Predictors of faking behavior on personality inventories in selection: Do indicators of the ability and motivation to fake predict faking?

Djurre Holtrop1,2 | Janneke K. Oostrom3 | Patrick D. Dunlop1 | Cecilia Runneboom1

1Future of Work Institute, Faculty of Business and Law, Curtin University, Bentley, WA, Australia
2Department of Social Psychology, Tilburg School of Social and Behavioral Sciences, Tilburg University, Tilburg, The Netherlands
3Department of Management and Organization, School of Business and Economics, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Abstract
This study investigated whether faking behavior on a personality inventory can be predicted by two indicators of the ability to fake (cognitive ability and the ability to identify criteria; ATIC) and two indicators of the motivation to fake (perceived faking norms and honesty–humility). Firefighter applicants first completed a personality inventory under high-stakes conditions and, three months later, under low-stakes conditions (n = 128). Analyses revealed very little faking behavior on average. Cognitive ability and ATIC were both negatively related to personality score elevation, but only cognitive ability exhibited a statistically significant association. Neither perceived faking norms nor honesty–humility were significantly related to personality score elevation and only perceived competition was positively related to overclaiming (a proxy of faking).

1 | INTRODUCTION

Despite being one of the most often-used instruments for personnel selection (e.g., Kantrowitz et al., 2018; König et al., 2010), personality inventories are often criticized because of their susceptibility to ‘faking’ (Morgeson et al., 2007). That is, because personality assessments rely on self-report, some respondents may adopt a response set that does not accurately describe their personality, but instead serves the goal of standing out among an applicant pool (e.g., portraying themselves as harder working than they truly are; Ziegler et al., 2012a,b). While many researchers and applied users of personality assessments have spent time contemplating the ‘faking’ problem (e.g., Griffith & McDaniel, 2006; Ziegler et al., 2012b), it remains a vexing phenomenon to study in applied settings. Indeed, multiple theoretical perspectives have been proposed regarding (a) who is most likely to fake and (b) the situational factors that will promote or reduce faking, but field examinations of these causal factors have been relatively scarce. Instead, much of the research into the antecedents of faking behavior has relied on experimental studies with hypothetical job applications or ‘fake-good’ instructions (e.g., MacCann, 2013); we term these types of studies collectively as ‘experimental faking studies’. While experimental faking studies do show how much faking could occur in principle, they are unable to provide an accurate test of the theorized antecedents of faking behavior in the field. Furthermore, the few field studies that have been conducted examined isolated antecedents of faking and not a combination of theorized predictors of faking. Hence, there is a clear need for field studies on the behavioral and motivational antecedents of faking behavior. In this study, we investigate the faking behavior observed among a sample of applicants to firefighter positions, who completed a personality inventory both under application settings and, 3 months later, research settings. In doing so, this is the
first known field study that examines how indicators of the ability and motivation to fake drive faking behavior among job applicants on personality inventories.

1.1 | Faking on personality inventories

Faking behavior was defined by Ziegler et al. (2012a; p. 8) as ‘… a response set aimed at providing a portrayal of the self that helps a person to achieve personal goals’ and added that ‘faking occurs when this response set is activated by situational demands and person characteristics to produce systematic differences in test scores that are not due to the attribute of interest’. To obtain accurate assessments of the extent to which a personnel selection context triggers a systematic change in response sets among applicants requires both low-stakes and high-stakes scores from the same applicants (Tett & Simonton, 2011). However, because it is extremely challenging to collect such data (see Donovan et al., 2014 for an excellent summary of these challenges), much of the faking research has relied on experimental faking studies (e.g., MacCann, 2013), or comparisons of data collected from applicants to those collected from non-applicants (e.g., Birkeland et al., 2006).

Viswesvaran and Ones (1999) meta-analyzed studies that compared honest and faked personality scores under ‘fake-good’ instructions and found that Likert-based measures of all Big Five personality domains are quite easily fakable. Similar results have been found for the HEXACO personality model, which captures an additional honesty–humility dimension beyond the Big Five (Grieve & De Groot, 2011; MacCann, 2013). For example, when instructed to fake good, MacCann (2013) showed that participants are able to increase their HEXACO domain scores by 0.49 (honesty–humility) to 1.08 of a standard deviation (conscientiousness). Research in which applicant scores are compared to those of non-applicants (e.g., employees or research participants) shows smaller effect sizes and more nuanced findings. In a meta-analysis, Birkeland et al. (2006) observed small to moderate group differences on the Big Five domains ($d = 0.11–0.45$). Furthermore, Anglim et al. (2017), comparing job applicants to age- and gender-matched non-applicants, found differences ranging from $d = 0.09$ (openness to experience) to 1.06 (agreeableness) on the HEXACO domains. Of note, Birkeland et al. also found that larger differences emerged on specific job-relevant domains (e.g., extraversion for sales) and, indeed, other research has confirmed that applicants fake according to the cognitive schema of a job (Geiger et al., 2018).

The above findings, however, merely provide proof-of-concept: They do not answer the question whether people actually fake when in real selection contexts (Smith & Ellingson, 2002; Snell & McDaniel, 1998). Within-subjects studies among actual applicants are very scarce. However, the few studies that exists suggest that faking is a pervasive problem. Indeed, Griffith et al. (2007) estimated that 30%–50% of job applicants fake on personality inventories (an estimate that is similar to other studies: Arthur et al., 2010; Donovan et al., 2014; Peterson et al., 2011).

1.2 | Individual differences as determinants of faking

Why, when, and in what form faking behavior emerges has been the topic of considerable theoretical discussion (Goffin & Boyd, 2009; Griffith & Peterson, 2006; Roulin et al., 2016; Tett & Simonton, 2011; Ziegler et al., 2012b). From these discussions, a range of dispositional and situational factors have been identified as potential drivers of faking. Three recurring factors that drive applicants’ decision to fake are: (a) their ability to fake, (b) their motivation to fake, and (c) the opportunity to fake (e.g., Goffin & Boyd, 2009; Roulin et al., 2016; Tett & Simonton, 2011). Whereas the ability and motivation to fake represent manifestations of various individual differences, the opportunity to fake is most often conceived through contextual factors (such as designing faking detection warnings: Fan et al., 2012; or having hard-to-fake tests; e.g., Roulin & Krings, 2016). In this research, we focus on a set of individual differences that are hypothesized to be predictive of applicant faking. In contrast, the contextual factors were held constant in our study because the applicants in our sample were exposed to the same procedures and were applying for the same, highly competitive, role. So far, to the best of our knowledge, few of these predictors have been investigated in the field (Roulin et al., 2016). By investigating the relevance of hypothesized predictors of faking behavior in a real applicant sample, this study contributes to sharpen existing theories of faking behavior.

1.2.1 | Ability to fake

Meta-analytic evidence revealed that nearly every individual possesses some ability to inflate their test scores, if instructed to make a good impression (Viswesvaran & Ones, 1999). However, as noted above, depending on the job context, faking requires some nuance, and research has found that individuals do differ from one another in their ability to fake in an effective way (i.e., to appear highly desirable among an applicant pool). In particular, the ability to fake has often been considered in relation to aspects of cognitive ability (e.g., Roulin et al., 2016; Tett & Simonton, 2011) and the Ability to Identify Criteria (ATIC; Klehe et al., 2012; Roulin et al., 2016).

Cognitive ability

Although theoretical models of faking often cite cognitive ability as a candidate predictor of faking, laboratory studies and field studies seem to yield quite varied results as to its role. On the one hand, faking experiments found that participants with higher cognitive abilities are better able to fake than those with lower cognitive abilities. For example, MacCann (2013) found that trait score elevation on the HEXACO-PI-R—a sign of ‘effective’ faking—was mostly (positively) related to crystallized intelligence and to a lesser degree to fluid intelligence. Geiger et al. (2018) found that faking more closely to a cognitive schema of a job was positively related to applicants’ crystallized intelligence, general cognitive ability, and interpersonal abilities. On the other hand, few field
studies that investigated the effect of cognitive abilities on faking found non-significant or mixed effects of it. Moreover, all these field studies used proxy measures of faking, such as Social Desirability (SD) scores, to estimate faking behavior. For example, De Fruyt et al. (2006) found that, in a large sample of applicants, SD scores (Paulhus, 2002) were unassociated with intelligence, and Levashina et al. (2014) found that applicants with higher cognitive abilities scored higher on an SD scale and lower on extreme responding frequency (i.e., choosing the most extreme response options on a Likert scale). Finally, while investigating faking on a biodata questionnaire (which also relies on accurate self-reporting and thus can be faked), Levashina et al. (2009) found that applicants with higher cognitive ability engaged less in faking behaviors, but among those applicants that chose to fake, those with higher cognitive ability faked more than those with lower ability. Together, previous research suggests that applicants with higher cognitive abilities may employ more strategic or subtle responding strategies when faking than those with lower cognitive abilities and—as a result—are better able to fake.

**Ability to identify criteria**

As a precondition for effective faking, a respondent must also possess some understanding about the construct(s) on which to fake (e.g., Roulin et al., 2016). Such an understanding has been conceptualized in past research as ATIC. Specifically, ATIC is defined as ‘... a person’s ability to correctly perceive performance criteria when participating in an evaluative situation ...’ Thus, the concept of ATIC is based on capturing the correctness of candidates’ perceptions of the performance criteria (i.e., candidates’ assumptions regarding what is being measured) in an actual evaluative situation’ (Kleinmann et al., 2011; p. 129). Confirming that ATIC is indeed a form of ability, previous research found that ATIC is moderately and positively related to cognitive ability (Melchers et al., 2009). Although ATIC is often proposed as a precursor of faking, to the best of our knowledge, only two studies—both of them using instructed ‘fake-good’ designs—have directly investigated the relationship between ATIC and faking, and both involved interviews rather than personality inventories (Buehl et al., 2019; Dürr & Klehe, 2017). Furthermore, together, these findings provide conflicting evidence for ATIC’s theorized relation with faking; Buehl et al. (2019) discovered a significant positive relation between interview ATIC and faking, whereas Dürr and Klehe (2017) did not find a significant relation between interview ATIC and self-reported faking.

Considering that personality inventories are the second most widely used psychometric assessment, surpassing even the popularity of cognitive ability tests (Kantrowitz et al., 2018), we argue that it is especially prudent to also investigate the principles of ATIC in relation to personality inventories and its potential to predict personality inventory faking behavior. Indeed, much like interviews, which typically comprise several questions, probes, and prompts, personality inventories contain a wide variety of stimuli, in the form of items. Furthermore, and again like interviews, responses to the multiple stimuli are aggregated to assess several higher-level characteristics (traits). The information regarding what traits are assessed, which traits are job-relevant, and how the items relate to these traits is rarely provided to applicants. Thus, applicants very likely vary in their understanding of (a) how personality inventories are structured (i.e., both first, and second, how items are aggregated into trait estimates), and (b) which of the measured traits will be considered.

Thus, we construed this variability as ‘Personality Inventory ATIC’, that is defined as an individual’s ability to identify which job-relevant personality traits are being assessed by a personality inventory. In other words, personality ATIC pertains to a person’s ability to correctly perceive the traits that are used to assess the suitability of individuals for the job.

Altogether, we expect that personality inventory ATIC will be an antecedent of faking behavior on personality inventories. Therefore, we developed and tested a novel method for personality inventory ATIC and examined its association with faking.

**Hypothesis 1**  
*a) Cognitive ability and b) personality inventory ATIC will be positively associated with faking.*

**1.2.2 Motivation to fake**

Although nearly all applicants are motivated to obtain job offers, the motivation to fake to receive a job offer might still differ between individuals. Indeed, Roulin et al. (2016) hypothesized that individuals’ motivation to fake depends on two individual differences: (1) the perceived competition and (2) their disposition toward deviant behavior.

**Perceived competition**

In a highly competitive selection situation, the extent to which an applicant believes that other applicants are likely to fake might trigger a stronger perceived need to fake, because not faking ‘may leave them at a competitive disadvantage’ (Griffith & McDaniel, 2006, p. 7). Thus far, the evidence for the effects of the perceived competition, which we term ‘faking norms’, is scarce, and the operationalization of this variable varies across studies. For example, Roulin and Krings (2016) asked people to retrospectively self-report about their most recent job application and found that people with competitive worldviews were more likely to report that they faked in job interviews. The effect of perceived competition for a job, operationalized as the selection ratio, on faking has also been investigated in a few studies. Ho et al. (2019) recently used a vignette study and showed that the perceived competition for a job affects how much applicants intend to fake. Similarly, Buehl and Melchers (2018) used vignettes to investigate the effect of competition but instead found no increase in faking intentions when perceived competition was high. In short, the emerging research on the effect of perceived faking norms and competition shows mixed results and none of this research used real applicants nor directly measured faking. In the present study, we build on this emerging stream of research and, to estimate their perceived
faking norms, asked applicants directly how much they believe that other applicants are willing to fake.

**Hypothesis 2** Perceived faking norms are positively associated with faking.

**Disposition toward deviant behavior**

Some studies indicated that a person’s stable attitudes toward deviant behaviors, such as Machiavellianism (MacNeil & Holden, 2006) or integrity (Wrensen & Biderman, 2005), are likely to affect their faking behavior. Ashton et al. (2004) introduced honesty–humility as a sixth major personality dimension, which describes a person’s tendency to be sincere, modest, fair, and avoid greed. Honesty–humility strongly and negatively correlates with ‘dark’ traits, such as Machiavellianism (Hodson et al., 2018; Muris et al., 2017), positively with integrity (Lee et al., 2008), and negatively with deviant behaviors (Pletzer et al., 2019). So far, however, only a few experimental studies have examined honesty–humility as a predictor of faking, and these either showed null-results or a negative relation. For example, Ho et al. (2019) showed that honesty–humility was negatively related to self-reported faking intentions, based on job interview vignettes. However, another study found no relation between honesty–humility and regression-adjusted difference scores (i.e., a direct measure of faking) on structured interviews (Buehl et al., 2019). Furthermore, among a sample of firefighter applicants, Dunlop et al. (2020) recently found no relation of honesty–humility with overclaiming knowledge of firefighting concepts in their applications. In short, although theoretical models of faking have included honesty–humility or similar ‘moral’ constructs in their model (Roulin et al., 2016; Snell et al., 1999), the evidence is not yet entirely conclusive that (low) honesty–humility is predictive of faking on personality inventories. In the present study, we will build on these previous studies by testing (low) honesty–humility as a predictor of faking among actual applicants.

**Hypothesis 3** Honesty–humility, as assessed in the low-stakes conditions, will be negatively related to faking on other personality dimensions and overclaiming.

All predictors are summarized in Figure 1. After testing the hypotheses, we will explore to what extent this study’s hypothesized predictors of faking overlap in terms of their relations with faking.

**Research Question 1**: To what extent do (a) cognitive ability, (b) ATIC, (c) perceived faking norms, and (d) honesty–humility overlap in their prediction of faking?

Other personality traits, in addition to honesty–humility, are known to affect faking behavior (e.g., McFarland & Ryan, 2000). Generally, these are measured with Big Five personality inventories, of which some dimensions somewhat overlap with honesty–humility. To further contribute the existing body of knowledge, we will explore which personality dimensions (as measured with the HEXACO personality inventory) are predictive of faking.

**Research Question 2**: To what extent do personality dimensions, other than honesty–humility, predict faking?

![FIGURE 1](image-url) Summary of hypotheses and theoretical model of this research
1.3 | Measuring faking

Accurate assessments of faking among applicants are extremely challenging to collect. In a relatively ideal scenario, regression-adjusted difference scores (RADS) are used to empirically estimate faking (Burns & Christiansen, 2011). RADS are the standardized residuals that emerge when high-stakes trait scores are regressed onto their low-stakes trait counterparts. These residuals, therefore, contain faking and error because they capture the part of high-stakes personality that cannot be explained by low-stakes (i.e., ‘honest’) personality.

In the absence of a repeated-measures design, however, researchers do not have access to both low- and high-stakes personality scores. Hence, considerable effort has been spent to devise alternative proxy measures of faking. SD scales (Paulhus, 2002), perhaps the most well-known proxy of faking, have received much criticism for this purported goal: SD scales are not very effective in identifying fakers (Tett & Christiansen, 2007); ‘correcting’ personality scores with SD scales adversely affects the quality of hiring decisions (Christiansen et al., 1994); and SD scales are confounded with meaningful variance of desirable personality traits (De Vries et al., 2014).

1.3.1 | Overclaiming as a proxy for faking

In response to the above-mentioned concerns about SD scales, Paulhus and colleagues proposed an alternative indicator of faking, which they termed the overclaiming technique (Paulhus, 2011; Paulhus & Harms, 2004; Paulhus et al., 2003). When applying the overclaiming technique, participants are asked to indicate their knowledge of items within themed sets. While most items in a set are legitimate (targets), bogus items (foils) are also included. Because it is not possible for participants to be truly knowledgeable of a foil (Dunlop et al., 2017), endorsement of these items is thought to be indicative of faking. Accordingly, if participants claim knowledge of the foils, they are considered to be overclaiming, and thus may also have distorted (overclaimed on) their personality inventory responses (Burns & Christiansen, 2011). Studies on the effectiveness of the overclaiming technique to identify faking behavior have provided mixed evidence (Bing et al., 2011; Feeney & Goffin, 2015; Ludeke & Makransky, 2016; Müller & Moshagen, 2018; O’Connell et al., 2011). Recently, however, Dunlop et al. (2020) showed that the overclaiming technique is most likely to indicate faking if the following two conditions are met: (a) the assessment context includes a high valence outcome (e.g., a desired job) and (b) the overclaiming measure contains relevant content such that participants perceive claiming knowledge as instrumental to attaining that outcome. To further research on overclaiming as a potential proxy for applicant faking, we (a) investigated the extent to which overclaiming relevant job knowledge is related to actual faking behavior and (b) tested the hypotheses and research questions with overclaiming as a dependent variable too.

1.4 | The present study

In the present study, we investigated the prevalence and predictors of faking on personality inventories among a substantial sample of applicants for firefighter positions. In this study, we measured HEXACO personality (Ashton et al., 2004) both during the selection process (high-stakes) and 3 months later (low-stakes), allowing us to measure actual faking behavior (through regression-adjusted difference scores; Burns & Christiansen, 2011). We also collected information on the hypothesized predictors of faking: cognitive ability, ATIC, and perceived faking norms. Finally, we measured overclaiming (Paulhus et al., 2003) as an emerging proxy of faking that is methodologically independent of our main conceptualization of faking.

Because the present study includes low- and high-stakes personality scores of real job applicants, we believe its methodology meets the ‘gold standard’ of faking research. The present study’s within-subjects design allows the strongest conclusions compared to other designs, such as controlled experiments with ‘fake-good’ instructions, applicant versus non-applicant sample comparisons, or within-subjects studies among repeat-applicants.

2 | METHOD

Although data were collected in 2016, hypotheses and analyses for an earlier version of this manuscript were preregistered prior to any hypothesis testing. For the full preregistration document, please refer to this study’s open science framework (OSF) webpage (https://osf.io/wg39h/).

2.1 | Sample and procedure

In 2016, 572 people were assessed in relation to their applications for a firefighter position in Western Australia. These ‘high-stakes’ assessments (to contrast with the ‘low-stakes’ follow-up assessment described below) were conducted online and included a personality inventory, an overclaiming questionnaire, and two cognitive ability tests. Following the assessments, 379 (67%) of the applicants provided permission to the researchers to be contacted for a follow-up survey. Analyses revealed no evidence of differences in demographic composition, personality, or cognitive ability between those who gave permission and those who did not.

Three months after the selection assessments, the 379 applicants were approached via email to complete a follow-up online questionnaire. The applicants were clearly informed that this follow-up survey had no bearing on their current or future job applications with the hiring organization, creating a ‘low-stakes’ assessment condition. Additionally, the applicants were informed that they had a chance of being randomly selected to receive one out of ten $50 gift vouchers from a retailer of their choice. The follow-up survey contained a personality inventory, the personality inventory ATIC measure, and a perceived faking norms measure. Finally, the survey also
included some applicant reaction measures that are not relevant to the research questions of this study. A total of 168 applicants commenced the follow-up survey, however, complete responses were only received from 130 applicants (35.2%). Of these, 41 indicated their application was still under consideration, 10 had been formally offered a position in the firefighter academy, 77 had been rejected, and 2 had withdrawn.

Following a preregistered protocol, the high-stakes and the low-stakes data sets were both inspected for careless responding. Based on the low-stakes questionnaire, two participants were identified with very low variability in item responses between scales (SD < 0.70) and high variability in item responses within scales (SD > 1.60; Barends & De Vries, 2019; Lee & Ashton, 2018) and subsequently excluded from further analyses. The final sample of 128 participants was 85% male (a significantly lower proportion of males than the non-participant group) and was of a mean age of 30.5 years (SD = 5.62, slightly but significantly older than the non-participants, \( d = 0.24, p = .017 \)). There was no evidence of differences between the sample and non-participant group in any of the personality traits, however, the participants tended to perform better than non-participants on the two cognitive tests (\( d = 0.36 \) for the comprehension test, and \( 0.27 \) for the technical test, \( p < .001 \) and \( =0.01 \), respectively).

### 2.2 Identification of job-relevant personality traits for firefighters

Because we know that the manifestation of faking depends on the occupational context (i.e., people tend to fake on measures of job-relevant traits rather than indiscriminately), and ATIC is also context-dependent, we need to reach an understanding of how the firefighter job context maps onto personality. In many countries, firefighting is a prestigious occupation. In Australia, obtaining a role as a career firefighter is so highly coveted that there are organizations and podcasts dedicated to coaching applicants on firefighter selection processes (Clayton, 2020). Generally, firefighter personnel selection is a rigorous process that involves physical testing, psychometric testing, and interviews. Therefore, it is somewhat surprising that, unlike for other emergency responders such as police officers, there is limited research on which personality traits are most predictive of firefighter performance (exceptions include Kwaske & Morris, 2015; Meronek & Tan, 2004). Considering this lack of information, we combined information from a number of sources to determine which personality traits are most relevant for a firefighter position: (a) the job description and selection criteria from the Australian firefighter agency the participants had applied to, (b) the firefighter job description on O’Net, (c) an unpublished data set containing ratings of the perceived social desirability of the items in the HEXACO-Personality-Inventory Revised in relation to two occupation types: emergency responders and other professionals (e.g., a typical office job), (d) the results from research that investigated the personality–performance relationship for firefighters (Kwaske & Morris, 2015; Meronek & Tan, 2004), and (e) general meta-analytic evidence about the personality–performance relationship (e.g., Barrick & Mount, 1991; Zettler et al., 2020).

Consulting the evidence above, the author team independently, but unanimously, identified the three most job-relevant traits for firefighters from the HEXACO model of personality: conscientiousness, (low) emotionality, and honesty–humility. In contrast, the traits agreeableness, extraversion, and openness to experience were viewed as being ambiguously job-relevant for firefighters. The authors, therefore, believe that firefighter applicants should expect the hiring organization to use conscientiousness, (low) emotionality, and honesty–humility as selection criteria, therefore, our measurement of ATIC solely focused on the identification of these job-relevant traits. Faking was only measured as the increase in these three traits.

### 2.3 Measures and operationalization of variables

#### 2.3.1 Cognitive ability

Two cognitive ability tests were administered online, unproctored. The tests were developed by Saville Assessments (Willis Towers Watson). One test was the ‘Swift Comprehension’, and it comprised three short sub-tests that assessed verbal, numerical, and checking ability. The second was the ‘Swift Technical’ test, which assesses, via three short sub-tests, diagrammatic, spatial, and mechanical reasoning. Cognitive ability was operationalized by the first retained factor scores, resulting from factor analyzing all six sub-tests (maximum likelihood EFA).

#### 2.3.2 Ability to identify criteria (ATIC)

For this study, we designed a novel method to measure ATIC in relation to a personality inventory. Typically, ATIC is measured in relation to an assessment that captures a single or small number of construct(s). In these cases, ATIC is measured by asking candidates what they believed was being assessed, after responding to, for example, a question or an assessment center exercise (e.g., Klehe et al., 2012; König et al., 2007). In contrast, personality inventories typically combine the assessment of multiple broad traits (describing broad behavioral tendencies), that consist of a set of underlying facets (which describe narrower behavioral tendencies). Accordingly, to allow the participants to describe all the criteria they thought were being measured with the HEXACO-60, we designed a two-step ATIC measure that was administered directly after the personality inventory in the low-stakes assessment.

First, ATIC was assessed immediately after participants completed the HEXACO-60 by asking them to ‘indicate the skills/characteristics— you think—were being measured with the previous inventory? Participants were provided with up to six text boxes, across two pages, to speculate which skills/characteristics had been measured (example responses include ‘integrity’, ‘honesty’, and ‘leadership’). These
responses were coded independently by two raters who were blind to the hypotheses. These raters compared the provided skills/characteristics to the six domains of the HEXACO and rated these on accuracy on a scale from 0 to 3, where 0 = Wrong, the entry did not match part of any HEXACO dimension, 1 = Entry resembles one facet of a dimension, 2 = Entry matches most of the same dimension, and 3 = Entry comes very close to the same dimension. Negatively keyed entries (e.g., introversion) were considered equally correct as positively keyed entries. In case the same dimension was described more than once by a participant, the description that was rated as most accurate was retained.

Second, for each time that participants provided a characteristic, they were asked to elaborate in a follow-up text box, by providing ‘examples of behavior for the skill/characteristic’. Each HEXACO personality dimension consists of four facets, and this follow-up question served to assess to what extent participants were able to identify all facets of the provided personality dimensions, also for participants who had initially provided an incorrect label. For example, one participant mistakenly labeled a skill/characteristic ‘Courage’ instead of (low) emotionality, evidenced by the description ‘Being able to stay calm and think clearly in dangerous, scary or high-pressure situations’, which relates to some facets of emotionality, especially fearfulness. The same two raters independently compared the provided example behaviors to the facets of the domains, to assess to which level of detail an applicant had correctly understood a dimension. Again, the accuracy of the examples was rated on a scale from 0 to 3, ranging from 0 = Wrong, no facets correctly identified to 3 = Almost entirely right—3 to 4 facets of the same dimension correctly identified. When different examples of behavior described the same facet, the facet was only counted once. If aspects of multiple dimensions were described simultaneously, the raters only counted the facet(s) that matched the dimension provided in the first response.

In short, all participants obtained two scores, per rater, for the six HEXACO traits (i.e., the three clearly job-relevant and three ambiguously relevant traits), indicating (a) to what extent they had identified the dimension correctly and (b) to what extent they identified the underlying facets correctly. The interrater agreement, computed as per the recommendation of Landers (2015), of the ATIC scores for the job-relevant traits showed high consistency, ICC(2, 2) = 0.93, as did the ATIC scores on all six HEXACO dimensions: ICC(2, 2) = 0.89. To calculate applicants’ final ATIC score, we only examined the ratings in relation to the job-relevant traits (McDonald’s ω = 0.66). Therefore, an ATIC score was computed separately for each of honesty–humility, emotionality, and conscientiousness (each ranging from 0 to 6) to form an ATIC score for each job-relevant trait; additionally, an average score for all three job-relevant traits was computed to form an overall personality ATIC score that could range from 0 to 6.2

### 2.3.3 Perceived faking norms

To assess the extent to which applicants thought other applicants would fake, they were asked to ‘Indicate how many other applicants to the Firefighter position you believe would take the three approaches described below when completing a personality inventory as part of their application: (a) Be as honest as possible throughout the personality questionnaire, admitting both flaws and strengths, (b) Be honest about some aspects of their personality but try to make a good impression on others, or (c) Focus only on making a good impression, even if it meant completely ignoring their true flaws and strengths’. Participants then assigned 100 points across all options, showing the % of the applicants that they thought would behave in each way. Response 1 describes honest responding, response 2 describes behavior that is more-or-less expected of job applicants, and response 3 can be considered deception or faking. Therefore, the percentage assigned to option 3 was taken as an estimate of the applicants’ perceived faking norms. On average, participants estimated that 20% of all applicants would only focus on making a good impression (min = 0% and max = 80%).

### 2.3.4 HEXACO personality inventory

High-stakes personality was measured with the full-length HEXACO-PI-R (Lee & Ashton, 2004) and low-stakes personality was measured via the HEXACO-60 (Ashton & Lee, 2009). When drawing comparisons between low- and high-stakes scores, we only used the subset of items from the HEXACO-60, which measures six personality dimensions with 10 items each: Honesty–Humility (H), Emotionality (E), Extraversion (X), Agreeableness (A), Conscientiousness (C), and Openness to experience (O). All items were rated on a 5-point scale ranging from 1 = Strongly disagree to 5 = Strongly agree. Among the 128 retained participants, the McDonald’s ω for the high-stakes scales ranged from 0.72 to 0.83 and for the low-stakes scales from 0.69 to 0.81.

### 2.3.5 Faking estimated with RADS

Faking was estimated empirically via RADS as described in Burns and Christiansen (2011), by regressing the high-stakes trait scores onto the low-stakes trait scores for each of the three traits identified as job-relevant (e.g., high-stakes conscientiousness onto low-stakes conscientiousness) and saving the standardized residuals. Next, the average RADS of all three job-relevant traits was computed. This average RADS served as an ‘omnibus’ measure of faking. When testing the correlations with specific personality dimensions (i.e., for Hypothesis 2 and Research Question 2), if the personality dimension was also a job-relevant trait, we recalculated the average RADS without the criterion that was used to predict faking to eliminate circularity.

### 2.3.6 Overclaiming

Overclaiming was measured with an overclaiming questionnaire that was tailored for firefighter selection, as described in Dunlop...
### TABLE 1  Means, standard deviation, and mean score differences of the HEXACO traits

| Trait                        | Current study | Comparison studies               |                   |
|------------------------------|---------------|----------------------------------|-------------------|
|                              | Low-stakes, N = 128 | High-stakes, N = 128 | Lee and Ashton (2018) | Firefighter incumbents, N = 20 |
|                              | M      | SD    | M      | SD    | t     | p    | r     | d    | M      | SD    | t(df) | p   | d     |
| Conscientiousness            | 4.07   | 0.41  | 4.11   | 0.43  | 1.54  | .13  | 0.74  | 0.13 | 3.54   | 0.63  | 14.42 | 0.00 | -0.85 |
| Emotionality                 | 2.46   | 0.45  | 2.48   | 0.46  | 0.57  | .57  | 0.74  | 0.06 | 2.74   | 0.64  | 6.95  | 0.00 | 0.45  |
| Honesty–humility             | 4.03   | 0.44  | 4.10   | 0.46  | 2.13  | .04  | 0.67  | 0.19 | 3.22   | 0.77  | 20.44 | 0.00 | -1.06 |
|                             |           |       |        |       |       |      |      |      |        |       |       |      |       |
| Agreeableness                | 3.73   | 0.45  | 3.75   | 0.45  | 0.80  | .42  | 0.73  | 0.06 | 3.14   | 0.68  | 14.63 | 0.00 | -0.86 |
| Extraversion                 | 3.95   | 0.41  | 3.99   | 0.40  | 1.96  | .05  | 0.79  | 0.19 | 3.28   | 0.74  | 18.13 | 0.00 | -0.91 |
| Openness to experience       | 3.37   | 0.58  | 3.38   | 0.61  | 0.28  | .78  | 0.83  | 0.03 | 3.71   | 0.65  | 6.58  | 0.00 | 0.52  |

Note: For data collected in the current study, high-stakes personality was measured with the full-length HEXACO-PI-R (Lee & Ashton, 2004) and low-stakes personality was measured via the abbreviated HEXACO-60 (Ashton & Lee, 2009). To ensure equivalence between the two measures, when drawing comparisons between low- and high-stakes scores, we only use and present scores on the items from the HEXACO-60. All traits are measured on a scale from 1 to 5. The Ms and SDs of the low-stakes HEXACO data were also compared to the HEXACO-60 scores of an Australian male subsample (Lee & Ashton, 2018) and to a small sample of U.K. firefighter incumbents (Francis et al., 2018). The t values represent the results of paired t-tests for the high-stakes condition and Welch’s t-tests for the comparison studies. The correlation (r) is the test–retest correlation between low- and high-stakes. The effect sizes (d) for the low- versus high-stakes are based on the pooled standard deviation and take into account the correlation between the two test scores (Morris, 2008).
### Table 2: Means, standard deviations, and correlations between study variables

|       | M   | SD  | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 | 13 |
|-------|-----|-----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 1. Age| 30.51| 5.62|    |    |    |    |    |    |    |    |    |    |    |    |    |
| 2. Gender| 1.15| 0.36| 0.02|    |    |    |    |    |    |    |    |    |    |    |    |
| 3. Current status| 1.40| 0.49| 0.06| -0.12|    |    |    |    |    |    |    |    |    |    |    |
| 4. Cognitive ability| 0.00| 0.81| -0.06| -0.20*| 0.46**|    |    |    |    |    |    |    |    |    |    |
| 5. ATIC| 1.32| 0.84| 0.02| -0.05| -0.05| 0.16| 0.66|    |    |    |    |    |    |    |    |    |
| 6. Perceived faking norms| 19.95| 16.72| 0.04| -0.18*| -0.07| 0.02| -0.09|    |    |    |    |    |    |    |    |
| 7. Honesty–Humility| 4.03| 0.44| 0.02| 0.11| 0.04| -0.01| 0.14| 0.02| 0.75|    |    |    |    |    |    |    |
| 8. Emotionality| 2.46| 0.45| -0.03| 0.21*| -0.17| -0.06| -0.09| -0.06| -0.14| 0.72|    |    |    |    |    |    |
| 9. Extraversion| 3.95| 0.41| 0.02| -0.12| 0.22*| 0.07| 0.14| 0.17| 0.12| -0.30**| 0.77|    |    |    |    |    |
| 10. Agreeableness| 3.73| 0.45| -0.01| 0.02| 0.09| 0.03| 0.00| 0.16| 0.27**| -0.24**| 0.39**| 0.77|    |    |    |    |
| 11. Conscientiousness| 4.07| 0.41| 0.20*| 0.07| 0.05| -0.02| 0.16| 0.00| 0.25**| -0.11| 0.41**| 0.23**| 0.82|    |    |    |
| 12. Openness to experience| 3.37| 0.58| 0.06| 0.09| -0.01| 0.05| -0.04| 0.07| -0.01| 0.19*| 0.11| 0.16| 0.10| 0.83|    |    |
| 13. Faking| 0.00| 0.71| 0.16| 0.03| 0.08| -0.19*| -0.11| -0.09| 0.04| -0.03| 0.11| 0.10| 0.17| -0.07|    |    |
| 14. Overclaiming| -0.37| 0.41| 0.09| -0.18*| 0.01| -0.01| 0.09| 0.19*| 0.08| -0.22*| 0.29**| 0.18*| 0.17| -0.05| 0.12|    |

Note: N = 128. Gender is coded as 1 = male and 2 = female. Current status is coded as 1 = out of the selection process (withdrew or rejected) and 2 = still in the selection process (under consideration or hired). Cognitive ability and faking are standardized residual scores. ATIC is measured on scale from 0 to 6. Perceived norm to fake is indicated by a percentage. The HEXACO personality traits are those measured in the low-stakes condition, on a scale from 1 to 5. Overclaiming is operationalized as the mean c score. For the correlations with the personality dimensions, faking was adjusted to not include the regression-adjusted difference score of the personality dimensions being correlated to faking. McDonald’s ω coefficients are italicized on the diagonal. *p < .01; **p < .05 (two-tailed).
et al. (2020). Applicants were asked to report their knowledge of sets of items, ostensibly describing: firefighting techniques, pieces of firefighting equipment, and household hardware and tools, using the following options: 0 (I have never heard of this item), 1 (I can understand what/who this item is when it is discussed), and 2 (I can talk intelligently about this item). In total, the questionnaire contained 36 real items (12 per item set) and 9 bogus items (3 per item set). The criterion location (c) index, representing the overclaiming bias (Paulhus & Harms, 2004), was calculated using the signal detection theory formula (Macmillan, 1993), essentially averaging the z-score that corresponds to the hit rate (the proportion of real items that the participants ‘knew’) and the z-score that corresponds to the false alarm rate (the proportion of bogus items that the participants ‘knew’), then multiplying that result by negative one (Stanislaw & Toddorov, 1999). The mean c scores on the firefighting techniques, firefighting equipment, and household items were then averaged as the final measure of overclaiming.

3 | RESULTS

3.1 | Preliminary analyses

We first examined the degree to which faking occurred on average among the applicants for the firefighter position. As an omnibus test of the effect of the assessment stakes, we conducted two ANOVAs, one for the means of the job-relevant traits and one for the means of all personality traits. The first 2 (stakes) × 3 (job-relevant traits) ANOVA found that the main effect of assessment stakes was negligible and non-significant, F(1,762) = 1.31, p = .25, partial η² = 0.002. The second 2 (stakes) × 6 (personality traits) ANOVA found a similar negligible effect of stakes, F(1,1524) = 1.72, p = .19, partial η² = 0.001. For the individual HEXACO traits, Table 1 presents the means, standard deviations, and score differences in the high-stakes versus the low-stakes testing condition. The sizes of the differences between low- and high-stakes, expressed via Cohen’s d, were very small, ranging from 0.03 to 0.19. In regards to the job-relevant traits, we found the largest difference between high- and low-stakes scores for honesty–humility, t(127) = 2.13, p = .04, d = 0.19. The other two job-relevant traits, conscientiousness (d = 0.13) and emotionality (d = 0.06) showed very small differences, and indeed the direction of the difference for emotionality was opposite to what one would expect given the trait’s relevance for a firefighter role. For extraversion, which was not a job-relevant trait, we found elevation comparable to honesty–humility, t(127) = 1.96, p = .052, d = 0.19. Altogether, the amount of response elevation (i.e., faking) that occurred from low- to high-stakes appeared to be low, and in one instance was it opposite to what is expected (i.e., emotionality).

The participants displayed an unexpected lack of response elevation from low- to high-stakes. Potentially, this could mean that the low-stakes personality scores were somehow capturing response elevation, although there is no obvious reason why. To further investigate this issue, we compared the low-stakes scores of our sample to the average HEXACO-60 scores of the male subsample in the original validation study of the personality inventory (Lee & Ashton, 2018) and to a small sample of U.K. firefighters (Francis et al., 2018) (see details in Table 1). This comparison showed that the average low-stakes scores of our sample were significantly higher than the average scores of the comparison samples.

Table 2 presents the means, standard deviations, and correlations between our study variables. We found a few significant correlations of demographic variables with our study variables. In line with previous research (Jackson et al., 2009), age correlated positively with conscientiousness, r = 0.20, p = .03. Compared to the male applicants (n = 109), the relatively small group of female applicants (n = 19) scored lower on cognitive ability, t(126) = −2.29, p = .02, d = −0.56, perceived weaker faking norms, t(62.90) = −3.61, p < .01, d = −0.63, scored higher on emotionality, t(126) = 2.47, p = .01, d = 0.61, and overclaimed to a lesser extent, t(126) = −2.02, p = .05, d = −0.45. We also checked whether there were any differences between participants who had been rejected or had withdrawn from the selection process (n = 77) versus those whose application was still under consideration or had been selected (n = 51). We only found significant differences for two variables, that is, cognitive ability, t(126) = 5.86, p < .01, d = 1.06, and extraversion, t(126) = 2.50, p = .01, d = 0.45. The overall faking mean RADS correlated positively, but not significantly, with overclaiming, r = 0.12, p = .18.

3.2 | Measurement of ATIC

Because this paper describes the first attempt to measure personality inventory ATIC, we further explored the properties of the measurement and made four observations. First, McDonald’s ω for all six separate ATIC ratings (i.e., three job-relevant traits X trait and facet level score) was 0.66, providing some evidence of a common (albeit moderate) factor driving ATIC. Second, after combining the two scores for each dimension to form one ATIC estimate per job-relevant trait, dependent t-tests showed that participants identified honesty–humility substantially more often than emotionality, t(127) = 6.21, p < .01, d = 0.55, and emotionality more often than conscientiousness, t(127) = 3.39, p < .01, d = 0.30. Additionally, the job-relevant traits were identified much more frequently than the ambiguously related traits, t(127) = 9.00, p < .01, d = 0.80 (see Figure 2 for all average ATIC levels). Third, participants scored 1.32 on average (SD = 0.84, min = 0.00, and max = 4.17), indicating that participants typically identified approximately one complete job-relevant trait or some partial job-relevant traits. Fourth, after completing the ATIC measure, participants were asked to indicate their level of confidence in their descriptions (not at all to extremely 5-point scale); these ratings were correlated with actual ATIC (r = 0.24, p < .01), indicating that participants could somewhat estimate the extent to which they had identified the criteria correctly.
Indicators of the ability to fake

Hypothesis 1 stated that (a) cognitive ability and (b) ATIC would be positively associated with faking, whereas we observed one significant and one non-significant negative correlation (cognitive ability: $r = -0.19, p = .03$; ATIC: $r = -0.11, p = .24$). Additionally, cognitive ability and ATIC were only correlated to a small and non-significant extent ($r = 0.16, p = .08$). We also conducted a regression analysis which showed that cognitive ability and ATIC together share a very small amount of variance (4%) with faking, $F(2, 125) = 2.77, p = .07$ (Cognitive ability: $B = -0.16, SE = 0.08, \beta = -0.18, t = -2.02, p = .045$; ATIC: $B = -0.07, SE = 0.08, \beta = -0.08, t = -0.87, p = .39$). Furthermore, cognitive ability and ATIC were correlated very weakly with overclaiming, respectively, $r = -0.01, p = .90$ and $r = 0.09, p = .33$. In short, Hypothesis 1—that the cognitive ability and personality inventory ATIC are positively predictive of faking—was not supported, instead we found weak evidence suggesting the reverse.

Indicators of the motivation to fake

Hypothesis 2, which stated that the perceived faking norms would be positively associated with (indicators of) faking, was partly supported. Perceived faking norms were positively correlated with overclaiming, $r = 0.19, p = .04$, but negatively and very weakly with faking behavior, $r = -0.09, p = .29$. Furthermore, we did not find support for Hypothesis 3, that honesty–humility is positively associated with faking and overclaiming. Honesty–humility, as measured in the low-stakes testing condition, was very weakly negatively correlated.
with faking (i.e., the average RADS of emotionality and conscientiousness), \( r = -0.04, p = .84 \), and positively with overclaiming, \( r = 0.08, p = .39 \).

### 3.5 Research questions

To examine the extent to which cognitive ability, ATIC, perceived faking norms, and honesty–humility overlap in their prediction of faking and overclaiming, we conducted two regression analyses (see Table 3). For faking (in this case the average RADS of just emotionality and conscientiousness), we found only negligible regression weights and, again contrary to expectations, the weights for cognitive ability and ATIC were negative. For overclaiming, we found a significant negative regression weight for perceived faking norms, \( B < 0.01, \beta = 0.19, t = 2.20, p = .03 \).

To examine whether any of the HEXACO traits, other than honesty–humility, predicted faking, we examined the correlations of each of the three low-stakes trait scores (i.e., the low-stakes scores) that were regarded as ambiguously job-relevant with faking. For emotionality and conscientiousness, we first created two other faking scores similar to the one created for honesty–humility (i.e., an average RADS of the other two job-relevant traits) and then correlated these traits with the faking score that excluded that particular trait. We did not find any significant correlations of emotionality \( (r = -0.03) \), extraversion \( (r = 0.11) \), agreeableness \( (r = 0.10) \), conscientiousness \( (r = 0.17) \), and openness to experience \( (r = -0.07) \) with faking. For overclaiming, however, we found more substantial correlations: emotionality \( (r = -0.22) \), extraversion \( (r = 0.30) \), agreeableness \( (r = 0.18) \), conscientiousness \( (r = 0.17) \), and openness to experience \( (r = -0.05) \).

Altogether, the results showed, on average, little evidence of faking, based on participants’ scores differences in the low-stakes and high-stakes assessments. Furthermore, both cognitive ability and ATIC were negatively associated with the individual differences in faking on the HEXACO measure, and only cognitive ability significantly so.

### 4 Discussion

The goal of the present study was to investigate the theorized predictors of faking on personality inventories among a sample of real applicants to a highly competitive and coveted role. To this end, we collected measures of personality from a sample of applicants both as part of the selection process (high-stakes) and 3 months later (low-stakes), allowing us to measure actual faking behavior directly through RADS. Our results showed low prevalence of faking overall on average. Specifically, participants only faked significantly on honesty–humility (a job-relevant trait) and marginally on extraversion (a trait that is less clearly job-relevant). Moreover, for one of the job-relevant traits, emotionality, the mean difference in scores was in the opposite of the expected direction—average levels increased from low- to high-stakes, albeit not to a statistically significant extent. Next, we tested the role of two sets of frequently theorized antecedents of faking, cognitive ability and ATIC as indicators of the ability to fake, and perceived faking norms and honesty–humility as indicators of the motivation to fake. Opposite to our expectations and existing theories, the indicators of the ability to fake were negatively related to actual faking behavior, but only cognitive ability significantly so. Furthermore, the indicators of the motivation to fake were only related to faking to a very modest extent; neither motivational antecedent was significantly related to faking behavior, and only perceived faking norms showed a small relation with overclaiming, a proxy measure of faking. All in all, our study shows a low prevalence of faking and very limited explanation of faking behavior by theoretically sound antecedents.

### 4.1 Theoretical implications

Our findings have several theoretical implications. First, our findings seem to suggest that faking might be less prevalent than often feared. Although the position of firefighter is highly coveted, the applicants in our sample, on average, faked much less than participants in experimental fake-good studies (e.g., MacCann, 2013) and somewhat less than applicants in some other field studies (e.g., Arthur et al., 2010; Birkeland et al., 2006; Ellingson et al., 2007). The difference in faking prevalence observed here with instructed faking experiments is to be expected, the applicants in this study were not explicitly instructed to manage impressions, thus not every applicant may have felt the need to fake. However, the contrast of our results against those from other field studies is more puzzling (e.g., Arthur et al., 2010; Donovan et al., 2003; Griffith et al., 2007; Peterson et al., 2011).

To understand the low faking prevalence in our sample, we compared the average low-stakes scores observed in this study’s sample to those from other low-stakes samples. We found mostly large score differences with the HEXACO-60 validation sample and with a small sample of firefighter incumbents. Considering the score differences with these other samples, it may be possible that the low-stakes scores in this study were contaminated by conscious or unconscious response elevation; in other words, even in the low-stakes condition, our participants may have been faking. This potential response elevation in the low-stakes scores may have reduced the mean difference between conditions. Still, despite the low mean differences between the low- and high-stakes condition, we observed individual differences in the degree of faking behavior; some participants showed more elevated scores than others in the high-stakes condition. To be precise, the SDs of the raw difference scores on the job-relevant traits ranged from 0.30 for conscientiousness to 0.37 for honesty–humility. To put this variance in faking into perspective, we compared it to the variances found in other samples. For example, Mason and Ryan (2000) reported a 0.63 difference score SD on conscientiousness in an experimental faking study and Peterson et al. (2011) reported a 0.72 difference score SD on...
conscientiousness in a field study. Compared to these two studies, this study’s sample showed less variance in faking behavior.

We offer three potential explanations for this study’s lower average levels of faking and lower variance in faking behavior. The first explanation pertains to the context of the high-stakes assessment condition. Specifically, it may be that different selection situations do not evoke faking behavior equally as they may provide different faking demands and opportunities. Indeed, the opportunity to fake was constant in this investigation and applicants’ perceptions thereof may have been different from other studies in this regard, evoking different or less faking behaviors here. Two alternative explanations for the low levels of faking pertain to the low-stakes condition. First, our participants completed the high-stakes assessment before the low-stakes assessment, therefore, we cannot rule out any order effects. Indeed, Ellingson et al. (2007) found larger difference scores for participants who first completed a personality inventory for development and then for selection purposes (\(d = 0.73–1.44\)) than for participants who first completed the inventory for selection and then for development purposes (\(d = 0.27–0.74\)). It might be that our participants wanted to appear consistent and thus tried to replicate their high-stakes response pattern in the low-stakes condition. We hasten to note that a lot of effort was put into making the low-stakes communication and instructions as unassuming as possible, and that considerable time (3 months) had passed since their high-stakes assessment. Second, it is possible that the participants completed the low-stakes assessments with their proverbial ‘firefighter applicant hat’ on. In other words, perhaps the context of our research may have prompted the participants to tailor their responses to how they see themselves as a firefighter and not how they are in general. Such a context effect would have brought the honest responses closer to the faked responses. To an extent, the difference between the mean scores of the firefighter incumbent sample (Francis et al., 2018) and the HEXACO community sample (Lee & Ashton, 2018; Table 1) support this notion. All firefighter incumbent mean scores were closer to our sample’s honest scores than the community sample’s scores.

Second, while keeping in mind that we found limited evidence of faking in the first place, our findings appear to contradict elements of theoretical models of faking (e.g., Roulin et al., 2016; Tett & Simonet, 2011) and previous findings from fake-good experiments (e.g., Geiger et al., 2018; MacCann, 2013) by showing that among actual applicants, cognitive ability was negatively correlated with faking. The phenomenon that findings from a lab environment do not replicate in the field is not uncommon. Indeed, upon comparing results from a large number of lab and field studies, Mitchell (2012) concluded ‘any psychological results found in the laboratory can be replicated in the field, but the effects often differ greatly in their size and less often (though still with disappointing frequency) differ in their directions’ (p. 114). Thus, we also do not wish to dismiss our findings as a mere anomaly simply because it makes sense theoretically that applicants with higher cognitive ability should be better able to fake. Instead, we believe that the need for field studies on the individual differences and circumstances that elicit faking is only further emphasized by this study’s incongruent findings.

In an attempt to address these puzzling findings, we explored two possible explanations. First, Levashina et al. (2009) showed that applicants with higher abilities fake less, but when they choose to fake they faked more than applicants with lower abilities. Following their approach, we also explored polynomial relations between the ability to fake and faking behavior, but found no evidence of any. Second, the negative relation might originate from applicants’ perceived need to fake. In this case, applicants were able to choose which assessments (personality or cognitive ability) to complete first. Thus, applicants who completed the cognitive ability tests prior to the personality inventory, and thought that they performed badly on the cognitive ability tests, might have felt a need to compensate their lower test scores on the cognitive ability tests by trying to elevate their scores on the personality inventory. Applicants who (think that they) performed well on the cognitive ability tests might have felt more confident about their hiring chances and, therefore, might not have felt such a need. Meanwhile, applicants who completed the cognitive ability tests after the personality inventory cannot use information about their cognitive test performance to guide their personality responding strategy. Our data show some tentative support for this explanation: Among the applicants who completed the cognitive ability tests before the personality inventory (\(n = 52\), the negative correlation between cognitive ability and faking was significant (\(r = -0.34, p = .02\)). However, among the applicants who completed the personality inventory before the cognitive ability tests (\(n = 76\)), the correlation between cognitive ability and faking was weaker (\(r = -0.07, p = .55\)). The difference between these two correlations is not significant (\(Z = 1.54, p = .12\)), however, although we note that our small sample offers only an under-powered test of that difference. Still, if this explanation would be true, this has implications for the design of selection procedures, as administering personality inventories at a later phase in the procedure might influence applicants’ faking behaviors.

Third, this study’s field sample presented a unique opportunity to investigate individual differences in indicators of the motivation to fake instead of a blanket experimental condition that aims to motivate all participants to fake. In most lab studies, there is little motivation to fake other than doing what the researcher has instructed, or perhaps a small monetary incentive. In contrast to expectations, perceived faking norms and honesty–humility, as indicators of the motivation to fake, played a minimal role in applicant faking. In fact, only perceived faking norms were related to overclaiming, as a proxy to faking, and this relation was modest (\(r = 0.19\)). Note, however, that these findings must still be considered with the fact that little (variance in) faking was observed in the first place. Additionally, the participants varied widely in how many applicants they thought engage in faking, ranging from 0% to 80% (\(M = 20\%\), median = 15%). The emerging body of literature that investigates applicants’ perceptions of the selection context (Buehl & Melchers, 2018; Ho et al., 2019; Roulin & Krings, 2016), approaches
these perceptions in various ways (such as competitive worldviews, selection ratio, and number of competing applicants). In our study, we asked to what extent the applicant thought that other applicants would focus only on making a good impression while completing the personality inventory. This study's method appears to overlap mostly with the competitive world views operationalization, which was found to be related to retrospectively reported applicant faking (Roulin & Krings, 2016), and appears to overlap less with the selection ratio operationalization, which was found to be both related (Ho et al., 2019) and unrelated (Buehl & Melchers, 2018) to faking. Considering these findings together with the finding in the present study, it seems that applicants' perception of the selection context somewhat affects their response biases. Future research could attempt to investigate if field interventions, for example, downplaying the competitive nature of a selection process, can reduce perceptions of competition and result in less faking behavior.

Lastly, we introduced a new conceptualization and measurement of ATIC for a multi-dimensional personality inventory. Previous studies have already measured multidimensional assessment exercises (e.g., König et al., 2007) or single dimension personality scales (König et al., 2006). In this study, we defined personality inventory ATIC as an individual's ability to identify which job-relevant personality traits are being assessed by a personality inventory. Accordingly, ATIC was assessed for each personality dimension by coding the responses to open-ended questions. Although the responses were sufficiently clear, as two independent raters generally agreed in their ratings (ICC2, 2 = 0.93), the Omega reliability was acceptable for a highly complex measure (ω = 0.66).

### 4.2 Practical implications

Faking behavior can have detrimental effects on personnel selection: Some research shows that it seems to undermine the construct and criterion validity of personality assessments (Donovan et al., 2014; Ellingson et al., 2001), as scores become contaminated with systematic, but construct-irrelevant variance (e.g., Heggestad et al., 2006). Hence, it is imperative to prevent faking. While keeping in mind that our sample showed less (variance in) faking behavior than usual, our findings also suggest that it can be difficult to predict individual differences in faking behavior in some settings. Not only did ATIC, perceived faking norms, and honesty–humility show negligible relations to faking and was cognitive ability unexpectedly negatively related to faking, but also overclaiming, as a proxy of faking behavior, failed to explain a meaningful amount of the variance in faking. Simply put, predicting faking was very difficult and even counter-intuitive in this sample. We thus challenge any broad claims of being able to detect faking with the same set of predictors in any setting. Importantly, because organizations cannot know the extent to which their applicant sample has faked, nor the variances in faking behavior (as low-stakes scores are not available in such a sample), it is nigh impossible to verify that faking can be detected in their setting. Thus, we believe that organizations should turn away from attempting to identify fakers and turn to interventions at the test-development and instruction level to shape the perceived opportunity to fake, such as adding detection warnings (Dwight & Donovan, 2003; Fan et al., 2012). We also encourage organizations to put the onus on test-developers to develop harder-to-fake personality inventories.

### 4.3 Limitations and suggestions for future research

This study has some limitations that are especially worth noting. First, we believe that the role of firefighter attracts a specific group of applicants, therefore, a self-selection bias is likely to have been present and may have affected our findings. Indeed, this suspicion is strengthened by the comparison with a normative data set (i.e., the male community sub-sample in Ashton & Lee, 2009) and the mean scores of firefighter incumbents (Francis et al., 2018). We note that the role of firefighter has strong stereotypes attached to it and O'Net job descriptions suggest that firefighter requirements do not match typical production, office, sales, or managerial roles. The unique attributes of the firefighter role and its strong stereotype may attract a specific type of person with a very positive and role-congruent self-image. As a result, this self-selection bias could potentially lead to higher low-stakes scores and leave less room for response elevation. As such, our results may not readily generalize to other, more common, roles. Additionally, because our analyses were conditional on the willingness of participants to respond to our invitation to our follow-up survey, there may have been variables relevant to this willingness that have biased the results. For example, perhaps 'true' honesty–humility determines both the willingness to fake in high-stakes (negatively), and the willingness to participate (positively). To the extent that this is true, the final sample would comprise only people who are truly high on honesty–humility, whereas those who faked their honesty–humility in high-stakes never completed the low-stakes assessment. Hence, there is a clear need for more research on the individual differences and circumstances that elicit faking in actual selection contexts.

Second, upon reflection, some adjustments might improve the quality of ATIC measurement. First, the ATIC task was very ‘difficult’. Indeed, some participants could apparently not even think of any characteristic that the personality inventory was measuring. Other research also found that not every applicant actively considers what personality inventories are trying to measure (König et al., 2012), perhaps meaning that participants are not equally motivated to look for job-relevant traits. However, if ATIC is indeed an ability, which is further supported by its (weak but) positive correlation with cognitive ability in our sample, it would be preferable that every participant is equally motivated to use their ATIC during the whole exercise (i.e., including completing the personality inventory)—so that the ability is not conflated with motivation to search for job-relevant traits. To that end, it might be preferable to inform participants before they complete the personality inventory that they will be asked to report the job-relevant traits that were measured afterward. Second, participants were provided with the opportunity to provide...
three characteristics per page and could indicate if they wanted to provide more. Only, a small number of participants (n = 19) provided characteristics on the second page. We would, therefore, propose to provide participants all response opportunities on the same page. Finally, to better estimate the effect of ATIC on faking, it seems prudent to also ensure that applicants themselves believe the traits that they have identified are important for the job. This belief is likely to be an important step in undertaking faking behavior and should thus make a worthwhile inclusion to map the process from ATIC to faking. As such, future research should instruct participants to report characteristics that are indicative of positive job-relevant behaviors.

Lastly, we wish to make some general observations regarding the prediction of faking behavior. First, much work has investigated a limited number of isolated antecedents of faking. The present study examined a larger set of antecedents and found that, collectively, these only predicted a very small amount of faking behavior. Still, we did not measure the opportunity to fake and did not have enough statistical power to explore multiplicative effects. We encourage future research to investigate these venues for these are noticeable omissions in our work and other fields. Second, we noticed that current models of faking behavior still forego some important considerations that applicants may have when completing a personality inventory. For example, applicants consider detrimental long-term effects of faking. Indeed, when interviewing real applicants, König et al. (2012) found, for example, that some applicants may be reluctant to fake out of fear of detection and that some do not want to present themselves differently than they really are. We, therefore, encourage future research to attempt the difficult task of mapping the cognitive processes that real applicants go through while completing personality inventories.

5 | CONCLUSION

Using a sample of firefighter applicants, this study shows that cognitive ability is negatively related to faking on personality inventories and that perceived faking norms are positively related to overclaiming. However, altogether, the applicants faked very little when their high-stakes assessment was compared to their low-stakes assessment. All in all, our study shows the importance of more within-subjects studies among real applicants to better understand the prevalence and predictors of faking.

ACKNOWLEDGMENTS

The authors thank Courtenay McGill for her efforts in coding the textual data, and two reviewers and the Associate Editor whose comments/suggestions greatly helped to improve this manuscript.

ORCID

Djurre Holtrop https://orcid.org/0000-0003-3824-3385
Janneke K. Oostrom https://orcid.org/0000-0002-0963-5016
Patrick D. Dunlop https://orcid.org/0000-0002-5225-6409

ENDNOTES

1 Please see this project’s OSF page for more details on the information sources that informed the rank-order the traits and the authors’ considerations on how to rank the traits.

2 This project’s OSF page includes a file with excerpts from our communications with the review team about the measurement of ATIC. We hope that the considerations in these communications may serve future researchers wishing to study ATIC in a multidimensional (personality) measure.

REFERENCES

Anglim, J., Morse, G., De Vries, R. E., MacCann, C., Marty, A., & Mõttus, R. (2017). Comparing job applicants to non-applicants using an item-level bifactor model on the HEXACO personality inventory. European Journal of Personality, 31(6), 669–684. https://doi.org/10.1002/per.2120

Arthur, W. Jr, Glaze, R. M., Villado, A. J., & Taylor, J. E. (2010). The magnitude and extent of cheating and response distortion effects on unproctored internet-based tests of cognitive ability and personality. International Journal of Selection and Assessment, 18(1), 1–16. https://doi.org/10.1111/j.1468-2389.2010.00476.x

Ashton, M. C., & Lee, K. (2009). The HEXACO - 60: A short measure of the major dimensions of personality. Journal of Personality Assessment, 91(4), 340–345. https://doi.org/10.1080/00223890902935878

Ashton, M. C., Lee, K., Perugini, M., Szatora, P., de Vries, R. E., Di Blas, L., Boies, K., & De Raad, B. (2004). A six-factor structure of personality-descriptive adjectives: Solutions from psycholinguistic studies in seven languages. Journal of Personality and Social Psychology, 86(2), 356–366. https://doi.org/10.1037/0022-3514.86.2.356

Barens, A. J., & De Vries, R. E. (2019). Noncompliant responding: Comparing exclusion criteria in MTurk personality research to improve data quality. Personality and Individual Differences, 143, 84–89. https://doi.org/10.1016/j.paid.2019.02.015

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. Personnel Psychology, 44(1), 1–26. https://doi.org/10.1111/j.1744-6570.1991.tb00688.x

Bing, M. N., Kluemper, D., Davison, H. K., Taylor, S., & Novicevic, M. (2011). Overclaiming as a measure of faking. Organizational Behavior and Human Decision Processes, 116(1), 148–162. https://doi.org/10.1016/j.obhdp.2011.05.006

Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. International Journal of Selection and Assessment, 14(4), 317–335. https://doi.org/10.1111/j.1468-2389.2006.00354.x

Buehler, A.-K., & Melchers, K. G. (2018). Do attractiveness and competition influence faking intentions in selection interviews? Journal of Personnel Psychology, 204–208. https://doi.org/10.1027/1866-588B/a000208

Buehler, A.-K., Melchers, K. G., Macan, T., & Kühnel, J. (2019). Tell me sweet little lies: How does faking in interviews affect interview scores and interview validity? Journal of Business and Psychology, 34(1), 107–124. https://doi.org/10.1007/s10869-018-9531-3

Burns, G. N., & Christiansen, N. D. (2011). Methods of measuring faking behavior. Human Performance, 24(4), 358–372. https://doi.org/10.1080/08999285.2011.597473

Christiansen, N. D., Goffin, R. D., Johnston, N. G., & Rothstein, M. G. (1994). Correcting the 16PF for faking: Effects on criterion-related validity and individual hiring decisions. Personnel Psychology, 47(4), 847–860. https://doi.org/10.1111/j.1744-6570.1994.tb01581.x

Clayton, B. (2020). https://firerecruitmentaustralia.com.au/

De Fruyt, F., Aluja, A., Garcia, L. F., Rolland, J. P., & Jung, S. C. (2006). Positive presentation management and intelligence and the personality differentiation by intelligence hypothesis in job applicants.
Zettler, I., Thielmann, I., Hilbig, B. E., & Moshagen, M. (2020). The nomological net of the HEXACO model of personality: A large-scale meta-analytic investigation. *Perspectives on Psychological Science, 15*(3), 723–760. https://doi.org/10.1177/1745691619895036
Ziegler, M., MacCann, C., & Roberts, R. D. (2012a). Faking: Knowns, unknowns, and points of contention. *New perspectives on faking in personality assessment* (pp. 3–16). Oxford University Press.
Ziegler, M., MacCann, C., & Roberts, R. D. (2012b). *New perspectives on faking in personality assessment*. Cambridge University Press.

**How to cite this article:** Holtrop D, Oostrom JK, Dunlop PD, Runneboom C. Predictors of faking behavior on personality inventories in selection: Do indicators of the ability and motivation to fake predict faking? *Int J Select Assess*. 2021;29:185–202. https://doi.org/10.1111/ijsa.12322