4.19	0.52	3.2%	4.04	0.57	0%	1.79*	0.28
Extraversion	3.96	0.37	0%	3.78	0.40	0%	3.22**	0.47
Openness	3.50	0.44	0%	3.43	0.44	0%	1.03	0.09
Agreeableness	3.98	0.41	0%	3.85	0.42	0%	2.09*	0.31

Note. * $p < .05$, one-tailed; ** $p < .01$, one-tailed. %max = percentage of participants who achieved the maximum score.

[a] On a scale from 1 (strongly disagree) to 5 (strongly agree)

[b] According to Cohen (1992), effect sizes of $d \approx .50$ can be considered as medium effects and effect sizes of $d \approx .20$ as small effects.