This preprint version is currently in press in the Journal of Business and Psychology and may not exactly replicate the final version published as:
Buehl, A.-K., Melchers, K. G., Macan, T., & Kühnel, J. (in press). Tell me sweet little lies: How does faking in interviews affect interview scores and interview validity? *Journal of Business and Psychology*. doi:10.1007/s10869-018-9531-3

Tell me sweet little lies:

How does faking in interviews affect interview scores and interview validity?

Anne-Kathrin Buehl and Klaus G. Melchers

Universität Ulm, Germany

Therese Macan

University of Missouri - St. Louis

Jana Kühnel

Universität Ulm, Germany

Author Note

Anne-Kathrin Buehl, Klaus G. Melchers, and Jana Kühnel, Institut für Psychologie und Pädagogik, Universität Ulm, Germany, Therese Macan, Department of Psychological Sciences, University of Missouri – St. Louis, USA.

We thank Barbara Körner, Karin Eigenseer, Jan-Philipp Schulz, Sina Bulling, Evelyn Schuwerk, Tina Gösel, and Cora Grässle for help with the data collection and/or coding of the videotapes.

Correspondence concerning this article should be addressed to Anne-Kathrin Buehl or Klaus Melchers, Institut für Psychologie und Pädagogik, Universität Ulm, Albert-Einstein-Allee 41, D-89069 Ulm, Germany, e-mail: anne-kathrin.buehl@alumni.uni-ulm.de or klaus.melchers@uni-ulm.de.

Abstract

Interviews are a prevalent technique for selection and admission purposes. However, interviews are also viewed as potentially fakeable, raising the question of whether interviewees' faking behavior impairs the quality of selection decisions. To address these concerns, our study examined whether interviewees can actually improve their interview score by faking and the role that interviewee ability factors play in interview faking. We also explored the effect of faking on criterion-related validity with regard to successfully predicting interviewees' task and contextual performance. We conducted simulated interviews in an honest and an applicant instruction condition using a within-subjects design. In line with our hypotheses, interviewees were able to improve their interview scores when asked to respond as an applicant. The size of the improvement of these interview scores correlated with interviewees' cognitive ability and their ability to identify the targeted interview dimensions. Concerning the effects of faking on criterion-related validity, we found that academic performance was better predicted in the applicant instruction condition whereas contextual performance was better predicted in the honest condition. Thus, it appears that claims that "faking impairs criterion-related validity" are too simplified and that we have to consider the kind of criterion predicted.

Keywords: selection interview; faking; criterion-related validity; contextual performance; ability to identify criteria

Tell me sweet little lies:

How does faking in interviews affect interview scores and interview validity?

Interviews are a prevalent and globally used technique for selection, including university admission decisions. Given that interviewers' judgments of applicants' qualities are largely based on what these applicants say and do during an interview, estimates that suggest over 90% of undergraduate job candidates show at least some forms of faking behavior during interviews seem disturbing (Levashina & Campion, 2007). But despite its prevalence, there are relatively few previous studies that have examined the degree to which faking in interviews actually leads to different evaluations and whether it has negative consequences for personnel selection.

To date we have limited knowledge of the effects of faking behavior in interviews (Levashina, Hartwell, Morgeson, & Campion, 2014). Specifically, even though it seems obvious that interviewees use faking behaviors to improve their performance in selection interviews, previous research has not been unequivocal with regard to the question whether faking indeed leads to better evaluations of individuals' interview performances (Allen, Facteau, & Facteau, 2004; Van Iddekinge, Raymark, & Roth, 2005). Furthermore, especially from an applied point of view, it is also important to know whether faking impairs interview criterion-related validity given that attempts to find strategies to reduce interviewees faking behavior have remained inconclusive (Levashina & Campion, 2007) and because interviewers are barely able to detect faking in interviews (Reinhard, Scharmach, & Müller, 2013; Roulin, Bangerter, & Levashina, 2014). Finally, despite the common concern that faking might be detrimental for the quality of selection decisions, such an impairment is only predicted by some models (e.g., Salgado, 2016) whereas others even suggest that faking might have a positive impact on criterion-related validity in certain cases (e.g., Marcus, 2009). That is,

ability-related factors that are needed to show effective faking behavior may also contribute to the prediction of job performance and result in higher criterion-related validity estimates.

As an attempt to reduce these gaps in our knowledge, the aim of the present study was to investigate the extent to which faking in selection interviews is possible and the corresponding role that faking plays on criterion-related validity. Specifically, we try to answer the following three questions: First, can interviewees actually improve their interview scores by faking? Second, what role do interviewee ability factors play with regard to faking in interviews? And third, what effect does faking in interviews have on our ability to successfully predict interviewees' job or academic performance?

Theoretical Background

Faking in interviews is defined as "an intentional distortion or a falsification of responses on measures in order to create a specific impression or provide the best answer" (Levashina & Campion, 2006, p. 300). Yet despite the limited empirical research on faking in interviews, there are several theoretical models on faking in general (Marcus, 2009; McFarland & Ryan, 2000, 2006; Roulin, Krings, & Binggeli, 2016; Snell, Sydell, & Lueke, 1999) and also on faking in interviews in particular (Levashina & Campion, 2006).

What most of these models have in common is that characteristics of the applicant and the situation determine whether, and if so, to what extent applicants engage in faking behavior. For the interviewee, influencing factors could be, for instance, personality, values, or beliefs (McFarland & Ryan, 2006). Furthermore, attractiveness of the position, faking warnings, or the current unemployment rate represent relevant situational influences (McFarland & Ryan, 2006). In a further step, an interviewee's ability determines whether this faking behavior is successful such that it leads to better evaluations in the interview and eventually a job offer. Thus, to be successful in an interview, interviewees may need certain abilities in order to adapt their faking behavior to the requirements of the interview. These

abilities could be oral communication skills, social skills, cognitive ability, or the ability to identify the constructs that are targeted in an interview (Levashina & Campion, 2006).

A construct that is related to faking is impression management (IM). IM means that an individual tries to create or convey a particular impression of him- or herself to another person (Leary & Kowalski, 1990). IM and faking in interviews are related but nevertheless not synonymous constructs. Specifically, IM can be conceptualized as a continuum of tactics going from complete honesty to outright lying about one's qualities (cf. Levashina & Campion, 2007). The lower part of the continuum could be defined as honest IM where applicants use IM tactics to create a positive impression but rely on facts and truthful descriptions of their qualities (cf. Roulin & Bourdage, 2017). The upper part of the continuum could be considered as deceptive IM where applicants use more or less deceptive forms of IM tactics to create a positive impression (e.g., they pretend to have certain skills or to hold certain views that they do not actually hold or possess) with slightly deceptive tactics being closest to honest IM and extremely deceptive tactics being closest to outright lies. Interview faking thus incorporates deceptive forms of IM (Roulin & Bourdage, 2017).

Various previous studies have examined IM in general and specifically in employment interviews. Even though these studies usually did not distinguish between honest versus deceptive IM, this research can be informative. Specifically, this research found that applicants use different verbal tactics including assertive approaches that aim at actively conveying a positive image (e.g., self-promotion, ingratiation) and defensive tactics that aim at protecting or repairing threatened images by apologizing for, excusing, or justifying one's actions or attributes. Furthermore, meta-analytic investigations about the use of verbal IM tactics revealed that they are positively correlated with interview scores (Barrick, Shaffer, & DeGrassi, 2009; Levashina et al., 2014). Thus, interviewees who use IM tactics to a stronger degree are evaluated more positively by interviewers. In addition, the effects of IM on interview performance are stronger in unstructured interviews than in structured interviews

(Barrick et al., 2009). Given the general effects of IM, we follow Gilmore's and Ferris's (1989) call for investigating the deceptive side of IM: faking. We begin with a brief summary of empirical findings regarding faking and move to develop our hypotheses.

Previous Research on Faking in Interviews and Development of Study Hypotheses

As already noted above, faking is a prevalent phenomenon in personnel selection. According to a study by Weiss and Feldman (2006), 81% of job applicants admitted they told at least one lie in their last interview, with an average of 2.19 lies per interview. Furthermore, Levashina and Campion (2007) found that at least 93% of their survey participants had used one faking tactic or more during a recent employment interview.

Whether applicants fake and to what extent they display such faking behavior is determined by aspects of the interviewee, the interview, and the interviewer. Most research has been conducted on the role of interviewees' personality. This research has found that individuals who are higher on self-monitoring, Machiavellianism, and extraversion (Hogue, Levashina, & Hang, 2013; Levashina & Campion, 2007; Roulin & Bourdage, 2017; Weiss & Feldman, 2006) and lower on integrity and honesty-humility tend to fake more (Buehl & Melchers, 2017; Law, Bourdage, & O'Neill, 2016; Levashina & Campion, 2007; Roulin & Bourdage, 2017). However, research is missing that investigates the relationship between interviewee ability and faking behavior. It has been found, however, that question type influences interviewees' faking behavior such that the use of situational questions that ask interviewees how they would handle a hypothetical situation as well as follow-up questioning foster faking behavior (Levashina & Campion, 2007).

A potential problem for applied contexts is that interviewers in general are not very good at detecting faking in interviews (Reinhard et al., 2013; Roulin et al., 2014). Even experience in hiring employees or the number of previously conducted interviews does not seem to improve the likelihood that interviewers are able to detect faking (Roulin et al.,

2014). Moreover, to date, there is little evidence for successful strategies to reduce applicant faking behavior (but see Law et al., 2016, for a recent exception where warning instructions reduced faking in a simulated interview). Instead, follow-up questioning, which had previously been suggested as a means to validate or falsify suspicious answers (Levashina & Campion, 2006), has been found to increase faking behavior (Levashina & Campion, 2007).

In sum, we know what faking is, that people report that they do fake, that certain personality characteristics are related to individuals' propensity to fake, and that at this moment, little can be done about faking. This leaves open a number of questions, specifically regarding: (a) the degree to which interviewees can actually augment their interview performance intentionally through faking, (b) the role of interviewees' ability to fake, and (c) whether faking matters with regard to criterion-related validity?

Effects of Faking on Interview Performance

Investigating whether interviewees who fake can improve how favorably interviewers evaluate them is not as easy as it may seem. In contrast to research on personality assessment, for example, there are very few studies that have actually dealt with the effects of faking on interview performance. To our knowledge, there are only two published studies that compared interviewees' performance in an honest versus a faking condition, and these two studies found diverging results. Allen et al. (2004) found that participants who were instructed to act as if they really wanted and needed the job could not increase their interview scores compared to a control group that was asked to answer honestly. In contrast, Van Iddekinge et al. (2005) found a significant mean difference in interview scores between a group that was instructed to put their "best foot forward" and another group instructed to answer honestly.

Both studies used a between-subjects design with an honest condition and an applicant or faking condition. In addition, Allen et al. (2004) had a second faking condition in which the dimensions of the interview were made transparent to interviewees (Klehe, König,

Richter, Kleinmann, & Melchers, 2008). However, participants in Allen et al.'s faking conditions were only asked to imagine being interviewed for an attractive job. By contrast, Van Iddekinge et al. (2005) tried to create an applicant setting as representative of an actual hiring situation as possible. Their study was conducted as part of a selection training program for students who would soon be on the job market. According to Van Iddekinge et al., students dressed as for an actual job interview, they appeared to be nervous and most of them requested feedback afterwards. Experienced interviewers conducted the interviews in a professional setting in offices within the university's career center.

In Van Iddekinge et al.'s (2005) applicant condition, interviewees were also asked to answer as if they were applying for a job that they really wanted. Furthermore, as an incentive to put their best foot forward, participants received $50 if they were one of the top scorers. With all these design provisions, Van Iddekinge et al. found a significant mean difference between the honest and applicant conditions, albeit of moderate size. Given the social evaluative nature of interviews, however, it is possible that the faking effect they found could be an underestimation.

A crucial element for faking studies using interviews that are conducted in a simulated setting is not only to motivate interviewees in the faking condition to put their best foot forward but also to make sure that they are telling the truth in the honest condition. In their honest condition, Van Iddekinge et al. (2005) applied all the design features for creating a realistic application atmosphere that we described above. However, instead of asking participants to put their best foot forward, they were asked to answer truthfully. Yet because participants in this condition knew their participation was part of a training program and professional interviewers conducted the interviews, it is possible that they did not answer as truthfully as possible. Instead, they may have wanted to practice for an actual interview and/or they may have perceived an evaluative component to the environment, limiting their ability to

be truthful. Thus, the size of the faking effect could be an underestimate if interviewees exaggerated their answers in the honest condition to make a good impression.

Martin and Nagao's (1989) study also underlines the social evaluative nature of interviews. They showed that participants engaged in less distortion in computerized interviews compared to face-to-face interviews. That is, participants scored lower on a scale for measuring socially desirable responding and inflated their GPAs less when they had to respond to the corresponding questions in a computerized interview compared to a face-to-face condition. Martin and Nagao argued that this difference exists because of less social pressure on interviewees in the computerized interviews. Similarly, it has been argued that lying is part of human nature, mostly as a protection mechanism against getting hurt or hurting others' feelings (DePaulo et al., 2003). Therefore, we know that for studies aiming to obtain true answers from participants, it is crucial to establish an environment of trust, where people are willing to reveal the truth about themselves.

Given this possibility, the mean difference obtained by Van Iddekinge et al. (2005) between the faking and the honest condition may be an underestimate of the true difference because the interview in the honest condition was also administered in the same application training context with the same group of professional interviewers. That is, even in the honest condition, interviewees might have been motivated to a certain degree to make a good impression when responding to interviewers. Accordingly, if it is possible to create an interview situation in which interviewees are willing to provide (at least relatively) truthful answers, we can better assess the effects of interview faking behavior. Therefore, and based on Van Iddekinge et al.'s findings, we propose that:

Hypothesis 1: Interviewees will receive more favorable interview scores in the faking condition than in the honest condition.

Ability to Fake

Even though the faking models theoretically consider ability as one of the key elements that influence the occurrence of faking (Levashina & Campion, 2007; Marcus, 2009; McFarland & Ryan, 2000, 2006; Roulin, Bangerter, & Levashina, 2015; Snell et al., 1999), research on the role of ability in interview faking is missing. The different models have suggested that aspects such as oral communication skills, social skills, cognitive ability, or the ability to identify the targeted constructs are relevant. In light of previous research that has especially supported the role of the latter two aspects, we focus on them for the present study.

Concerning the role of cognitive ability in interviews, it is expected that interviewees who have higher cognitive ability should be able to improve their interview scores more compared to those with lower cognitive ability because faking places high cognitive demands on the interviewee (Van Iddekinge et al., 2005). First of all, interviewees have to adapt their faked answers to anything the interviewer might already know about the interviewee or could possibly find out about the interviewee (DePaulo, Stone, & Lassiter, 1985). Second, the answers need to be specifically targeted to the interviewer and the respective questions (Cataldi, 1996). Third, interviewees only have a few seconds to prepare or adjust their answers, which means that they have to think of possible answers and to evaluate different options quickly. And fourth, interviewees need to monitor their nonverbal and verbal behavior constantly so not to give any hints of their faking to the interviewer (Van Iddekinge et al., 2005). In line with this, previous interview research found a positive correlation between cognitive ability and interview performance (Berry, Sackett, & Landers, 2007; Huffcutt, Roth, & McDaniel, 1996). Additional indirect support comes from research on faking in other selection tools. Levashina, Morgeson, and Campion (2009), for example, found that applicants with higher cognitive ability were more successful than applicants with lower cognitive ability when they faked a biodata measure – even though they were less likely to fake at all.

The other relevant factor that we want to consider is the ability to identify the targeted constructs being measured, which has been established in the literature as ATIC (= ability to identify criteria, Kleinmann et al., 2011). Specifically, ATIC is defined as a person's ability to correctly identify the performance criteria when participating in an evaluative situation. In the context of a selection interview, these criteria are the performance dimensions that are targeted by the different interview questions (Kleinmann et al., 2011).

Interviewees with higher ATIC scores should be better at identifying the performance dimensions measured in the interview. As a consequence, they should be better able to adapt their answers and show more dimension-relevant behavior. This in turn should lead to higher interview scores. In line with this suggestion, previous research found a positive correlation between ATIC and interview performance (e.g., Griffin, 2014; Melchers, Bösser, Hartstein, & Kleinmann, 2012; Oostrom, Melchers, Ingold, & Kleinmann, 2016) and related research on assessment centers also found similar correlations (e.g., Jansen et al., 2013; König, Melchers, Kleinmann, Richter, & Klehe, 2007; Preckel & Schüpbach, 2005).

Based on current theories of faking and the available empirical evidence, we predict that interviewees' abilities should be related to the degree to which they are able to fake during an interview when they are asked to present themselves in the best possible light compared to an honest condition. Thus, we posit the following hypotheses:

Hypothesis 2: Interviewees' cognitive ability will have a positive correlation with the degree to which they can deliberately improve their interview performance in comparison to a condition in which they answer honestly.

Hypothesis 3: Interviewees' ability to identify the criteria (ATIC) will have a positive correlation with the degree to which they can deliberately improve their interview performance in comparison to a condition in which they answer honestly.

Faking in Interviews and Criterion-Related Validity

There is widespread concern that faking in interviews has a negative effect on the quality of selection decisions. Therefore, if interviewees can fake in interviews, it is important to examine the possible effects of this faking behavior. Unfortunately, there is only limited evidence that addresses the effects of faking on the criterion-related validity of selection procedures and most of the available research has been conducted on personality tests. In addition, most research on criterion-related validity has dealt with predicting task performance. However, in recent years, there have been calls to expand the criterion domain to include other performance aspects such as contextual performance (Van Iddekinge & Ployhart, 2008), which is also an important aspect for the functioning of organizations. Therefore, it is essential to consider both aspects of performance because both are important for organizational success.

The available evidence with regard to the effects of faking on interview criterion-related validity is rather sparse. In the only study that addresses this issue to some degree, Ingold, Kleinmann, König, and Melchers (2015) found that self-reported faking[1] was positively (but not significantly) related to job performance (r = .15). Observer ratings of the same faking items were not significantly related to job performance. But whether, and if so, to what degree, interview scores in Ingold et al.'s study were higher in comparison to a situation in which interviewees would have provided honest answers was not examined.

[1]It should be noted that Ingold, Kleinmann, König, and Melchers (2015) called their questionnaire a measure of IM. However, given that they used a subset of items from Levashina and Campion's (2007) Interview Faking Behavior Scale (i.e., of deceptive IM), we consider this assessment a measure of self-reported faking.

Results from studies that investigated effects of faking on criterion-related validity in other selection procedures are somewhat mixed. On the one hand, several older, large-scale studies on personality measures did not find negative effects for criterion-related validity. That is, Ones, Viswesvaran and Schmidt (1993) as well as Hough et al. (1990) did not find evidence of detrimental effects of faking on personality measures with regard to their criterion-related validity. In their meta-analysis on the criterion-related validity of integrity tests, Ones et al. (1993) compared, among other things, applicant samples to job incumbent samples (i.e., two kinds of samples that should differ considerably with regard to their motivation to make a good impression in the integrity tests). They found that applicants' scores were even more highly correlated with job performance than incumbents' scores, meaning that there was no detrimental effect of faking on criterion-related validity. However, the observed difference was relatively small. Similarly, Hough et al. (1990) also found no detrimental effects on criterion-related validity of personality tests in a faking good condition in comparison to a condition in which test takers were told to respond like an applicant.

On the other hand, most of the recent research has found decreases in criterion-related validity. This research consists primarily of simulation studies (Converse, Peterson, & Griffith, 2009; Komar, Brown, Komar, & Robie, 2008; Marcus, 2006; Schmitt & Oswald, 2006) and two field-studies (Donovan, Dwight, & Schneider, 2014; Peterson, Griffith, Isaacson, O'Connell, & Mangos, 2011) and indicate that faking could have negative effects on criterion-related validity.

Finally, there is also at least one recent study by Ingold, Kleinmann, König, Melchers, and Van Iddekinge (2015) that surprisingly found a positive effect of faking on criterion-related validity. In this study, personality scores in an applicant instruction condition had higher criterion-related validity than personality scores from an honest condition.

How can these diverging findings with regard to criterion-related validity be explained? Unfortunately, even though a considerable number of faking models exist (e.g.,

Levashina & Campion, 2006; Marcus, 2009; McFarland & Ryan, 2000, 2006; Roulin et al., 2015; Snell et al., 1999) only three of them mention the effects of faking on validity (Marcus, 2009; McFarland & Ryan, 2000; Salgado, 2016). McFarland and Ryan (2000) note that faking could have an influence on validity – but they do not specify the nature of these effects. In contrast, the models by Salgado (2016) and by Marcus (2009) make specific predictions about criterion-related validity.

According to Salgado's (2016) model, faking increases the mean but decreases the variance in applicants' scores on a selection procedure so that faking leads to a homogenization of scores. Because of this homogenization and the reduction of systematic variance, Salgado's model predicts decreases in the reliability and the criterion-related validity of a selection procedure as a consequence of faking. To evaluate his model, Salgado conducted a meta-analysis on personality test data. He found support for the predictions that faking would increase mean scores, decrease standard deviations, and decrease internal consistencies in comparison to a control condition. However, Salgado did not evaluate the assumed effects on criterion-related validity.

In contrast to Salgado's (2016) model, Marcus's (2009) process theory of self-presentation describes conditions under which positive effects of faking on criterion-related validity are possible. Specifically, his model assumes that two key self-presentation resources are important during the selection process: (1) applicants' motivation to fake and (2) their skill or ability to fake. Dispositional (e.g., personality) and situational factors (e.g., attractiveness of the position) determine whether an individual is motivated to fake during a selection procedure but the applicant's ability determines whether this faking behavior is successful (e.g., that it results in better test scores or more favorable interview ratings). Thus, to be successful in an interview, interviewees need certain abilities in order to adapt their faking behavior to the requirements of the interview. Among other variables, Marcus

mentions cognitive ability and ATIC as possible influencing factors on individuals' ability to fake.

In addition to the relevance of cognitive ability and ATIC for faking in interviews, these abilities might also be important for the positive effects of faking on criterion-related validity because they are also positively related to job performance. Specifically, cognitive ability (cf. Berry et al., 2007, or Huffcutt et al., 1996, for meta-analytic evidence) and ATIC (e.g., Griffin, 2014; Melchers et al., 2009) are positively correlated with interview performance but also with job performance (see for example, Ingold et al., 2015, for evidence concerning ATIC; and Schmidt & Hunter, 1998, for evidence concerning cognitive ability). Thus, when interviewees are motivated to fake in an interview, their interview performance should be saturated to a higher degree with ability-related factors that contribute to criterion-related validity than when interviewees are not motivated to fake. In contrast, we propose that decreases in criterion-related validity have not been found for faked personality scores because faking in such tests is less cognitively demanding and much less saturated with ability-related factors. As a consequence of this line of reasoning, we posit the following hypothesis:

> Hypothesis 4: Interview performance in a faking condition will show higher criterion-related validity for predicting task performance than interview performance in an honest condition.

As an extension to Marcus's (2009) theory, it is not only important to examine task performance but to also consider the effect that faking has on the prediction of contextual performance. Contextual performance (also called citizenship behavior) is defined as "extra-role, discretionary behavior that helps other organization members perform their jobs or that shows support for and conscientiousness toward the organization" (Borman & Motowidlo, 1993, p. 75). Because contextual performance is important for organizational success (Podsakoff, Whiting, Podsakoff, & Blume, 2009), it is also relevant to consider the extent to

which interviews can predict contextual performance (Latham & Skarlicki, 1995) – and also whether faking influences this prediction.

Theoretical arguments based on Mischel's (1977) distinction between strong versus weak situations suggest that citizenship behaviors may be predicted differently than task performance in faking versus honest conditions. Specifically, according to Mischel, strong situations limit behavioral options and provide clear signals on what to do. In contrast, weak situations do not give individuals explicit clues how to behave. Thus, individuals "project" their personality into the situation and behave as they typically would. In this way, personality should be a more influential predictor in weak situations than in strong situations because the former restrict the degree of behavioral variability across individuals, which is comparable to the homogenization that is assumed in Salgado's (2016) model.

The distinction between weak versus strong situations and its effects on behavioral variability is important in the present context because a faking or job application condition represents a strong situation where interviewees are expected to provide answers that show themselves in a favorable light. In contrast, an honest condition represents a weak situation in which interviewees' answers should reflect personality differences to a stronger degree than in a faking or job application condition. As a consequence, honest interviews should allow a better prediction of criteria that are influenced more strongly by personality and motivational differences. Furthermore, it is also known from previous research that personality predicts citizenship behaviors (Hattrup, O'Connell, & Wingate, 1998). According to Borman and Motowidlo (1997), this is because contextual behaviors such as altruism, team work, or volunteering reflect individual differences in motivation and interpersonal style rather than differences in skills or abilities. We therefore posit the following final hypotheses:

Hypothesis 5: Interview performance in an honest condition will show higher criterion-related validity for predicting contextual performance than interview performance in a faking condition.

Hypothesis 6: Interviewees' personality is a better predictor of interview performance in an honest condition than in a faking condition.

Method

Participants

One hundred and eleven undergraduate psychology students from a German university voluntarily participated in our study.[1] Their ages ranged from 18 to 39 years with a mean of 22.66 (*SD* = 3.23). The overall sample consisted of 93 females and 18 males. Most of them had interview experience (84.7 %) with having participated in *M* = 3.58 previous interviews. Participants were recruited from a departmental subject pool and could participate in this study to partially fulfill a course requirement. Only students who were in their second semester or above were accepted to take part in the study, because we used their grade point average (GPA) as one of our criterion variables.

Procedure

After signing up for the study, participants were sent a link to an online questionnaire containing questions on demographic variables, personality, GPA, and university citizenship behavior (UCB). Participants were then invited for a structured interview that was designed to predict academic performance (see below for more information on the interview and the different variables). Thus, we collected data to determine the concurrent validity of the interview. An important aspect in this regard is the structure of study programs in the German higher education system. Specifically, in contrast to the North American system (but in line with the system from the majority of the European countries) the general structure and the kind of courses of the 3-year Bachelor's and the 2-year Master's program in Psychology are rather similar. They are both based on courses that have a clear focus on Psychology from the very first year. Students do not have to take general courses outside of Psychology (with the exception of only one or two courses), and the kind of courses (lectures, seminars, lab classes,

etc.) and necessary coursework to earn credits (exams, presentations, lab reports, essays, empirical research theses, etc.) are comparable between a Bachelor's and a Master's program. In line with this, a study with German Psychology students found a rather high (uncorrected) correlation of $r = .75$ between students' undergraduate GPA and their final GPA from the Master's program (Wedler, Troche, & Rammsayer, 2008). Thus, the main difference is that the Master's level courses are more specialized and advanced in content so that a BSc degree in Psychology is required to apply for a Master's program but apart from this difference, similar skills are required by students.

With regard to our study, all participants completed the interview twice; once in an honest condition, in which they were instructed to answer as honestly as possible, and a second time in an applicant instruction condition, in which they were told to present themselves in the best possible way as if they were applying for a highly attractive Master's program. The order in which participants took part in the two interview conditions was counterbalanced across participants. There were at least 10 days between the two interviews to reduce the likelihood that participants recalled the answers given in the previous interview.

Before the start of their first interview, participants completed a cognitive ability test and after their second interview, they completed a questionnaire to assess ATIC. Furthermore, after each interview, participants responded to manipulation check items.

The interviews were usually conducted by an interviewer and a second person who served as an additional rater. Participants never saw an interviewer or a rater twice to ensure that interview ratings in the second interview were not biased by ratings from the first interview and also to avoid any tendency for interviewees to be consistent in their answers. Furthermore, all interviews were videotaped.

Applicant Instruction Condition. Participants were instructed to be dressed as if they were invited for a selection interview for an attractive graduate program. The interview was

conducted in a conference room to make the setting as realistic as possible. The interviewer was a faculty member and participants were addressed formally by their last names.

For the applicant instruction interview, we patterned our approach after the procedure used by Van Iddekinge et al. (2005). We did not ask the participants to "fake good" but to put their best foot forward in order to appear as an ideal candidate. We told them they should imagine applying for an attractive Master's program in Psychology. This is a realistic scenario in Germany as most students apply for a Master's program because even though an undergraduate degree in Psychology is based on a three-year full-time Psychology program, a Master's degree is required by most employers and to get access to advanced postgraduate or PhD programs as well as for postgraduate training to become a psychotherapist. Furthermore, for a substantial number of the Master's programs, interviews are used as part of the admission process ("Liste Master Psychologie", 2016). Moreover, according to a recent survey of the German Psychological Association the number of applications for these programs outnumbers the number of places at the different universities by a factor of 10 so that interviewees who took part in our study should be motivated to perform and get an idea of how well they can do (Abele-Brehm, 2017). In line with this evidence, participants seemed to be nervous and appeared to view the interview as training for their future application process. To further increase their motivation, participants were told that the top 5% of the interviewees in this condition would each receive 50 Euros (approximately $60 US dollars) as an incentive to put their best foot forward.

Honest condition. For this condition, participants were instructed to be dressed as they would normally be when they attended class at the university. We conducted this interview in a departmental kitchen area with comfortable chairs and ensured that no one else would be present during the interview to make the setting as relaxed as possible. The interviewer for the honest condition was an industrial/organizational Master's student. The rationale for this was that the choice of interviewer should reduce participants' motivation to

selectively present him- or herself in a more positive way and especially to reduce any fears that the interviewer might have any impact on participants' future course evaluations. Participants were addressed in an informal manner by their first name only. For the honest instructions, participants were told that they were part of a study in which we wanted to validate an interview examining their study behavior and that they should therefore answer as truthfully as possible. We assured them that their answers would remain confidential and would be used for research purposes only.

Measures

Structured interview. The interview consisted of 20 questions (two questions on career and program choice, seven past behavioral questions, and eleven future-oriented questions). The interview was developed from a large set of critical incidents that were collected prior to this study. Critical incidents measured different behaviors necessary for academic success such as teamwork, problem solving, conscientiousness, persistence, and planning and organizing. In a pilot study with 66 student participants, we found that the interview predicted university citizenship behavior, $r = .49$, $p < .01$, as well as peer-ratings of students' academic performance, $r = .29$, $p < .05$.

As noted above, the interviewer was a faculty member in the applicant instruction condition and an industrial/organizational Master's student in the honest condition. The additional raters in each of the conditions were all industrial/organizational Master's students. All interviewers and raters received several hours of frame-of-reference training (Melchers, Lienhardt, von Aarburg, & Kleinmann, 2011; Roch, Woehr, Mishra, & Kieszczynska, 2012). Furthermore, the Master's student was also experienced as an interviewer because she had already conducted more than 50 interviews either as a student member on a departmental search committee for the selection of a new professor or as the previous head of the student union where she had to interview applicants for positions as (paid) acting managers for various divisions of the student union.

For each question, the interviewer and the second rater took notes of interviewees' answers and independently evaluated the answer to each question directly before proceeding to the next question. Evaluations were made on 5-point scales ranging from 1 = *poor performance* to 5 = *excellent performance* for which descriptive anchors were provided for poor, medium, and excellent answers. When individual ratings for a question differed two or more points on the 5-point scale, the interviewer and the second rater discussed their observations and ratings after the completion of the interview. Although interviewers did not have to agree with each other, most differences could be resolved after a short discussion so that the ratings for individual questions did not usually differ by more than one point.

To determine interrater reliability, we calculated intraclass correlations (ICC 1.1) between the overall interview ratings, which were based on the respective mean across all 20 interview questions. The mean interrater reliability was .99 for the honest condition and .96 for the applicant instruction condition after the discussion of the interviewer and the second rater. Before this discussion, the mean interrater reliability was .90 for the honest condition and .78 for the applicant instruction condition.

Faking. To operationalize interview faking, we used the regression-adjusted difference score (RADS) between the applicant instruction condition and the honest condition. Many previous studies on faking in personality tests have used the raw difference score between the faking and the honest condition (e.g., Viswesvaran & Ones, 1999). Because of the psychometric problems that occur with such raw difference scores (Burns & Christiansen, 2011), we used RADS proposed by Burns and Christiansen (2011) as the main indicator of faking. The RADS avoids these psychometric problems, namely that a difference score will always be negatively related with scores from the honest condition in addition to its positive relationship with the scores from the faking condition. Instead of this, regression-adjusted difference scores are "derived by regressing scores from the faking condition onto scores from the honest condition and computing the residual," so that "the regression-adjusted

difference score is interpreted as that part of the scores from the faking condition that cannot be explained by honest scores" as was explained by Burns and Christiansen (2011, p. 361/362).

Personality. We measured the Big Five (with the exception of Openness) with the short version of the Big Five Inventory (BFI-K, Rammstedt & John, 2005) and honesty-humility with the corresponding scale from the HEXACO questionnaire (Ashton & Lee, 2009). The honesty-humility scale contains the four subfacets sincerity ("I wouldn't use flattery to get a raise or promotion at work, even if I thought I would succeed"), fairness ("I would never accept a bribe, even if it were very large"), greed avoidance ("Having a lot of money is not especially important to me"), and modesty ("I am an ordinary person who is no better than others"). All personality items had to be answered on a 5-point rating scale from 1 (= *I do not agree at all*) to 5 (= *I fully agree*). Because of a mistake, Big Five data were not collected for the first 24 participants.

GPA. GPA was used as an indicator of participants' task performance. Participants were asked to report their current GPA. According to the German grading system, GPA can vary between 1.0 and 4.0, with lower scores representing better academic performance. GPA, however, was recoded for the present study to make interpretation of the results easier so that higher values represent better academic performance.

Peer Ratings. A peer rating of participants' academic performance was used as an additional indicator of their task performance. When signing up for our study, participants were asked to name two fellow students who would be willing to answer an online questionnaire about the participant's study performance. We contacted these peers via email and sent them a link to an online questionnaire. Peers were informed that these ratings were for research purposes and would not be revealed to the participant. We measured task performance with three items, which we developed for the pilot study mentioned above. Those three items were: "How would you rate the performance at this university of [name of

participant]?" "How would you rate the performance of [name of participant] in written exams?" and "How would you rate the performance of [name of participant] in oral exams?". We used the relative percentile method suggested by Goffin, Jelley, Powell, and Johnston (2009) so that we instructed peers to compare the participant to all students of the respective cohort with help of the percentile rank. Cronbach's alpha for this scale was .78. To determine interrater reliability, we calculated intraclass correlations (ICC 1.1) between the peer ratings, which were based on the respective mean across the three questions. The mean interrater reliability (i.e., the reliability of a single rater) was .32. This is comparable to the average interrater reliability of peer ratings for employees' job performance (Conway & Huffcutt, 1997; Viswesvaran, Ones, & Schmidt, 1996).

University citizenship behavior. We measured UCB as an indicator of contextual performance. To do so, we used items from the UCB scale by Zettler (2011). These 23 items (e.g., "I encourage fellow students when they feel low") were answered on a 5-point rating scale ranging from 1 = *I do not agree at all* to 5 = *I fully agree*. Cronbach's alpha for this scale was .76.

Cognitive ability. Cognitive ability was measured with the Wonderlic Cognitive Ability Test (Wonderlic, 2002). This 12-minute, 50-item test assesses vocabulary, arithmetic reasoning, and spatial relations and is a commonly used measure of cognitive ability.

ATIC. To measure ATIC, we closely followed procedures used in previous studies (Ingold, Kleinmann, König, Melchers, et al., 2015; Jansen et al., 2013). Specifically, after their second interview, participants completed a questionnaire that contained all 20 interview questions. Space was provided underneath each question and participants were instructed to write the dimension that they believed was assessed with each interview question. In addition, participants were instructed to provide behavioral examples for each of these dimensions (e.g., they wrote down teamwork as the targeted dimension and provided an example that someone helped a fellow student). Participants could write down as many dimensions and

behavioral examples as they wanted. To ensure that participants understood the task, we gave them a neutral example on creativity.

After completion of the interviews, trained raters examined the dimensions and behavioral examples that had been written by the interviewees and rated the fit with the targeted dimensions on a scale ranging from 0 = *no fit* to 3 = *fits completely*. Raters discussed disagreements that exceeded one point. The interrater reliability of the averaged ATIC ratings from the two raters (ICC 1.1) across all 20 interview questions was .87 before discussion and .97 after discussion. The overall ATIC score was calculated based on the average score of the two raters across the 20 questions after discussion of their ratings.

Manipulation Check. We examined whether the two conditions indeed led to differences in faking behavior. To do so, we used a shortened version of the Levashina and Campion (2007) Interview Faking Behavior Scale that was also used by Ingold, Kleinmann, König, and Melchers (2015). The 11 items from this scale (e.g., "I have covered something up in order to be able to give better interview responses", a complete list of the items is provided in the Appendix) were answered on a 5-point rating scale ranging from 1 = *I do not agree at all* to 5 = *I fully agree*. Cronbach's alpha for this scale was .88.

Results

Table 1 displays correlations, means, and standard deviations for all study variables. To ensure our manipulations worked as intended, we analyzed data from the manipulation check scale. We found that participants in the applicant instruction condition had a significantly higher mean on the Interview Faking Behavior scale ($M = 2.23$, $SD = 0.70$) compared to the honest condition ($M = 1.32$, $SD = 0.26$), $t(109) = 13.93$ $p < .01$, $d = 1.33$ (d was calculated according to the formula by Cohen, 1988, for within-subjects data). Thus, the manipulation was successful.

Hypothesis 1 stated that interviewees would be rated more favorably in the applicant instruction condition than in the honest condition. In line with this, interviewees received higher mean interview ratings across all 20 questions in the applicant instruction condition (M = 3.62, SD = 0.24) compared to the honest condition (M = 3.20, SD = 0.32), $t(110) = 12.86$ $p < .01$, $d = 1.49$. Thus, our data supported Hypothesis 1 and showed a clear faking effect.

Hypothesis 2 stated that interviewees' cognitive ability would be positively correlated with the degree to which individuals could improve their performance in the applicant instruction versus honest condition. As can be seen in Table 1, the correlation between cognitive ability and the regression-adjusted mean difference was $r = .18$, $p < .05$ (one-tailed). Given the directional nature of our hypothesis, we consider this as support for Hypothesis 2.

Next, Hypothesis 3 stated that interviewees' ability to identify criteria would be positively correlated with the degree to which individuals could improve their interview performance in the applicant instruction versus the honest condition. Given the directional nature of our hypothesis, we again used a one-tailed test. As can be seen in Table 1, ATIC was positively correlated with the regression-adjusted mean difference, $r = .16$, $p < .05$ (one-tailed). Thus, our data supported Hypothesis 3.

With regard to examining the effect on criterion-related validity of the interview, Hypothesis 4 stated that interview performance in the applicant instruction condition would better predict task performance than interview performance in the honest condition. In line with this, the interview scores were a significant predictor of GPA in the applicant instruction condition, $r = .34$, $p < .01$, but not in the honest condition, $r = .15$, $p = .13$. Steiger's (1980) test showed that the difference between these correlations was significant, $z = 1.77$, $p < .05$ (one-tailed), $N = 111$, which provides support for Hypothesis 4.

Furthermore, we also examined peer ratings of participants' academic performance. The interview scores were a significant predictor of the peer ratings in the applicant instruction condition, $r = .24$, $p < .05$, but not in the honest condition, $r = .01$, $p = .96$.

Steiger's (1980) test showed that the difference between these correlations was significant, z = 1.91, $p < .05$ (one-tailed), $N = 95$, which provides additional support for Hypothesis 4.

With regard to Hypothesis 4, we reasoned that faking in our interview required abilities that were also relevant for task performance. In line with this thinking, the RADS showed to be a significant predictor of task performance. Specifically, the RADS had a significant correlation of $r = .31$, $p < .01$, with GPA and of $r = 25$, $p < .05$, with peer ratings of academic performance.

Next, Hypothesis 5 stated that interview performance in the honest condition would better predict contextual performance than interview performance in the applicant instruction condition. Consistent with this hypothesis, the interview scores were not significantly correlated with UCB in the applicant instruction condition, $r = .13$, $p = .18$, but were a significant predictor of UCB in the honest condition, $r = .51$, $p < .01$. Again, Steiger's (1980) test showed that the difference between these correlations was significant, z = -3.77, $p < .01$ (one-tailed), $N = 111$, which provides support for Hypothesis 5.

Finally, Hypothesis 6 predicted that interviewees' personality would be a better predictor of their interview performance in the honest condition than in the applicant instruction condition. Inspection of Table 1 suggests that the correlations between personality and interview performance are higher in the honest condition than in the applicant condition. However, to test Hypothesis 6 formally, we conducted multiple regression analyses for which we used interview performance as the dependent variable and the different personality variables as predictors. In the first set of these analyses, we considered all personality variables (i.e., extraversion, conscientiousness, agreeableness, and emotional stability as well as the four honesty-humility facets) together. Because of the missing Big Five data for the first 24 participants, this reduced our sample size to 87. In line with our hypothesis, the different personality variables only accounted for a limited (and non-significant) amount of variance in participants' interview performance in the applicant instruction condition, $R = .21$,

R^2 = .05, p = .87. In contrast, the personality variables accounted for a considerable amount of variance in interview performance in the honest condition, R = .45, R^2 = .20, p < .05. To compare the multiple correlations between personality and interview performance in the two conditions, we again used Steiger's (1980) test, which revealed a significant difference, z = -1.75, p < .05 (one-tailed).

Additional Analyses

Our attempts to create conditions that encouraged interviewees to present themselves in the best possible way in the applicant instruction condition but as honestly as possible in the honest condition led to the use of two different interviewers and thus the corresponding confounding of the two interviewers with the two interview conditions. Even though both were experienced interviewers and had undergone extensive frame-of-reference training, they may have shown differences in their interviewing behavior and these differences may be responsible for the obtained differences in interviewees' performance.

To rule out this alternative explanation that mean differences in the two conditions (applicant versus honest condition) are solely attributable to differences in interviewer behavior, we analyzed video recordings of the interviews. Two independent observers rated whether interviewers adhered to the wording of the interview questions and whether interviewers gave nonverbal feedback during the interview in the form of nodding or signs of encouragement (interviewers did not give verbal feedback concerning the quality of interviewees' answers). Paired sample t-tests showed that the interviewer in the honest condition in fact adhered to the wording of the interview questions to a greater extent than did the interviewer in the applicant instruction condition, t(100) = 18.77, p < .01, and that the interviewer in the applicant instruction condition gave relatively more nonverbal feedback than the interviewer in the honest condition, t(100) = 17.25, p < .01.

To test further whether these differences between the interviewers might account for the mean difference in interview scores (Hypothesis 1), we tested a set of hierarchical linear

models (HLM 7.01, Raudenbush, Bryk, Cheong, Congdon, & Toit, 2011). For these HLM analyses, we specified and compared two nested hierarchical models. In Model 1, we predicted interview scores from adherence to wording of questions and from nonverbal feedback. Both were significant predictors of the interview score (Estimate = -0.73, *SE* = 0.13, $t(99) = -5.31$, $p < .01$, and Estimate = 0.24, *SE* = 0.06, $t(99) = 3.76$, $p < .01$, respectively). Thus, across conditions (honest and applicant instruction), less adherence to wording and more feedback were beneficial for the interview score. In the next step (Model 2), we additionally entered the predictor condition (coded 0 = *applicant condition*, 1 = *honest condition*). Condition was a significant predictor of the interview score (Estimate = -0.35, *SE* = 0.07, $t(98) = -4.70$, $p < .01$) and the model including the predictor condition did fit the data significantly better than the previous model (Δ -2 × log likelihood = 20.75, $df = 1$, $p < .01$). Thus, participants achieved better interview scores in the applicant instruction condition than in the honest condition even after controlling for interviewer behavior, yielding support for Hypothesis 1.

Next, we also tested whether differences in interviewers' behaviors affected the results concerning criterion-related validity (i.e., Hypotheses 4 and 5). To do so, we predicted interview scores in both conditions from adherence to wording and nonverbal feedback, and used the residuals of this regression to calculate correlations between GPA, peer ratings of academic performance, UCB, interview $score_{residual}$ (honest condition), and interview $score_{residual}$ (applicant instruction condition). The interview $score_{residual}$ was a significant predictor of peer ratings of academic performance in the applicant instruction condition, $r = .23$, $p < .05$, but not in the honest condition, $r = -.03$, $p = .77$. Steiger's (1980) test showed that the difference between these correlations was significant, $z = 1.97$, $p < .05$ (one-tailed), $N = 87$. Furthermore, the interview $score_{residual}$ was a significant predictor of UCB in the honest condition, $r = .50$, $p < .01$, but not in the applicant instruction condition, $r = .14$, $p = .17$. Steiger's test showed that the difference between these correlations was significant, $z = -3.30$,

$p < .01$ (one-tailed), $N = 101$. Finally, the interview $score_{residual}$ was a significant predictor of GPA in the applicant instruction condition, $r = .33$, $p < .05$, but not in the honest condition, $r = .14$, $p = .14$. However, Steiger's test for the difference between these correlations just failed to reach significance, $z = 1.57$, $p = .058$ (one-tailed), $N = 101$. Nevertheless, these results suggest that the overall pattern of results concerning Hypotheses 4 and 5 remained largely the same even after taking differences in interviewers' behaviors into account.

Discussion

We aimed to answer the three following questions with this study: First, can interviewees actually improve their interview scores by faking? Second, what role do interviewee ability factors play with regard to faking in interviews? And third, what effect does faking in interviews have on our ability to successfully predict interviewees' job or academic performance?

Concerning these questions, the present study found that interviewees improved their interview performance when they faked compared to when they responded honestly. Furthermore, the degree of the improvement of their interview performance was related to ability factors, namely, cognitive ability and ATIC. Finally, we found that task performance was better predicted from interview performance in the applicant instruction condition, whereas contextual performance was better predicted from interview performance in the honest condition.

Our finding that interviewees can improve their scores contributes to the question of whether people can fake in interviews. Thus, our results are in contrast to Allen et al.'s (2004) findings but confirm evidence by Van Iddekinge et al. (2005) that interviewees are able to improve their scores if they are motivated to do so. Furthermore, the effect size that we found was substantially higher than in Van Iddekinge et al.'s study ($d = 1.26$ compared to $d = .54$ in the study by Van Iddekinge et al.). We suggest that the larger effect size in the present study

can be attributed to the experimental conditions that we created and that the smaller effect in Van Iddekinge et al.'s study might be an underestimate of the effects of interviewees' faking behavior. Moreover, the null effect in Allen et al.'s study might be due to aspects of the experimental procedure they used that limited the degree to which interviewees put their best foot forward in the faking condition and/or the extent to which they provided truthful answers in the honest condition. As noted above, we tried to assure that interviewees in the present study were, on the one hand, motivated to put their best foot forward in the applicant instruction condition but, on the other hand, also willing to reveal the truth about themselves in the honest condition.

Moreover, our results stress the important role of interviewees' individual differences (specifically, their cognitive ability and their ability to identify criteria, i.e., ATIC) when it comes to performing well in selection interviews. Our findings that interviewees' cognitive ability and ATIC both were positively correlated with the regression-adjusted mean difference between interview performance in the applicant versus honest condition is consistent with previous research on faking in biodata measures (Levashina et al., 2009) and provides support for the propositions laid out in several theories of faking (e.g., Ellingson & McFarland, 2011; Goffin & Boyd, 2009; Griffith, Lee, Peterson, & Zickar, 2011; Marcus, 2009; McFarland & Ryan, 2000, 2006; Mueller-Hanson, Heggestad, & Thornton, 2006; Roulin et al., 2016; Snell et al., 1999; Tett & Simonet, 2011). With regard to ATIC, our results expand previous findings of the relation between ATIC and interview performance (Griffin, 2014; Melchers et al., 2009) and demonstrate that ATIC influences the extent to which interviewees can improve their interview scores.

ATIC and cognitive ability are both part of the analytical skills in the self-presentation model by Marcus (2009). Understanding the role of abilities in faking can help explain the effects of faking on criterion-related validity. The self-presentation model by Marcus (2009) suggests a positive effect of faking on criterion-related validity. As described previously, this

effect is proposed to occur because the ability-related factors that are relevant for improving one's score in selection interviews through faking are also related to one's job performance (Jansen et al., 2013; Schmidt & Hunter, 2004; Schmidt & Hunter, 1998). Thereby, the predictions from his model are different from and more encompassing than the predicted negative effect of faking on the basis of Salgado's (2016) model. Salgado's model only assumes a homogenization effect that reduces criterion-relevant variance in scores from a selection procedure but does not assume a positive effect of other, criterion-relevant factors such as ability variables that are captured in those scores to a stronger degree when applicants attempt to present themselves in a positive light.

In line with Marcus's (2009) model, our findings demonstrate that interview performance in the applicant instruction condition better predicted task performance than interview performance in the honest condition. Thus, faking contributed to predicting task performance in our study instead of impairing criterion-related validity. Therefore, these findings suggest that we should not necessarily regard faking as always resulting in a negative effect for interview validity. However, we would assume that positive effects on criterion-related validity depend on situations where abilities are of sufficient importance for faking.

In contrast to the positive effect of faking for the prediction of academic performance, we also found that interview performance in the honest condition was a better predictor of contextual performance than in the applicant instruction condition. As described above, this can be explained by referring to the distinction between strong versus weak situations (Mischel, 1977). Specifically, the interview performance in the honest condition should have been a better predictor of contextual performance, because whereas the applicant instruction condition represents a strong situation in which interviewees try to perform well, the honest condition is a weak situation in which personality influences decisions about behavior to a stronger degree (Mischel, 1977). This is in line with the homogenization effect assumed by Salgado's (2016) model and also with the results related to Hypothesis 6 that showed that

personality was a better predictor of interview performance in the honest condition than in the applicant condition. In this way, honest interviews should allow for a better prediction of criteria that are influenced more strongly by personality and motivational differences (Hattrup et al., 1998) as we found.

By considering contextual performance in addition to task performance, the present study also responded to Van Iddekinge's and Ployhart's (2008) call for an expansion of the performance domain for studies that deal with criterion-related validity. By doing so, we gained more insight into the effects of faking on the criterion-related validity of selection interviews. Based on the present results, we suggest that the claim that "faking impairs criterion-related validity" is too simplified and that it is also important to consider the kind of criterion predicted.

Limitations and Lines for Future Research

One possible limitation of this study is that it was conducted in a simulated setting. However, we tried to make sure that the applicant condition was as realistic as possible and we chose official rooms, a faculty member as the main interviewer, and instructions for participants to dress in professional interview attire. Moreover, interviewees seemed to be nervous and requested feedback for their performance in this condition. In contrast, for the honest condition, it was important that participants understood it was being conducted in a simulated setting to make sure that they felt comfortable truly answering honestly without fear of any negative consequences.

Another limitation that is related to the simulated setting is that we conducted our study with student participants. However, as was noted by other researchers from the personnel selection domain, "the admissions problems of academic administrators are very similar to those of private and public employers" (p. 1480, Schmitt et al., 2009). In line with this, having to face an interview similar to the present one is a realistic possibility as explained in the Methods section. Moreover, we adapted the applicant situation to students'

reality. For example, our interview was conducted by a faculty member and was designed to predict academic success. Furthermore, students took it as preparation for their application to a Master's program. Nevertheless, future research should be conducted in organizational settings in order to replicate our results and generalize our findings to actual hiring settings. To date, we are not aware of any study on faking in employment or admission interviews and its influences on criterion-related validity that have used such field data. In order to conduct a criterion-related validity study in the field, researchers could rely on designs that have been used in research of faking in personality tests; comparing interview data from incumbents (e.g., from a pilot test run) with data from actual applicants.

With regard to our OCB measure, another possible limitation concerns the use of self-ratings instead of other-ratings. Specifically, there is a chance that self-ratings of OCB may be inflated in comparison to other-ratings (cf. Allen, Barnard, Rush, & Russell, 2000), which might also influence the relationships between OCB measures and other variables. However, recent meta-analytic work revealed that the degree to which self-ratings of OCB are higher than other ratings is relatively moderate (Carpenter, Berry, & Houston, 2014) and that self-rated and other-rated OCBs also have similar patterns of relationships with other variables (Carpenter et al., 2014; Gonzalez-Mulé, Mount, & Oh, 2014). Thus, we assume that the qualitative pattern of results concerning the criterion-related validity of OCB should not be influenced by the source of the OCB ratings. Nevertheless, future research is necessary to corroborate this argument.

Another limitation that arises from our measures is that we concentrated on cognitive ability and ATIC (a social effectiveness construct that is relatively specific to the selection situation, cf. Melchers et al., 2009) but did not take other ability-related constructs into account such as social skills or communication skills. Future research should also consider such variables to get a fuller picture of the abilities that are relevant for faking to be effective.

Furthermore, with regard to Van Iddekinge and Ployhart's (2008) call for an expansion of the criterion domain, future research should look at additional criteria in the context of faking in interviews such as counterproductive work behavior or additional forms of productive work behavior (Frese, Fay, Hilburger, Leng, & Tag, 1997) or adaptive behavior (Pulakos, Arad, Donovan, & Plamondon, 2000)). Moreover, such an expansion of the criterion domain would also be valuable with regard to evaluations of the effects of faking for other tests and selection procedures (e.g., personality tests or SJTs).

It would also be important for future research to examine the generalizability of our results to other kinds of interviews. To date, previous research on the relationship between IM and interview performance revealed larger effects of IM in unstructured as compared to structured interviews (Barrick et al., 2009), but this research did not differentiate honest from deceptive IM (i.e., faking). Therefore, future research should investigate whether there are also larger effects of faking in less structured interviews and also how faking affects the criterion-related validity of such interviews. Such a suggestion would also be supported by research from Blackman and colleagues (Blackman, 2002; Townsend, Bacigalupi, & Blackman, 2007). They found that applicants' personality was better predicted by unstructured interviews compared to structured interviews. Their explanation for this finding was that unstructured interviews represent weak situations in which interviewees' personality has a more marked effect in comparison to structured interviews that represent stronger situations. However, none of the studies from Blackman and colleagues used an applicant instruction interview condition. Nevertheless, on the basis of the results from Blackman and colleagues as well as from Barrick et al. (2009), it might well be the case that differences between honest versus applicant conditions with regard to interview performance as well as with criterion-related validity might be more pronounced for unstructured than for structured interviews like the present one.

A final aspect that we want to mention here concerns our use of an applicant instruction instead of a fake-good condition to manipulate faking. It might be that this kind of instruction did not lead to actual faking behavior (i.e., to deceptive IM) by all of our interviewees, but instead to more honest IM in some cases. In line with this, data from our manipulation check suggest a significant increase in faking behavior compared to our honest condition but on average interviewees did not reach the maximum level of deceptive behavior that can be captured by the scale that we used. Accordingly, our study might actually underestimate the impact of faking behavior on interview scores and on criterion-related validity. With regard to the latter, it might well be the case that faking impairs the prediction of relevant criteria once interviewees reach a certain level of clearly deceptive behavior. However, we consider our applicant instruction condition as a more realistic condition than a fake-good instruction condition because research on faking in personality tests has revealed that actual applicants fake in a more reluctant way than participants in fake-good studies (see Birkeland et al., 2006; and Viswesvaran & Ones, 1999). This was also confirmed in a recent study by Krammer, Sommer, and Arendasy (2017) who found that individuals faked much more in an instructed fake-good condition then in a real selection situation in which they had completed the same personality scale earlier. However, similar to the personality domain, it might be possible that future research reveals evidence that the difference between a faking condition and an honest condition becomes larger when a fake-good instruction is used than when an applicant instruction is utilized. Furthermore, even though additional analyses from our study did not support curvilinear relationships between faking and task performance, indirect support for negative effects of extreme faking behavior on criterion-related validity comes from a study on faking in situational judgment tests by Peeters and Lievens (2005). In contrast to the present study, they used a rather explicit fake-good instruction in their faking condition and found that the situational judgment test lost its criterion-related validity with regard to predicting academic performance.

Practical Implications

Results of our study provide practical advice to organizations and to interviewers. We tried to examine the extent to which organizations should be concerned about applicant faking in selection interviews and thus, about hiring the wrong people. For task performance, our results seem to indicate that organizations may not necessarily need to be concerned. In fact, interview scores in the applicant instruction condition correlated higher with task performance than interview scores in the honest condition. However, when predicting contextual performance we found the opposite effect. Though the correlation between the interview in the applicant instruction condition and UCB was not negative, it was smaller than in the honest condition, suggesting that the interview in the applicant instruction condition still predicted UCB but not as well as in the honest condition.

Given that a large part of the research in faking in interviews has concentrated on detecting faking (e.g.,Reinhard et al., 2013; Roulin et al., 2015) and that the existing evidence so far suggests that interviewers have difficulties detecting faking, we can offset concerns that limiting faking is absolutely necessary in order to ensure the quality of the interview.

Conclusion

The aim of our study was to shed light on the relationship between faking in interviews and criterion-related validity. Our results revealed that it is possible to improve interview scores by faking. Our results also showed that faking can have positive effects on criterion-related validity in terms of predicting task performance but not with regard to predicting contextual performance. These findings, therefore, reveal that the question whether faking impairs criterion-related validity cannot be answered with a simple Yes or No but that we need to differentiate between the facets of performance that we want to predict.

References

Abele-Brehm, A. (2017). Zur Lage der Psychologie [The state of psychology]. *Psychologische Rundschau, 68*, 1-19. doi:10.1026/0033-3042/a000346

Allen, T. D., Barnard, S., Rush, M. C., & Russell, J. E. A. (2000). Ratings of organizational citizenship behavior: Does the source make a difference? *Human Resource Management Review, 10*, 97-114. doi:10.1016/S1053-4822(99)00041-8

Allen, T. D., Facteau, J. D., & Facteau, C. L. (2004). Structured interviewing for OCB: Construct validity, faking, and the effects of question type. *Human Performance, 17*, 1-24. doi:10.1207/S15327043HUP1701_1

Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment, 91*, 340-345. doi:10.1080/00223890902935878

Barrick, M. R., Shaffer, J. A., & DeGrassi, S. W. (2009). What you see may not be what you get: Relationships among self-presentation tactics and ratings of interview and job performance. *Journal of Applied Psychology, 94*, 1394-1411. doi:10.1037/a0016532

Berry, C. M., Sackett, P. R., & Landers, R. N. (2007). Revisting interview-cognitive ability relationships: Attending to specific range restriction mechanisms in meta-analysis. *Personnel Psychology, 60*, 837-874. doi:10.1111/j.1744-6570.2007.00093.x

Blackman, M. C. (2002). Personality judgment and the utility of the unstructured employment interview. *Basic and Applied Social Psychology, 24*, 241-250. doi:10.1207/S15324834BASP2403_6

Borman, W. C., & Motowidlo, S. J. (1997). Task performance and contextual performance: The meaning for personnel selection research. *Human Performance, 10*, 99-109. doi:10.1207/s15327043hup1002_3

Borman, W. C., & Motowidlo, S. M. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel Selection in Organizations* (pp. 71-98). San Francisco, CA: Jossey-Bass.

Buehl, A.-K., & Melchers, K. G. (2017). Individual difference variables and the occurrence and effectiveness of faking behavior in interviews. *Frontiers in Psychology, 8*, 1-15. doi:10.3389/fpsyg.2017.00686

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

Carpenter, N. C., Berry, C. M., & Houston, L. (2014). A meta-analytic comparison of self-reported and other-reported organizational citizenship behavior. *Journal of Organizational Behavior, 35*, 547-574. doi:10.1002/job.1909

Cataldi, A. E. (1996). Cognitive activity and deception: Target, referent, and the chance of detection. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 56*, 6460.

Converse, P. D., Peterson, M. H., & Griffith, R. L. (2009). Faking on personality measures: Implications for selection involving multiple predictors. *International Journal of Selection and Assessment, 17*, 47-60. doi:10.1111/j.1468-2389.2009.00450.x

Conway, J. M., & Huffcutt, A. I. (1997). Psychometric properties of multisource performance ratings: A meta-analysis of subordinate, supervisor, peer, and self-ratings. *Human Performance, 10*, 331-360. doi:10.1207/s15327043hup1004_2

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*, 74-118. doi:10.1037/0033-2909.129.1.74

DePaulo, B. M., Stone, J. I., & Lassiter, G. D. (1985). Deceiving and detecting deceit. In B. R. Schlenker (Ed.), *The self and social life* (pp. 323-370). New York, NY: McGraw-Hill.

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

Ellingson, J. E., & McFarland, L. A. (2011). Understanding faking behavior through the lens of motivation: An application of VIE theory. *Human Performance, 24*, 322-337. doi:10.1080/08959285.2011.597477

Frese, M., Fay, D., Hilburger, T., Leng, K., & Tag, A. (1997). The concept of personal initiative: Operationalization, reliability and validity in two German samples. *Journal of Occupational and Organizational Psychology, 70*, 139-161. doi:10.1111/j.2044-8325.1997.tb00639.x

Gilmore, D. C., & Ferris, G. R. (1989). The effects of applicant impression management tactics on interviewer judgments. *Journal of Management, 15*, 557-564. doi:10.1177/014920638901500405

Goffin, R. D., & Boyd, A. C. (2009). Faking and personality assessment in personnel selection: Advancing models of faking. *Canadian Psychology/Psychologie canadienne, 50*, 151-160. doi:10.1037/a0015946

Goffin, R. D., Jelley, R. B., Powell, D. M., & Johnston, N. G. (2009). Taking advantage of social comparisons in performance appraisal: The relative percentile method. *Human Resource Management, 48*, 251-268. doi:10.1002/hrm.20278

Gonzalez-Mulé, E., Mount, M. K., & Oh, I.-S. (2014). A meta-analysis of the relationship between general mental ability and nontask performance. *Journal of Applied Psychology, 99*, 1222-1243. doi:10.1037/a0037547

Griffin, B. (2014). The ability to identify criteria: Its relationship with social understanding, preparation, and impression management in affecting predictor performance in a high-stakes selection context. *Human Performance, 27*, 147-164. doi:10.1080/08959285.2014.882927

Griffith, R. L., Lee, L. M., Peterson, M. H., & Zickar, M. J. (2011). First dates and little white lies: A trait contract classification theory of applicant faking behavior. *Human Performance, 24*, 338-357. doi:10.1080/08959285.2011.597475

Hattrup, K., O'Connell, M. S., & Wingate, P. H. (1998). Prediction of mulitdimensional criteria: Distinguishing task and contextual performance. *Human Performance, 11*, 305-319. doi:10.1207/s15327043hup1104_1

Hogue, M., Levashina, J., & Hang, H. (2013). Will I fake it? The interplay of gender, Machiavellianism, and self-monitoring on strategies for honesty in job interviews. *Journal of Business Ethics, 117*, 399-411. doi:10.1007/s10551-012-1525-x

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75*, 581-595. doi:10.1037/0021-9010.75.5.581

Huffcutt, A. I., Roth, P. L., & McDaniel, M. A. (1996). A meta-analytic investigation of cognitive ability in employment interview evaluations: Moderating characteristics and implications for incremental validity. *Journal of Applied Psychology, 81*, 459-473. doi:10.1037/0021-9010.81.5.459

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

Ingold, P. V., Kleinmann, M., König, C. J., Melchers, K. G., & Van Iddekinge, C. H. (2015). Why do situational interviews predict job performance? The role of interviewees' ability to identify criteria. *Journal of Business and Psychology, 30*, 387-398. doi:10.1007/s10869-014-9368-3

Jansen, A., Melchers, K. G., Lievens, F., Kleinmann, M., 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

Klehe, U.-C., König, C. J., Richter, G. M., Kleinmann, M., & Melchers, K. G. (2008). Transparency in structured interviews: Consequences for construct and criterion-related validity. *Human Performance, 21*, 107-137. doi:10.1080/08959280801917636

Kleinmann, M., Ingold, P. V., Lievens, F., Jansen, A., Melchers, K. G., & König, C. J. (2011). A different look at why selection procedures work The role of candidates' ability to identify criteria. *Organizational Psychology Review, 1*, 128-146. doi:10.1177/2041386610387000

Komar, S., Brown, D. J., Komar, J. A., & Robie, C. (2008). Faking and the validity of conscientiousness: A Monte Carlo investigation. *Journal of Applied Psychology, 93*, 140-154. doi:10.1037/0021-9010.93.1.140

König, C. J., Melchers, K. G., Kleinmann, M., Richter, G. M., & Klehe, U.-C. (2007). Candidates' ability to identify criteria in nontransparent selection procedures: Evidence from an assessment center and a structured interview. *International Journal of Selection and Assessment, 15*, 283-292. doi:10.1111/j.1468-2389.2007.00388.x

Krammer, G., Sommer, M., & Arendasy, M. E. (2017). The psychometric costs of applicants' faking: Examining measurement invariance and retest correlations across response conditions. *Journal of Personality Assessment, 99*, 510-523. doi:10.1080/00223891.2017.1285781

Latham, G. P., & Skarlicki, D. P. (1995). Criterion-related validity of the situational and patterned behavior description interviews with organizational citizenship behavior. *Human Performance, 8*, 67-80. doi:10.1080/08959289509539857

Law, S. J., Bourdage, J. S., & O'Neill, T. A. (2016). To fake or not to fake: Antecedents to interview faking, warning instructions, and its impact on applicant reactions. *Frontiers in Psychology, 7*, 1-13. doi:10.3389/fpsyg.2016.01771

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin, 107*, 34. doi:10.1037/0033-2909.107.1.34

Levashina, J., & Campion, M. A. (2006). A model of faking likelihood in the employment interview. *International Journal of Selection and Assessment, 14*, 299-316. doi:10.1111/j.1468-2389.2006.00353.x

Levashina, J., & Campion, M. A. (2007). Measuring faking in the employment interview: Development and validation of an interview faking behavior scale. *Journal of Applied Psychology, 92*, 1638-1656. doi:10.1037/0021-9010.92.6.1638

Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology, 67*, 241-293. doi:10.1111/peps.12052

Levashina, J., Morgeson, F. P., & Campion, M. A. (2009). They don't do it often, but they do it well: Exploring the relationship between applicant mental abilities and faking. *International Journal of Selection and Assessment, 17*, 271-281. doi:10.1111/j.1468-2389.2009.00469.x

Marcus, B. (2006). Relationships between faking, validity, and decision criteria in personnel selection. *Psychology Science, 48*, 226-246.

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

Martin, C. L., & Nagao, D. H. (1989). Some effects of computerized interviewing on job applicant responses. *Journal of Applied Psychology, 74*, 72-80. doi:10.1037/0021-9010.74.1.72

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

McFarland, L. A., & Ryan, A. M. (2006). Toward an integrated model of applicant faking behavior. *Journal of Applied Social Psychology, 36*, 979-1016. doi:10.1111/j.0021-9029.2006.00052.x

Melchers, K. G., Bösser, D., Hartstein, T., & Kleinmann, M. (2012). Assessment of situational demands in a selection interview: Reflective style or sensitivity? *International Journal of Selection and Assessment, 20*, 475-485. doi:10.1111/ijsa.12010

Melchers, K. G., Klehe, U.-C., Richter, G. M., Kleinmann, M., König, C. J., & Lievens, F. (2009). “I know what you want to know”: The impact of interviewees' ability to identify criteria on interview performance and construct-related validity. *Human Performance, 22*, 355-374. doi:10.1080/08959280903120295

Melchers, K. G., Lienhardt, N., von Aarburg, M., & Kleinmann, M. (2011). Is more structure really better? A comparison of frame-of-reference training and descriptively anchored rating scales to improve interviewers' rating quality. *Personnel Psychology, 64*, 53-97. doi:10.1111/j.1744-6570.2010.01202.x

Mischel, W. (1977). The interaction of person and situation. In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads: Current issues in interactional psychology* (pp. 333–352). Hillsdale, NJ: Erlbaum.

Mueller-Hanson, R. A., Heggestad, E. D., & Thornton, G. C., III. (2006). Individual differences in impression management: An exploration of the psychological processes underlying faking. *Psychology Science, 48*, 288-312.

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology, 78*, 679-703. doi:10.1037/0021-9010.78.4.679

Oostrom, J. K., Melchers, K. G., Ingold, P. V., & Kleinmann, M. (2016). Why do situational interviews predict performance? Is it saying how you would behave or knowing how you should behave? *Journal of Business and Psychology, 31*, 279-291. doi:10.1007/s10869-015-9410-0

Peeters, H., & Lievens, F. (2005). Situational judgment tests and their predictiveness of college students' success: The influence of faking. *Educational and Psychological Measurement, 65*, 70-89. doi:0.1177/0013164404268672

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

Podsakoff, N. P., Whiting, S. W., Podsakoff, P. M., & Blume, B. D. (2009). Individual- and organizational-level consequences of organizational citizenship behaviors: A meta-analysis. *Journal of Applied Psychology, 94*, 122-141. doi:10.1037/a0013079

Preckel, D., & Schüpbach, H. (2005). Zusammenhänge zwischen rezeptiver Selbstdarstellungskompetenz und Leistung im Assessment Center [Correlations between receptive self-presentation competence and performance in an assessment center]. *Zeitschrift für Personalpsychologie, 4*, 151-158. doi:10.1026/1617-6391.4.4.151

Pulakos, E. D., Arad, S., Donovan, M. A., & Plamondon, K. E. (2000). Adaptability in the workplace: Development of a taxonomy of adaptive performance. *Journal of Applied Psychology, 85*, 612-624. doi:10.1037/0021-9010.85.4.612

Rammstedt, B., & John, O. P. (2005). Kurzversion des Big Five Inventory (BFI-K) [Short version of the Big Five Inventory (BFI-K)]. *Diagnostica, 51*, 195-206. doi:10.1026/0012-1924.51.4.195

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T., & Toit, M. (2011). *HLM 7: Hierarchical linear and nonlinear modeling*. Chicago, IL: Scientific Software International.

Reinhard, M.-A., Scharmach, M., & Müller, P. (2013). It's not what you are, it's what you know: Experience, beliefs, and the detection of deception in employment interviews. *Journal of Applied Social Psychology, 43*, 467-479. doi:10.1111/j.1559-1816.2013.01011.x

Roch, S. G., Woehr, D. J., Mishra, V., & Kieszczynska, U. (2012). Rater training revisited: An updated meta-analytic review of frame-of-reference training. *Journal of Occupational and Organizational Psychology, 85*, 370-395. doi:10.1111/j.2044-8325.2011.02045.x

Roulin, N., Bangerter, A., & Levashina, J. (2014). Interviewers' perceptions of impression management in employment interviews. *Journal of Managerial Psychology, 29*, 141-163. doi:10.1108/JMP-10-2012-0295

Roulin, N., Bangerter, A., & Levashina, J. (2015). Honest and deceptive impression management in the employment interview: Can it be detected and how does it impact evaluations? *Personnel Psychology*, 395-444. doi:10.1111/peps.12079

Roulin, N., & Bourdage, J. S. (2017). Once an impression manager, always an impression manager? Antecedents of honest and deceptive Impression Management use and variability across multiple job interviews. *Frontiers in Psychology, 8*, 1-13. doi:10.3389/fpsyg.2017.00029

Roulin, N., Krings, F., & Binggeli, S. (2016). A dynamic model of applicant faking. *Organizational Psychology Review, 6*, 145-170. doi:10.1177/2041386615580875

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

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology, 86*, 162-173. doi:10.1037/0022-3514.86.1.162

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262-274. doi:10.1037/0033-2909.124.2.262

Schmitt, N., Keeney, J., Oswald, F. L., Pleskac, T. J., Billington, A. Q., Sinha, R., & Zorzie, M. (2009). Prediction of 4-year college student performance using cognitive and noncognitive predictors and the impact on demographic status of admitted students. *Journal of Applied Psychology, 94*, 1479-1497. doi:10.1037/a0016810

Schmitt, N., & Oswald, F. L. (2006). The impact of corrections for faking on the validity of noncognitive measures in selection settings. *Journal of Applied Psychology, 91*, 613-621. doi:10.1037/0021-9010.91.3.613

Snell, A. F., Sydell, 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

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*, 245. doi:10.1037/0033-2909.87.2.245

Tett, R. P., & Simonet, D. V. (2011). Faking in personality assessment: A “multisaturation” perspective on faking as performance. *Human Performance, 24*, 302-321. doi:10.1080/08959285.2011.597472

Townsend, R. J., Bacigalupi, S. C., & Blackman, M. C. (2007). The accuracy of lay integrity assessments in simulated employment interviews. *Journal of Research in Personality, 41*, 540-557. doi:10.1016/j.jrp.2006.06.010

Van Iddekinge, C. H., & Ployhart, R. E. (2008). Developments in the criterion-related validation of selection procedures: A critical review and recommendations for practice. *Personnel Psychology, 61*, 871-925. doi:10.1111/j.1744-6570.2008.00133.x

Van Iddekinge, C. H., Raymark, P. H., & Roth, P. L. (2005). Assessing personality with a structured employment interview: Construct-related validity and susceptibility to response inflation. *Journal of Applied Psychology, 90*, 536-552. doi:10.1037/0021-9010.90.3.536

Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement, 59*, 197-210. doi:10.1177/00131649921969802

Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology, 81*, 557-574. doi:10.1037/0021-9010.81.5.557

Wedler, B., Troche, S., & Rammsayer, T. (2008). Studierendenauswahl – Eignungsdiagnostischer Nutzen von Noten aus Schule und Studium [Student selection at colleges and universities: The significance of grades in aptitude diagnosis]. *Psychologische Rundschau, 59*, 123-125. doi:10.1026/0033-3042.59.2.123

Weiss, B., & Feldman, R. S. (2006). Looking good and lying to do it: Deception as an impression management strategy in job interviews. *Journal of Applied Social Psychology, 36*, 1070-1086. doi:10.1111/j.0021-9029.2006.00055.x

Wonderlic Inc. (2002). *Wonderlic Personnel Test and Scholastic Level Exam: User's manual*. Libertyville, IL: Author.

Zettler, I. (2011). Self-control and academic performance: Two field studies on university citizenship behavior and counterproductive academic behavior. *Learning and Individual Differences, 21*, 119-123. doi:10.1016/j.lindif.2010.11.002

Table 1

Descriptive Statistics and Intercorrelations of Study Variables

	Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Interview (applicant condition)	3.62	0.25															
2	Interview (honest condition)	3.20	0.32	.32**														
3	Regression-adjusted-difference score	0.00	0.23	.95**	.00													
4	Cognitive ability	29.27	5.18	.16†	-.05	.18†												
5	ATIC	1.67	0.52	.20*	.16†	.16†	.09											
6	GPA	3.10	0.35	.34**	.15	.31**	.04	.15										
7	Peer rating of academic performance	69.23	11.81	.24*	.01	.25*	.09	.25*	.44**									
8	University citizenship behavior (UCB)	3.47	0.45	.13	.51**	-.03	-.13	-.02	.14	-.04								
9	Sincerity	3.14	0.86	-.04	.00	-.04	.14	-.18	-.16	-.09	-.01							
10	Fairness	3.73	0.94	.08	.15	.04	.08	-.13	-.01	.02	.15	.38**						
11	Greed avoidance	3.51	0.78	-.09	-.11	-.06	.01	-.12	.00	-.04	-.02	.44**	.30*					
12	Modesty	4.06	0.63	-.05	.20*	-.12	.05	-.23*	-.10	-.15	.30*	.33**	.29*	.25*				
13	Extraversion	3.86	0.87	-.01	.06	-.03	-.08	.19†	-.01	-.11	.20†	-.05	-.11	-.14	.09			
14	Conscientiousness	3.91	0.62	.02	.29**	-.07	-.18†	-.03	.25*	.09	.47*	.03	.01	-.09	.13	-.03		
15	Agreeableness	3.08	0.77	-.06	.12	-.10	-.06	-.19†	.06	-.06	.35*	.09	.27*	.13	.40**	.09	.23*	
16	Emotional stability	3.00	0.84	.07	.09	.05	-.22*	.03	.24*	.06	.04	-.16	-.20†	-.13	-.05	-.31**	.09	-.27*

Note. ATIC = ability to identify criteria, GPA = grade point average. N = 111 for all variables except for peer rating of academic performance (N = 95) and for the Big Five (N = 87).

**$p < .01$, *$p < .05$, † $p < .10$ (two-tailed)

Appendix

Items from the short form of the Interview Faking Behavior Scale from Levashina and Campion (2007) in the German translation from Ingold et al. (2015) which we adapted to the university context:

- I have overstated or embellished answers beyond a reasonable description of the truth.
- I have modified or adapted answers to fit the degree program.
- I have created the impression of a fit with the degree program or university in terms of beliefs, values, or attitudes.
- I built stories by combining or arranging study experiences to provide better answers.
- I cooked up better answers.
- I answered based on the experiences or accomplishments of others.
- I did not mention some things in order to improve answers.
- I disguised or concealed aspects of my background to create better answers.
- I improved answers by separating from negative events or experiences.
- I expressed beliefs, values or attitudes held by the interviewer or the university.
- I insincerely praised or complimented the interviewer or organization.

Footnote

1. In the interest of transparency, we want to acknowledge that we have recently published an article in which one of the studies used the same sample as the present paper but with a decidedly different focus, different research questions, and hypotheses. In this other article (Buehl & Melchers, 2017), we described two studies that focused on individual difference variables as antecedents of the occurrence and effectiveness of interview faking. Data from the present study were used for Study 2 in that article. Specifically, we used the same indicator of faking (the regression-adjusted difference score that is described later in this section) and interview performance in the applicant condition as our DVs and tested GMA as a moderator for the effect of self-reported faking behavior (used as a manipulation check in the present article) on faking and on interview performance. However, neither the difference score between the two interview conditions nor any of the criterion variables from the present article were used in the other article nor were any issues related to criterion-related investigated.