

Röhner, Jessica; Degro, Mia; Holden, Ronald. R.; Schütz, Astrid

**Article — Published Version**

## A Registered Report to Disentangle the Effects of Frame of Reference and Faking in the Personnel-Selection Scenario Paradigm

International Journal of Selection and Assessment

**Provided in Cooperation with:**

John Wiley & Sons

*Suggested Citation:* Röhner, Jessica; Degro, Mia; Holden, Ronald. R.; Schütz, Astrid (2025) : A Registered Report to Disentangle the Effects of Frame of Reference and Faking in the Personnel-Selection Scenario Paradigm, International Journal of Selection and Assessment, ISSN 1468-2389, Wiley, Hoboken, NJ, Vol. 33, Iss. 2, <https://doi.org/10.1111/ijsa.70012>

This Version is available at:

<https://hdl.handle.net/10419/323790>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<http://creativecommons.org/licenses/by/4.0/>

## REGISTERED REPORT OPEN ACCESS

# A Registered Report to Disentangle the Effects of Frame of Reference and Faking in the Personnel-Selection Scenario Paradigm

Jessica Röhner<sup>1</sup>  | Mia Degro<sup>1</sup>  | Ronald. R. Holden<sup>2</sup>  | Astrid Schütz<sup>1</sup> 

<sup>1</sup>Department of Psychology, University of Bamberg, Bamberg, Bavaria, Germany | <sup>2</sup>Department of Psychology, Queen's University, Kingston, Ontario, Canada

**Correspondence:** Jessica Röhner ([jessica.roehner@uni-bamberg.de](mailto:jessica.roehner@uni-bamberg.de))

**Received:** 7 April 2025 | **Revised:** 7 April 2025 | **Accepted:** 10 April 2025

**Funding:** This research was partly funded by a grant from the Equal Opportunities Office and by a grant from the Internal Research Fund at the University of Bamberg.

**Keywords:** contextualization | faking | frame of reference | instructions in faking research | internal validity

## ABSTRACT

In laboratory faking research, participants are often instructed to respond honestly (generic instructions [GIs], control condition) or to fake (personnel-selection scenario [PSS], faking condition). Considering the research on instruction-level contextualization, a PSS might not only motivate participants to fake but might also promote the adoption of a work frame of reference (FOR). Thus, differences in responses between faking and control conditions could partly result from FOR effects. (Full) item-level contextualization can also be used to promote the adoption of a work FOR, and the adoption through this route is stronger than through instruction manipulation. We combined the two approaches to disentangle FOR and faking, conducted a 4-wave longitudinal study with a 2 (instructions: GIs vs. PSS) × 2 (full item-level work contextualization absent vs. present) repeated-measures design ( $N = 309$ ), and compared the effects of these conditions on three HEXACO-PI-R scales (Conscientiousness, Emotionality, Honesty-Humility). Irrespective of the investigated personality trait, the ANOVAs revealed significant main effects. As expected, compared with GIs, the PSS increased the adoption of a work FOR, and the effects were smaller than the effects of full item-level work contextualization present (vs. absent). Also, as expected, the PSS (vs. GIs) and full item-level work contextualization present (vs. absent) changed participants' scale mean scores. However, importantly, there were no interaction effects. Exploratory mediation analyses indicated direct rather than indirect (mediator: adoption of a work FOR) effects of instructions on participants' scale mean scores. In conclusion, the internal validity of faking research is *not threatened* by confounding FOR effects.

Imagine taking part in a study. You are asked to respond honestly to a self-report personality measure. Afterward, you retake the measure as part of a hypothetical personnel-selection scenario (PSS) and are asked to fake your responses in a way that would help you get a job offer. When asked to answer honestly, you respond while keeping various contexts in mind (i.e., you

respond with respect to how you think you generally behave). While faking your responses in the hypothetical PSS, you think predominantly about a work context (i.e., how you think you behave at work). Does the difference in your responses between the two conditions reflect differences in the adoption of a work frame of reference (FOR), faking behavior, or both?

We thank Djurre Holtrop for sending the fully contextualized personality questionnaires. We thank Theresa Waclawek and Jane Zagorski for language editing. Last but not least, we are grateful for the helpful comments we received from Olivier Corneille.

Unfortunately, Mia Degro was unable to finish the project but approved the submission.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *International Journal of Selection and Assessment* published by John Wiley & Sons Ltd.

## Summary

- Faking impacts decision-making (e.g., in application contexts), and self-report personality measures are particularly subject to faking.
- Faking is often studied by comparing participants' scale mean scores under generic instructions (GIs) versus a personnel-selection scenario (PSS), and the latter might promote the adoption of a work frame of reference (FOR).
- FOR and faking may thus be confounded when using a PSS—a conjecture that we tested by replicating this procedure while varying full item-level work contextualization and investigating faking independent of FOR effects.
- The current study revealed that the mechanisms by which instructions (GIs vs. PSS) and full item-level work contextualization (absent vs. present) operate are independent.
- Exploratory mediation analyses indicated direct rather than indirect (mediator: adoption of a work FOR) effects of instructions on participants' scale mean scores, and the current study thus demonstrates that the internal validity of faking research is *not threatened* by confounding FOR effects.

We argue that FOR and faking both impact the difference between the “honest” condition and the PSS and that this issue has often been neglected in previous research. In the present study, we thus attempted to disentangle the effects of FOR and faking.

## 1 | The Importance of Valid Personality Assessment

Self-report personality measures are important and frequently used measures in psychology (e.g., Dang et al. 2020). They are essential not only for research purposes (e.g., Sassenberg and Ditrich 2019) but also for applied contexts (e.g., personnel-selection; Schmidt-Atzert and Amelang 2012). Scientific conclusions as well as significant decisions (e.g., personnel decisions) are often primarily or at least partly based on information obtained from self-report personality measures (Diekmann and König 2015; Sassenberg and Ditrich 2019). Thus, the information that is obtained must validly predict the criteria of interest (e.g., Sassenberg and Ditrich 2019). However, participants' dishonest self-presentation (i.e., *faking*; Mueller-Hanson et al. 2003; Röhner et al. 2022; Salgado 2016) undermines the validity of psychological measures, and self-report personality measures are particularly subject to faking (Birkeland et al. 2006; Hu and Connelly 2021; Viswesvaran and Ones 1999).

## 2 | Participants' Dishonest Self-Presentation: The Phenomenon of Faking

### 2.1 | Defining Faking

Faking is part of the broader concept of self-presentation (e.g., Marcus 2009). Whereas honest self-presentation can be

defined as “legitimate behavior inevitably triggered by situational demands” (Marcus 2009, p. 427), dishonest self-presentation is described as the “intentionally deceptive presentation of attributes [that respondents] do not truly believe to possess” (Marcus 2009, p. 419). Faking (i.e., an intentional, goal-oriented, and [socially] motivated behavior, emerging from an interaction between personal and situational variables to create a deceptive and inaccurate impression; Röhner and Schütz 2020; see also Ziegler et al. 2012) can be understood as the latter, the dishonest facet of self-presentation. Therefore, changes in participants' responses due to faking do not or only temporarily translate into actual behavior (see also Tett and Simonet 2021).

### 2.2 | The Importance of Faking Research

Many studies have demonstrated the negative consequences of faking for psychological assessment (e.g., Birkeland et al. 2006; see also Tett and Simonet 2021). For example, faking can affect statistical characteristics (e.g., means, standard deviations, and factor structures; Salgado 2016; Ziegler et al. 2012) and subsequent decisions (Jeong et al. 2017; Schmitt and Bradburn 2018; Sisco and Reilly 2010; Stewart et al. 2010). Nevertheless, the antecedents, the consequences, and the process of faking are not yet fully understood (Bensch et al. 2019; Röhner and Ewers 2016; Röhner et al. 2022), which makes further research of this phenomenon essential.

## 3 | The Dilemma When Investigating Faking

### 3.1 | Assessing Faking Without (Valid) Faking Indicators

To investigate faking, researchers need to ascertain who faked and who did not. However, faking has proven to be very difficult to detect (e.g., Röhner et al. 2023; Ziegler and Bensch 2015). Not only has the traditional way of detecting faking with the use of so-called lie scales (e.g., Crowne and Marlowe 1960; Lambert et al. 2016) been criticized because the validity of such scales has been demonstrated to be questionable (e.g., Connelly and Chang 2016; Lanz et al. 2022; Röhner et al. 2024), but also, alternative types of valid faking indicators are rare (see Röhner and Schütz 2019, for an overview). Moreover, in most cases, faking indicators are restricted to a subset of psychological measures (e.g., the IAT, Röhner et al. 2023; the d2, Schmidt-Atzert et al. 2004), or their application is useful only under certain measurement conditions (e.g., when combined with warnings; Röhner and Holden 2022). As researchers have a need to ascertain who faked and who did not when valid faking indicators are still missing, they have systematically induced faking in the laboratory by manipulating participants' motivation to fake (e.g., Roulin et al. 2016).

According to faking models, three prerequisites are typically considered necessary for faking to occur: the *motivation*, the *ability*, and the *opportunity* to fake (e.g., Roulin et al. 2016; Tett and Simonet 2011). However, because participants' abilities and opportunities to fake have been demonstrated in several studies

(e.g., Rasheed and Robie 2023; see also Ziegler et al. 2012 for an overview), even when measures that are supposedly immune to faking are administered (e.g., Röhner and Thoss 2018; Teige-Mocigemba et al. 2015), these two factors are generally assumed to be given (e.g., Geiger et al. 2021; Geiger et al. 2018). Consequently, to elicit faking, researchers typically focus on manipulating participants' motivation to fake.

### 3.2 | The PSS as a Solution

To manipulate faking motivation, researchers often generate settings in which participants are instructed to answer the items of the to-be-faked measures as if they were applying for their dream job and to respond in a manner that will maximize their chances of getting hired (e.g., Davison et al. 2021; Harris et al. 2021). Such conditions (i.e., the faking conditions) are then compared with conditions in which participants are, for example, asked to respond honestly (i.e., the control conditions). The mean differences between these conditions are then used to estimate the degree of faking (e.g., Bensch et al. 2019; Birkeland et al. 2006; Griffith et al. 2007; Hartman and Grubb 2011; Salgado and Lado 2018; Tonković et al. 2012). This approach allows researchers to manipulate faking motivation and thus to examine faking in experimental laboratory settings. We will refer to this paradigm as the PSS.

## 4 | Are the Differences Between Faking and Control Conditions Elicited Only by Faking?

The abovementioned approach appears straightforward. Nevertheless, the differences between faking and control conditions might be elicited not only by the motivation to fake but also by the different framing of item content, an idea suggested in previous research. For example, in their meta-analysis, Birkeland et al. (2006) investigated differences between applicants and nonapplicants and suggested that “an alternative explanation for higher applicant mean scores concerns item context (Robie et al. 2000; Schmit et al. 1995)” (p. 318; see also, e.g., Blickle et al. 2009).<sup>1</sup> In their argument, Birkeland et al. (2006) referred to Robie (2001), who suggested that “nonapplicants might consider their behavior across a *wide range of life contexts* [emphasis added] when responding to a personality measure (i.e., How do I behave in general?), whereas applicants may complete the measure focused specifically *on how they behave at work* [emphasis added] (i.e., How do I behave at work?). This frame-of-reference difference might cause applicants to score higher on a personality measure due to the finding that individuals who take a personality measure that is composed of items contextualized to a work environment obtain higher scores than individuals taking a personality test that is composed of general, noncontextualized items” (Birkeland et al. 2006, p. 318).

On the basis of the results of another study, Robie et al. (2001) also concluded that participants who are asked to respond to a self-report personality measure as if they are applying for a job may mentally form a work FOR and respond accordingly. Thus, for over 20 years, several researchers have argued that the FOR

may impact studies on faking. Still, empirical investigations of this conjecture are very rare.

## 5 | The FOR

### 5.1 | Defining the FOR

The FOR represents the (predominant) context that participants consider when responding to items (Schmit et al. 1995). For example, participants who pursue their hobby of stamp collecting and work conscientiously to expand their collection (private context) might not be that conscientious at work. When using a self-report personality measure (e.g., in an application setting) to assess such a participant's conscientiousness, it is relevant to know whether or not the participant predominantly adopts the FOR “at work.” Depending on which FOR is predominant in a participant's mind, different responses may be given, each of which may be true but may provide information about the participant's behavior in different contexts. In this case, changes in the participant's responses on the self-report personality measure result from the adoption of a different dominant FOR and should not be considered faking but “honest” self-presentation (e.g., Robie et al. 2001).

### 5.2 | How the FOR Can Be Influenced

The adoption of a certain FOR can be promoted by manipulating the instructions or by manipulating the items (i.e., instruction-level vs. item-level contextualization). Using the target of increasing the work FOR as an example, the instructions can be manipulated by adding the context to them (e.g., “Remember, think about how you are AT WORK in general when responding to these questions”; see p. 547, Hunthausen et al. 2003), whereas the items can be manipulated either by adding the context to the original item as a tag (e.g., “at work”; e.g., “I pay attention to details at work,” see p. 268, Lievens et al. 2008) or by fully contextualizing the item (e.g., by using “I am an ordinary employee who is no better than others” instead of “I am an ordinary person who is no better than others”; Holtrop et al. 2014).

Research has indicated that when it comes to the adoption of a certain FOR, the effect of item-level contextualization exceeds the one of instruction-level contextualization, which is why the use of the former over the latter has been increasingly recommended for research in recent years (e.g., Holtrop et al. 2014; Lievens et al. 2008; Schlotzhauer et al. 2024; Swift and Peterson 2019).<sup>2</sup> In addition, fully contextualized items have been found to elicit more positive reactions from participants than tagged items and have therefore become increasingly preferred over their tagged counterparts (Holtrop et al. 2014; Holtrop et al. 2014; but see also Holtz et al. 2005).

### 5.3 | How the Adoption of the FOR Changes

Research has demonstrated that noncontextualized items elicit the adoption of different FORs, whereas contextualized items

lead participants to adopt the context-related FOR to a greater degree (e.g., Fisher et al. 2017). For example, Fisher et al. (2017) demonstrated that applicants reported that they adopted a work FOR to a greater degree when items were contextualized with a work FOR ( $M = 4.62$  across samples) than when items were noncontextualized ( $M = 4.46$  across samples;  $d = 0.24$  across samples).<sup>3</sup>

Research has not yet investigated whether the adoption of a work FOR also differs in samples that are asked to fake their responses within the context of a PSS (in experimental laboratory faking research). Based on the presented results on applicant samples, it is very plausible that participants who are asked to fake in accordance with the PSS will adopt a work FOR to a greater degree when the items are contextualized with a work FOR than when the items are noncontextualized (e.g., Fisher et al. 2017). In addition, according to previous research, the adoption of a work FOR should be stronger when the work contextualization is realized via a (full) item-level contextualization rather than via a mere manipulation of instructions (see Holtrop et al. 2014; Lievens et al. 2008; Schlotzhauer et al. 2024; Swift and Peterson 2019).

5.4 | How the FOR Changes Scale Mean Scores on Self-Report Personality Measures

Many studies have shown that participants' responses differ depending on which FOR is most salient (e.g., Bowling and Burns 2010; Fisher et al. 2017; Holtrop et al. 2014; Pathki et al. 2022). For example, incumbents who took a self-report personality measure that was contextualized with a work FOR obtained scores that may be considered more positive (e.g., higher scores on Conscientiousness and lower scores on Emotionality) than when they took a general, noncontextualized one (e.g., Bowling and Burns 2010; Holtrop et al. 2014; Pathki et al. 2022).<sup>4</sup> Not only were such scale mean score differences found in incumbent samples in which participants were most likely not motivated to fake, but they were also found in applicant samples in which participants may have been motivated to fake. For example, applicants who took a self-report personality measure that was contextualized with a work FOR scored higher on Conscientiousness than those who took a noncontextualized one (although differences in Conscientiousness scores were nonsignificant in three out of four samples;

Fisher et al. 2017).<sup>5</sup> There were no significant changes in Honesty-Humility (Holtrop et al. 2014).

These findings demonstrate that the FOR can account for scale mean score differences and that these differences vary depending on the personality trait under investigation (e.g., higher scores in Conscientiousness vs. lower scores in Emotionality vs. no significant changes in Honesty-Humility). Thus, it is important to test and compare the results across different personality traits.

6 | How FOR and Faking May Be Confounded

It has been suggested that the ways in which instructions are typically manipulated in experimental laboratory faking research (i.e., generic instructions, GIs vs. PSS) may be associated with the adoption of a work FOR (see Birkeland et al. 2006; Robie 2001; Robie et al. 2001). Although the work FOR that is promoted by the manipulation of instructions is comparably weaker than the work FOR that is promoted by (full) item-level contextualization (e.g., Holtrop et al. 2014; Lievens et al. 2008; Schlotzhauer et al. 2024; Swift and Peterson 2019), this potential confounding of FOR and faking is not trivial: The increased adoption of a work FOR in the PSS (compared with GIs) may bias the results (e.g., scale mean score changes on self-report personality measures) so that they are falsely attributed to faking only (i.e., internal validity is impaired).

Knowing that (full) item-level contextualization can also be used to promote the adoption of a work FOR (Fisher et al. 2017) and that the adoption of a work FOR through this route is stronger than through mere manipulation of instructions, combining the two approaches should disentangle the effects of FOR and faking because it allows the impact of FOR to be isolated from the impact of faking (see Table 1).

7 | Finding Common Ground

To summarize, there are two processes (i.e., FOR and faking) that can be influenced by manipulating the instructions on a self-report personality measure, and the manipulations that are typically used in experimental laboratory faking research (e.g., faking condition: "Imagine you are applying for a job; answer

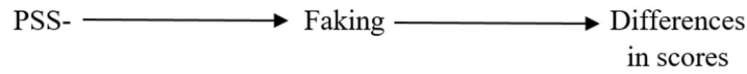
TABLE 1 | Factors that may impact participants responses in self-report personality measures.

Instructions	Full item-level work contextualization		Factor(s) for differences between conditions
	–	+	
GIs	GIs–	GIs+	Absence (–) versus presence (+) of full item-level work contextualization in GIs
PSS	PSS–	PSS+	Absence (–) versus presence (+) of full item-level work contextualization in PSS
Factor(s) for differences between conditions	Faking and FOR in GIs versus PSS	Faking	

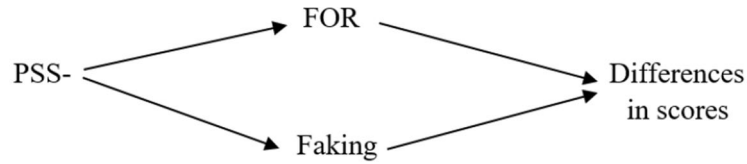
Note: GIs– (generic instructions, full item-level work contextualization absent), GIs+ (generic instructions, full item-level work contextualization present), PSS– (personnel-selection scenario, full item-level work contextualization absent), PSS+ (personnel-selection scenario, full item-level work contextualization present).



## Interpretation



## Process



**FIGURE 1** | Previous Interpretations and Assumed Process in Experimental Laboratory Faking Research Designs. *Note:* PSS– (personnel-selection scenario, full item-level work contextualization absent).

the items to increase your chances of being hired” vs. control condition: “Answer the items honestly”) might therefore do more than just induce faking in the faking condition and “honest” responding in the control condition. They might also elicit a work FOR to different degrees in the two conditions (see also Anglim et al. 2021; Bing et al. 2004; Birkeland et al. 2006; Robie 2001; Robie et al. 2001). Thus, in the PSS paradigm that is typically used in experimental laboratory faking research, FOR and faking may be confounded (see Figure 1). If so, this undermines internal validity and calls into question previous conclusions about faking as the only factor that accounts for changes in participants’ responses. To the best of our knowledge, this idea has often been stated (e.g., Birkeland et al. 2006; Robie 2001; Robie et al. 2001) but has yet to be empirically investigated. To do so and to determine the “true” extent to which faking is elicited by the PSS, FOR effects need to be controlled for (Table 1).

## 8 | The Present Study

We argue that FOR and faking effects are both elicited by the PSS and that both impact the differences between GIs and the PSS (see Table 1; Figure 1). Thus, due to potentially confounding FOR effects (as elicited by the PSS), the effects of faking may have been incorrectly estimated in previous research. By using fully work-contextualized items, we investigated faking while controlling for the FOR. To do so, we varied both the instructions (i.e., GIs vs. PSS) and the full item-level work contextualization (i.e., absent vs. present). We examined the effects of instructions and full item-level work contextualization on the adopted FOR (i.e., mean changes in the adoption of a work FOR) and on response behavior (i.e., mean changes in a self-report personality measure assessing Conscientiousness, Emotionality, and Honesty-Humility). We decided to investigate these traits because they are the three personality dimensions that are most predictive of behavior at work (e.g., De Vries et al. 2014; Holtrop et al. 2014; Lee et al. 2005) and because a self-report personality measure with full item-level work contextualization is available for them (Holtrop et al. 2014).<sup>6</sup> We focused on changes in scale mean scores on a self-report personality measure because FOR effects and faking effects have both typically been evaluated using mean score changes in these measures (e.g., Shaffer and Postlethwaite 2012; Ziegler et al. 2012). On the basis of the arguments presented above, we derived the following hypotheses.

**Hypothesis 1 (H1; changes in the adoption of a work FOR):** In line with assumptions from previous research (e.g., Birkeland et al. 2006; Holtrop et al. 2014; Robie et al. 2001), we expected participants to report that they adopted a work FOR to a greater degree when a PSS was presented than when GIs were presented; and in line with research by Fisher et al. (2017), we expected participants to report that they adopted a work FOR to a greater degree when work-contextualized items were used than when noncontextualized items were used (i.e., full item-level work contextualization present vs. absent). Because the impact of instructions has been shown to be weaker than the impact of item-level contextualization (e.g., Holtrop et al. 2014; Lievens et al. 2008; Schlotzhauer et al. 2024; Swift and Peterson 2019), and because with a fully work-contextualized measure, one might expect an adoption of the work FOR with each item, we expected a pattern similar to the one plotted in A1 (see Appendix). Thus, we expected both of the main effects, that is, the main effect of instructions (GIs vs. PSS) and the main effect of full item-level work contextualization (absent [–] vs. present [+]) on the adoption of a work FOR, to be significant but qualified by a significant interaction for each personality trait under investigation (i.e., Conscientiousness, Emotionality, and Honesty-Humility).

**Hypothesis 2 (H2; changes in response behavior):** Both manipulations of instructions (i.e., PSS vs. GIs; e.g., Birkeland et al. 2006) and items (i.e., work-contextualized items vs. noncontextualized items; e.g., Bowling and Burns 2010; Pathki et al. 2022; Schmit et al. 1995; Shaffer and Postlethwaite 2012) have been found to have main effects on the scale mean scores of self-report personality measures. We expected to replicate these findings and extend them by revealing a significant interaction: We hypothesized that the effect of instructions when comparing the conditions without full item-level work contextualization (i.e., GIs– vs. PSS–) would be elicited by both the instruction-inherent FOR and faking (e.g., Birkeland et al. 2006; Robie 2001; Robie et al. 2001). In other words, only when comparing the conditions with full item-level work contextualization (i.e., GIs+ vs. PSS+) should the effect of instructions be elicited by faking without FOR effects biasing the results (Table 1; Figure 1). This would increase the effect of instructions when full item-level work contextualization is absent (i.e., FOR and faking) compared with when it is present (i.e., faking only). Thus, we expected both of the main effects, that is, the main effect of instructions (GIs vs. PSS) and the main effect of full item-level work contextualization (absent

[−] vs. present [+]) on the scale mean scores of the self-report personality measure, to be significant but qualified by a significant interaction (see A2 in the Appendix) for each personality trait under investigation (i.e., Conscientiousness, Emotionality, and Honesty-Humility).

## 9 | Pretest

To test the suitability of the self-report personality measure for our study a priori, we followed the procedure described by Robie et al. (2000) and Lievens et al. (2008). Thus, we administered a pretest to assess whether the participants adopted the different FORs (see below) to different degrees when responding to the self-report personality measure in a noncontextualized condition (i.e., GIs−).

### 9.1 | Methods

#### 9.1.1 | Sample

We recruited an online community sample ( $N = 53$ ) via Biliendi. Two participants spent less than 2 s per item and were thereby excluded from further analyses (i.e., careless responding; e.g., Edwards 2019; Huang et al. 2012). Thus, the final sample consisted of 51 participants (25 women, 25 men, 1 diverse/no response) with an average age of 45.02 years ( $SD = 14.04$ ). Their main occupations were: computer science/related occupations (11.32%); business management/organizational occupations, purchasing/sales/trading occupations (each 9.43%); legal/administrative occupations (5.66%); cleaning occupations, financial services occupations, mechanical/energy-related/electrical occupations, medical care occupations, security occupations, teaching occupations (each 3.77%); social/theological occupations, food manufacturing/processing occupations, nonmedical health/personal care/medical technology occupations, transportation occupations, entertainment occupations, plastic/wood manufacturing/processing occupations, sales occupations, development/design/production control occupations, hospitality occupations, logistics occupations (each 1.87%), and other (18.87%).<sup>7</sup>

#### 9.1.2 | Procedure

Participants completed the noncontextualized version of the self-report personality measure and rated the extent to which the items that they had responded to elicited the thought of specific contexts (adopted FORs; see below). Thus, for our pretest analyses, we assessed participants' demographics (see A3 in the Appendix), their adopted FORs, and their responses to the noncontextualized self-report personality measure.

#### 9.1.3 | Measures

**Adopted FORs.** The adoption of FORs was measured with four items translated and adapted from Fisher et al. (2017).<sup>8</sup> After

each page of the self-report personality measure (i.e., after every four items out of a total of 32 items that measured each personality trait), participants were asked to indicate the extent to which they agreed or disagreed (1 = *strongly disagree* to 5 = *strongly agree*) that they had thought of each of the following specific contexts (i.e., educational, social, work, faking; see A4 in the Appendix) while responding to the items. The contexts that we used were largely in line with Fisher et al. (2017) and represent the typical FORs for personality items used in previous theorizing and research (Davison and Bing 2009; Hunthausen et al. 2003; Lievens et al. 2008).

**The Self-Report Personality Measure.** Participants completed the Conscientiousness, Emotionality, and Honesty-Humility scales from the German HEXACO-PI-R (Lee and Ashton 2004; Lee et al. 2009; see A5 in the Appendix), which corresponds to the noncontextualized Dutch version of the measure that Holtrop et al. (2014) used. The scales were presented to participants in a randomized order. Each scale consisted of 32 items that were rated on 5-point Likert scales ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). All scales (Conscientiousness:  $M = 3.76$ ,  $SD = 0.48$ , Emotionality:  $M = 3.01$ ,  $SD = 0.57$ , and Honest-Humility:  $M = 3.75$ ,  $SD = 0.57$ ) showed high levels of reliability with Cronbach's alphas of 0.90, 0.92, and 0.91, respectively.

## 9.2 | Results

To analyze whether participants differed in their adoption of the FORs, we computed the means of the eight ratings for each FOR and each personality trait. Thus, we had 51 raters ( $k = 51$ ; i.e., participants) and four FORs ( $n = 4$ ; i.e., the means of the eight ratings for each FOR), which we analyzed for each of the three personality traits. The results of our pretest were consistent with previous research (e.g., Schulze et al. 2021) and supported our assumption that the GIs−condition would elicit the four FORs to different extents in our participants. The scale-specific ICCs were all  $< 0.50$  (see Table 2), indicating poor interrater reliability (Koo and Li 2016). A total of 19% to 25% of the variance in ratings was attributable to actual differences between the adopted FORs, and 75% to 81% of the variance was attributable to differences between the participants (and other unsystematic sources of variance). Thus, there were substantial differences regarding the adopted FORs between participants, and a necessary precondition for our study was met.<sup>9</sup>

## 10 | Main Study

Our main study included four waves. We investigated our hypotheses on the changes in the adoption of a work FOR (H1) and the changes in response behavior (i.e., changes in scale mean scores on the self-report personality measure; H2). The independent variables were the instructions (GIs vs. PSS) and the full item-level work contextualization (absent vs. present). The dependent variables were the adoption of a work FOR versus the mean scores on Conscientiousness, Emotionality, and Honesty-Humility.

**TABLE 2** | Intraclass correlation coefficients for all FORs under GIs–.

	<i>ICC</i>	<i>F</i>	<i>df1</i>	<i>df2</i>	95% CI	
					<i>LL</i>	<i>UL</i>
Conscientiousness	0.24	29.64	3	150	0.08	0.82
Emotionality	0.25	33.77	3	150	0.09	0.83
Honesty-Humility	0.19	27.03	3	150	0.06	0.77

Note: *ICC* = *ICC*(2,1) according to Shrout and Fleiss (1979), *n* = 4 (i.e., adopted FORs: educational, social, work, faking), *k* = 51 (i.e., participants). CI = confidence interval; LL = lower limit; UL = upper limit.

## 10.1 | Methods

### 10.1.1 | Power Analyses

For an alpha level of 0.008, an a priori power analysis computed with G\*Power 3.1.9.6 (Faul et al. 2009) revealed that the sample size required to detect a small effect size (partial  $\eta^2 = 0.01$ ) with a power of 0.80 for the two-way repeated-measures ANOVAs was *N* = 203.<sup>10</sup> Considering potential outliers, we planned to oversample and targeted a minimum sample size of *N* = 250 in Wave 4.

On the basis of previous research (e.g., Davison et al. 2021; Lee et al. 2019; Liu and Zhang 2020; Roess and Roche 2017; Terry et al. 1999), we expected a dropout rate of approximately 50%. Moreover, we expected that approximately 20% of the participants would not meet the inclusion criteria (see Appendix A7). We included an additional safety margin. Thus, we planned to oversample and increase the number of participants necessary for the first wave of measurement, aiming to collect data from at least *N* = 700.

### 10.1.2 | Sample

**Data Exclusion.** For the first wave, we recruited an online community sample (*N* = 3661) via Bilendi.<sup>11</sup> On the basis of our inclusion criteria and the testing of data assumption requirements, 3352 participants were excluded from further analyses. Of those excluded, 1810 did not complete all four waves of measurement, 1245 responded incorrectly to at least one attention check (i.e., careless responders), 42 indicated a lack of motivation (i.e., rated *neutral*, *disagree*, or *strongly disagree* when asked whether they were motivated), 18 indicated that they did not participate in a quiet environment (i.e., rated *neutral*, *disagree*, or *strongly disagree* when asked whether they had participated in a quiet environment), 100 indicated that they could not identify with the PSS (i.e., rated *neutral*, *disagree*, or *strongly disagree* when asked for their ability to identify with the PSS), and 71 failed the predefined markers for nonblinding (see Table 3). Finally, outliers on the dependent variables were excluded from further analyses (i.e., exclusion of 66 participants; see A8 and A9 in the Appendix).

**Final Sample.** Thus, the final sample consisted of 309 participants (98 women, 211 men) with an average age of 47.90 years (*SD* = 13.10). Their main occupations were: business management/organizational occupations (17.15%); computer science/related occupations (8.09%); mechanical/

energy-related/electrical occupations (7.77%); financial services occupations, purchasing/sales/trading occupations (each 4.85%); teaching occupations, social/theological occupations (each 3.88%); security occupations, medical care occupations (each 3.24%); legal/administrative occupations, advertising/marketing occupations (each 2.54%); hospitality occupations (2.27%); sales occupations (1.94%); transportation occupations, logistics occupations, metal manufacturing/processing occupations, building/supply engineering occupations, mechanical/automotive engineering occupations (each 1.29%); nonmedical health/personal care/medical technology occupations, math/biology/chemistry/physics occupations (each 0.97%); construction planning/architecture/surveying occupations, horticultural occupations/floristry, cleaning occupations, food manufacturing/processing occupations, (interior) finishing occupation (each 0.65%); building construction/civil engineering occupations, entertainment occupations, paper/printing occupations/technical media design (each 0.32%), and other (21.04%).<sup>12</sup>

### 10.1.3 | Procedure

We addressed the research questions by conducting a four-wave longitudinal study with a 2 × 2 repeated-measures design using an interval of approximately 2 weeks between measurements. We applied this approach to assess participants' faking (e.g., Brown and Böckenholt 2022) and their adopted FORs (e.g., König et al. 2012; Lievens et al. 2008; Robie et al. 2000) while simultaneously trying to minimize carry-over effects (e.g., Burns and Christiansen 2006; Davison et al. 2021) and demand characteristics (e.g., Orne 1962; see also Corneille and Lush 2023). Our procedure is shown in Table 4.

To vary the adoption of a work FOR independent of faking, we used full item-level work contextualization as a manipulation, an approach that has been used successfully in past research (e.g., Robie et al. 2017). Thus, the instructions (i.e., GIs vs. PSS) and the full item-level work contextualization (i.e., absent [–] vs. present [+]) were manipulated, resulting in four within-person conditions (i.e., GIs–, GIs+, PSS–, PSS+; see Table 1). To avoid sequence effects, we fully counterbalanced our conditions (see Table 4), thus following a procedure that many previous studies have employed (e.g., Bing et al. 2004; Bowling and Burns 2010; Lievens et al. 2008; Robie et al. 2001).

**Manipulation of the Instructions.** To elicit faking and to manipulate the adoption of a work FOR at the level of the



**TABLE 3** | Coding of participants' blinding to the purpose of the study.

Code	Assumed purpose
0 ( <i>blind</i> ; i.e., no markers mentioned)	<ul style="list-style-type: none"> <li>• None, no idea, empty textbox, etc.</li> <li>• Assumed purposes that do not fall in the categories below</li> </ul>
1 ( <i>not blind</i> ; i.e., one or more markers mentioned)	<ul style="list-style-type: none"> <li>• Consistency of responding, stability of personality, change across time</li> <li>• Behavioral differences between contexts, change across contexts, adaptation</li> <li>• Faking, socially desirable responding, lying, detection of lies, impression management, distortion of answers, response distortion, biased self-presentation</li> </ul>

instructions, participants were instructed either to respond honestly (i.e., GIs) or to fake their responses while imagining a personnel-selection scenario (i.e., PSS). The manipulation was applied within participants.

*The GIs read: "On the following pages, you will find a series of statements about you. Please read each statement and decide the extent to which you agree or disagree with that statement. Please answer every statement, even if you are not completely sure about your response. Please be as honest as possible."*

*The PSS read: "You are applying for a very popular and well-paid position as a pharmacy assistant. This has always been your dream job. As part of the personnel selection, you are asked to respond to a measure that will assess your personal characteristics. You will be asked to indicate the extent to which you agree with several statements. Your responses will influence whether or not you are offered the position. To get your dream job, you have to make the impression that you are very emotionally stable, very conscientious, and very honest and service-oriented. Please answer the following questionnaire in a way that will maximize your chances of getting the job."*

To ensure that all participants had the same knowledge about the pharmacy assistant job, we provided a description of this job along with the PSS (see Appendix A10 for the description).

**Manipulation of the Full Item-Level Work Contextualization.** To manipulate the adoption of a work FOR at the item level, participants responded to either the noncontextualized or the fully work-contextualized version of the self-report personality measure (i.e., full item-level work contextualization absent vs. present; see below). Again, the manipulation was applied within participants.

Within each condition, the scales were administered to participants in a randomized order. Each participant was presented with both versions of the self-report personality measure (i.e., noncontextualized and fully work-contextualized), once with GIs and once with the PSS.

#### 10.1.4 | Measures

**Adopted Frames of Reference.** The adoption of FORs was assessed as in our pretest (see above).

**Versions of the Self-Report Personality Measure.** We included two versions of the self-report personality measure, each measuring Conscientiousness, Emotionality, and Honesty-Humility. In the noncontextualized version, the full item-level work contextualization was absent (i.e., the respective scales from the German HEXACO-PI-R; Lee and Ashton 2004; Lee et al. 2009; see A5 in the Appendix). In the fully work-contextualized version, the full item-level work contextualization was present. This version was an adaptation of the Conscientiousness, Emotionality, and Honesty-Humility scales from the HEXACO-PI-R, involving full item-level work contextualization, kindly provided to us by Djurre Holtrop (Holtrop et al. 2014; see A11 in the Appendix). We used a German version of this measure because our sample comprised native German speakers. As on our pretest, the scales were administered to participants in a randomized order.

#### 10.1.5 | Analytical Approach

We computed our analyses in R (version 4.4.0; R Core Team 2024) with the packages *boot* (Canty and Ripley 2024), *car* (Fox and Weisberg 2019), *effectsize* (Ben-Shachar et al. 2020), *ez* (Lawrence 2016), *fs* (Hester et al. 2024), *furrr* (Vaughan and Dancho 2022), *ggExtra* (Attali and Baker 2023), *ggh4x* (Van den Brand 2024), *ggrepel* (Slowikowski 2024), *gt* (Iannone et al. 2024), *haven* (Wickham et al. 2023), *patchwork* (Pedersen 2024), *psych* (Revelle 2023), *rstatix* (Kassambara 2023), *scales* (Wickham et al. 2023), *skimr* (Waring et al. 2022), *tidyverse* (Wickham et al. 2019), and *xlsx* (Dragulescu and Arendt 2020). The R scripts were made available online before the final analysis.

**Descriptive Statistics.** Descriptive statistics were computed on the demographic information and dependent variables. To describe and plot the sample distributions and correlations, tables and boxplots were used.

**Test of Requirements.** The requirements for the preregistered two-way repeated-measures ANOVAs included a normal distribution of scores on the dependent variable in each wave of measurement, homogeneity of variance, sphericity, and the

**TABLE 4** | Study procedure.

Wave 1	Wave 2	Wave 3	Wave 4
GIs–	GIs+	PSS–	PSS+
		PSS+	PSS–
		GIs+	PSS+
	PSS–	PSS+	GIs+
		GIs+	PSS–
		PSS–	GIs+
GIs+	GIs–	PSS–	PSS+
		PSS+	PSS–
		GIs–	PSS+
	PSS–	PSS+	GIs–
		GIs–	PSS–
		PSS–	GIs–
PSS–	GIs–	GIs+	PSS+
		PSS+	GIs+
		GIs–	PSS+
	GIs+	PSS+	GIs–
		GIs–	GIs+
		GIs+	GIs–
PSS+	GIs–	GIs+	PSS–
		PSS–	GIs+
		GIs–	PSS–
	GIs+	PSS–	GIs–
		GIs–	GIs+
		GIs+	GIs–

*Note:* Demographics were assessed in Wave 1 only. Inclusion criteria were assessed in Wave 4 only. The adopted FORs and the response behaviors on the self-report personality measures were assessed in each wave. GIs– (generic instructions, full item-level work contextualization absent), GIs+ (generic instructions, full item-level work contextualization present), PSS– (personnel-selection scenario, full item-level work contextualization absent), PSS+ (personnel-selection scenario, full item-level work contextualization present).

nonexistence of outliers. These requirements were tested with histograms, pp-plots, qq-plots, Levene's test, the Mauchly test (if it had been necessary, we would have used the Huynh-Feldt and Greenhouse-Geisser corrections), and boxplots before running the analyses.

**Hypothesis Testing.** We computed two-way repeated-measures ANOVAs to analyze main and interaction effects for Hypotheses 1 and 2. In each ANOVA, the independent variables (i.e., instructions: GIs vs. PSS and full item-level work contextualization absent vs. present) were entered as dummy-coded main effects, including an interaction term.

**Open Practices.** This is a Registered Report. Thus, the theory, hypotheses, method, and R scripts were preregistered and peer-reviewed before the analyses. All materials, including the analysis scripts, (translated) measures, R codes, data sets, supplemental material, and the Stage 1 protocol are available on the OSF (<https://osf.io/57n4e/>). All analyses were either preregistered or marked as exploratory. We highlighted all

deviations from the preregistration and the analysis scripts. We encourage other researchers to use our data and materials or to conduct different replication studies to clarify the findings reported in this study.

## 10.2 | Results

### 10.2.1 | Descriptive Statistics

**10.2.1.1 | Scale Mean Scores, Standard Deviations, and Cronbach's Alphas. Adopted Work FOR.** Mean scores, standard deviations, and Cronbach's alphas for the adopted work FOR across the four conditions (i.e., GIs–, GIs+, PSS–, PSS+) and for the personality traits under investigation are presented in Table 5 (upper half). In line with the descriptive statistics presented by Fisher et al. (2017), and irrespective of the personality trait that was investigated, the mean scores for the adoption of a work FOR increased when full item-level work contextualization was present (GIs+ and PSS+) compared with when it was absent (GIs– and PSS–).<sup>13</sup> The respective effect sizes of the main effects of full item-level work contextualization (absent vs. present) were moderate (Conscientiousness) to large (Emotionality and Honesty-Humility; see Table 6, upper half). No previous studies have computed descriptive statistics for the adopted work FOR when a PSS was presented compared with when GIs were presented. Cronbach's alpha was excellent in all conditions.

**Self-Report Personality Measure.** Mean scores, standard deviations, and Cronbach's alphas across the four conditions (i.e., GIs–, GIs+, PSS–, PSS+) and for each personality trait under investigation are presented in Table 5 (lower half). Concerning GIs–, scale score characteristics and Cronbach's alpha reliabilities for the personality trait that was investigated were comparable to Lee and Ashton's (2004) values (Conscientiousness:  $M = 3.32$ ,  $SD = 0.51$ ,  $\alpha = 0.89$ , Emotionality:  $M = 3.20$ ,  $SD = 0.55$ ,  $\alpha = 0.90$ , and Honesty-Humility:  $M = 3.36$ ,  $SD = 0.60$ ,  $\alpha = 0.92$ ). Concerning GIs+, they were largely comparable to Holtrop et al. (2014) values (Conscientiousness:  $M = 3.72$ ,  $SD = 0.30$ ,  $\alpha = 0.81$ , Emotionality:  $M = 2.87$ ,  $SD = 0.32$ ,  $\alpha = 0.81$ , and Honesty-Humility:  $M = 3.93$ ,  $SD = 0.29$ ,  $\alpha = 0.82$ ).<sup>14</sup> The PSS– and PSS+ led to typical effects demonstrated in faking research (e.g., Salgado 2016): The means changed in line with the faking directions, and with the exception of Honesty-Humility, the standard deviations and Cronbach's alphas increased.<sup>15</sup>

### 10.2.2 | Correlations

For correlations between the adoption of a work FOR and the scale mean scores of the self-report personality measure (see A12 in the Appendix), we computed Spearman rank-order correlations because participants' responses on the adoption of a work FOR were not normally distributed (see A13 to A15 in the Appendix).

The correlations showed patterns that were in line with both previous theorizing and our expectations. First, regardless of the personality trait under investigation (i.e., Conscientiousness, Emotionality, and Honesty-Humility), the adoption of a work

TABLE 5 | Mean scores, standard deviations, and Cronbach's alphas for the adopted work FOR as well as for the self-report personality measure by experimental condition and personality trait.

Scale	GIs–		GIs+		PSS–		PSS+	
	M (SD)	Cronbach's $\alpha$ [95% CI]	M (SD)	Cronbach's $\alpha$ [95% CI]	M (SD)	Cronbach's $\alpha$ [95% CI]	M (SD)	Cronbach's $\alpha$ [95% CI]
<b>Adopted work FOR</b>								
Conscientiousness	3.98 (0.72)	0.95 [0.94, 0.96]	4.10 (0.74)	0.95 [0.95, 0.96]	4.06 (0.71)	0.95 [0.94, 0.96]	4.20 (0.72)	0.95 [0.95, 0.96]
Emotionality	3.47 (0.95)	0.96 [0.95, 0.96]	4.08 (0.71)	0.95 [0.94, 0.96]	3.61 (0.85)	0.95 [0.94, 0.95]	4.13 (0.68)	0.96 [0.95, 0.96]
Honesty-Humility	3.72 (0.83)	0.94 [0.93, 0.95]	4.04 (0.72)	0.94 [0.93, 0.95]	3.80 (0.78)	0.94 [0.92, 0.95]	4.14 (0.72)	0.95 [0.94, 0.96]
<b>Self-report personality measure</b>								
Conscientiousness	3.73 (0.48)	0.90 [0.88, 0.91]	3.91 (0.47)	0.91 [0.88, 0.92]	3.88 (0.49)	0.92 [0.90, 0.93]	4.08 (0.47)	0.92 [0.90, 0.93]
Emotionality	3.12 (0.43)	0.85 [0.82, 0.87]	2.86 (0.36)	0.78 [0.74, 0.81]	3.04 (0.46)	0.87 [0.85, 0.89]	2.77 (0.37)	0.80 [0.76, 0.83]
Honesty-Humility	3.74 (0.59)	0.92 [0.91, 0.93]	3.61 (0.50)	0.89 [0.87, 0.91]	3.77 (0.58)	0.92 [0.91, 0.93]	3.67 (0.48)	0.88 [0.86, 0.90]

Note: GIs– (generic instructions, full item-level work contextualization absent), GIs+ (generic instructions, full item-level work contextualization present), PSS– (personnel-selection scenario, full item-level work contextualization absent), PSS+ (personnel-selection scenario, full item-level work contextualization present). N = 309. CI = confidence interval.

**TABLE 6** | Results of the two-way repeated-measures ANOVAs on the adoption of a work FOR and on the mean scores of the self-report personality measure for each personality trait.

Scale	Effect	<i>F</i> (1, 308)	<i>p</i>	Partial $\eta^2$	95% CI (bootstrapped)	
					LL	UL
<b>Adoption of a work FOR</b>						
Conscientiousness	I	11.97	0.001	0.04	0.01	0.09
	C	20.20	< 0.001	0.06	0.02	0.12
	I $\times$ C	0.18	0.668	0.00	0.00	0.01
Emotionality	I	9.99	0.002	0.03	0.01	0.08
	C	160.86	< 0.001	0.34	0.28	0.40
	I $\times$ C	2.69	0.102	0.01	0.00	0.04
Honesty-Humility	I	11.77	0.001	0.04	0.01	0.09
	C	83.29	< 0.001	0.21	0.15	0.27
	I $\times$ C	0.09	0.766	0.00	0.00	0.00
<b>Mean scores of the self-report personality measure</b>						
Conscientiousness	I	103.65	< 0.001	0.25	0.18	0.32
	C	103.36	< 0.001	0.25	0.17	0.33
	I $\times$ C	0.35	0.555	0.00	0.00	0.01
Emotionality	I	46.73	< 0.001	0.13	0.07	0.20
	C	226.29	< 0.001	0.42	0.36	0.49
	I $\times$ C	0.64	0.424	0.00	0.00	0.02
Honesty-Humility	I	13.76	< 0.001	0.04	0.01	0.09
	C	43.12	< 0.001	0.12	0.06	0.19
	I $\times$ C	1.64	0.201	0.01	0.00	0.03

Note: *N* = 309. I = main effect of instructions (GIs vs. PSS). C = main effect of full item-level work contextualization (absent vs. present). I  $\times$  C = interaction effect. CI = confidence interval; LL = lower limit; UL = upper limit.

FOR between these personality traits was strongly positively correlated within a particular condition (i.e., GIs–, GIs+, PSS–, or PSS+Is–; see the adoption of a work FOR concerning Conscientiousness, Emotionality, and Honesty-Humility for GIs– for an example). Otherwise, correlations remained positive but had more variation (small to large effect sizes). In this case, they were somewhat smaller when the presence/absence of full item-level work contextualization differed (e.g., PSS+ and GIs–) than when it was identical (e.g., PSS+ and GIs+), although this difference was less pronounced when Conscientiousness was assessed and full item-level work contextualization was absent. Thus, adopting a work FOR in a particular condition and for a particular personality trait was positively related to adopting a work FOR in other conditions, for other personality traits, or both.

Second, correlations between the personality traits within a certain condition were small to moderate and positive for Conscientiousness with Honesty-Humility, whereas for Emotionality with Honesty-Humility as well as for Emotionality with Conscientiousness, correlations were mostly negative, with small effect sizes. In addition, these correlations varied across the conditions, a finding that is congruent with the assumption that instructions and full item-level (work) contextualization impact estimates of construct validity (e.g., Lievens et al. 2008; Salgado 2016; Shaffer and Postlethwaite 2012). For example,

and as expected, the observed correlations for GIs– were similar to those reported by Lee and Ashton (2004), whereas the observed correlations for GIs+ mirrored those found by Holtrop et al. (2014). As expected, faking (PSS– and PSS+) affected these correlations (see e.g., Salgado 2016). The correlations between Conscientiousness and Honesty-Humility remained positive (small to moderate in size) when computed between the conditions. The correlations between Emotionality and Honesty-Humility as well as between Emotionality and Conscientiousness remained negative in size but reached a small effect size only when PSS+ was compared with GIs–, GIs+, or PSS–. The correlations between Emotionality and Honesty-Humility were positive (but did not reach a small effect size) when GIs– were compared with the other conditions. The correlations of the personality traits between the conditions, in each case, showed positive and strong correlations within the same personality trait (e.g., Conscientiousness with Conscientiousness). In summary, the correlations between the personality traits were as expected, and the correlations were affected, again as expected, by the conditions.

Third, the correlations between the adoption of a work FOR and the personality traits were generally small in size or did not even reach a small effect size (see e.g., Emotionality and the adoption of a work FOR). Exceptions predominantly involved the correlations between Conscientiousness and the adoption of

a work FOR (small to large effect sizes). When participants rated the adoption of a work FOR in conditions where full item-level work contextualization was absent (i.e., GIs– and PSS–), the correlations were positive and moderate in size between Conscientiousness and the adoption of a work FOR for the Conscientiousness items. They were positive but smaller in size (i.e., small to moderate) between Conscientiousness and the adoption of a work FOR for the Emotionality and Honesty-Humility items. These differences were less pronounced when full item-level work contextualization was present (i.e., GIs+ and PSS+). For Honesty-Humility and the adoption of a work FOR (Conscientiousness, Emotionality, or Honesty-Humility items), there were some small correlations that were primarily restricted to the conditions in which full item-level work contextualization was present (i.e., GIs+ and PSS+).

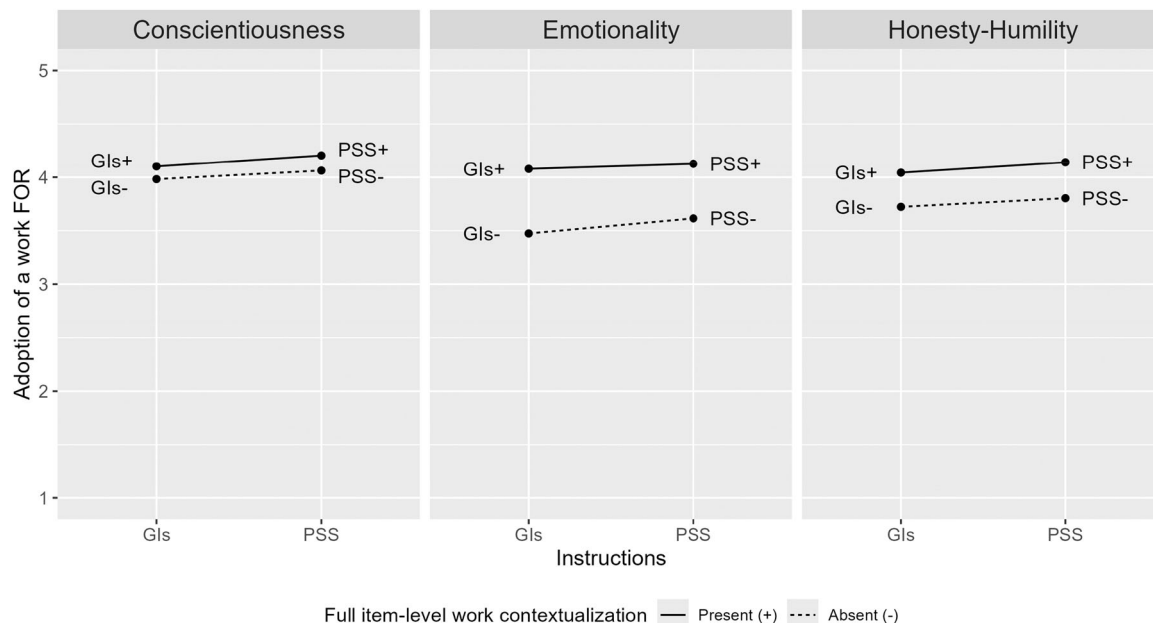
### 10.2.3 | Two-Way Repeated-Measures ANOVAs

**The Interaction Between Instructions and Full Item-Level Work Contextualization Had No Effect on the Adoption of a Work FOR.** The results of the two-way repeated-measures ANOVA (Table 6, upper half) showed significant but small main effects of instructions (GIs vs. PSS) on the adoption of a work FOR for Conscientiousness, Emotionality, and Honesty-Humility.<sup>16</sup> Thus, as hypothesized, participants reported the adoption of a work FOR to a greater degree when a PSS was presented than when GIs were presented (see Figure 2). In addition, there were moderate (Conscientiousness) to large (Emotionality and Honesty-Humility) main effects of full item-level work contextualization (absent [–] vs. present [+]). Thus, in line with our hypothesis, participants reported the adoption of a work FOR to a greater degree when using fully

work-contextualized items than when using noncontextualized items, and as expected, these main effects were stronger than the effects of instructions.

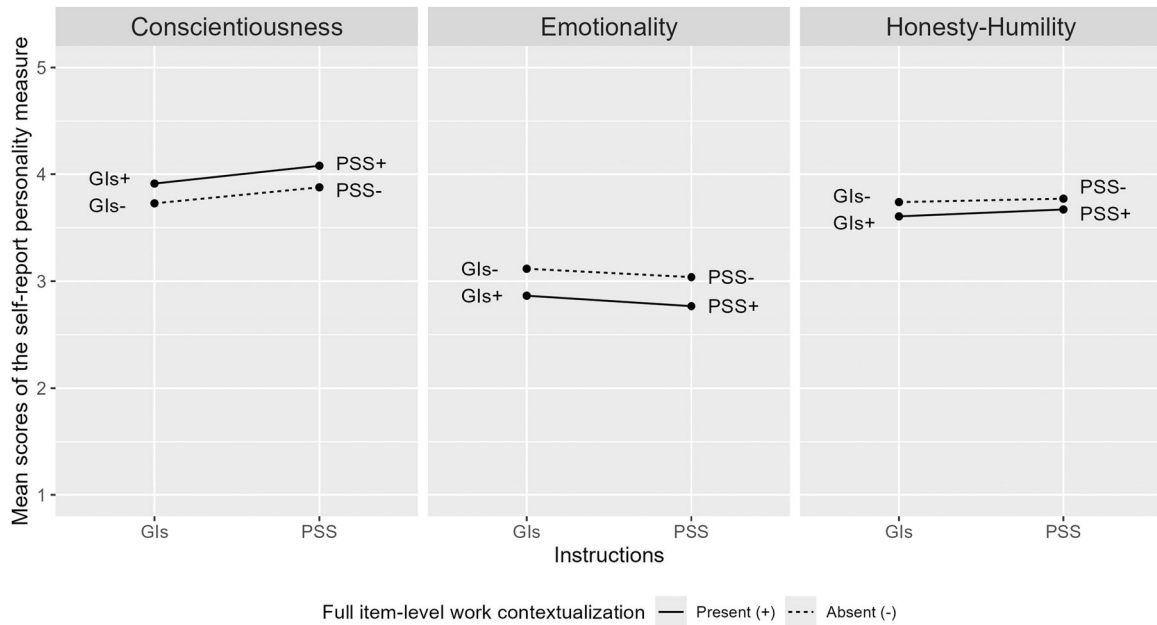
Irrespective of the personality trait under investigation (Conscientiousness, Emotionality, or Honesty-Humility), and in contrast to our hypothesis, the main effects of instructions and the main effects of full item-level work contextualization were not qualified by significant interactions. That is, the impact of instructions on the adoption of a work FOR was independent of the full item-level work contextualization, and vice versa.<sup>17</sup>

**The Interaction Between Instructions and Full Item-Level Work Contextualization Had No Effect on the Scale Mean Scores of the Self-Report Personality Measure.** The results of the two-way repeated-measures ANOVA (see Table 6, lower half) showed significant, large (Conscientiousness), moderate (Emotionality), and small (Honesty-Humility) main effects of instructions (GIs vs. PSS).<sup>18</sup> Thus, as hypothesized, compared with when GIs were presented, participants increased (Conscientiousness and Honesty-Humility) or decreased (Emotionality) their scale mean scores on the self-report personality measure when a PSS was presented (see Figure 3). In addition, there were significant and moderate (Honesty-Humility) to large (Conscientiousness and Emotionality) main effects of full item-level work contextualization (absent [–] vs. present [+]). Thus, also as hypothesized, compared with when noncontextualized items were used, participants increased (Conscientiousness) or decreased (Emotionality and Honesty-Humility) their scale mean scores on the self-report personality measure when fully work-contextualized items were used.



**FIGURE 2** | Effects of Instructions and Full Item-Level Work Contextualization on the Adoption of a Work FOR by Experimental Condition and Personality Trait. *Note:* GIs– (generic instructions, full item-level work contextualization absent), GIs+ (generic instructions, full item-level work contextualization present), PSS– (personnel-selection scenario, full item-level work contextualization absent), PSS+ (personnel-selection scenario, full item-level work contextualization present). The adoption of a work FOR was measured with a 5-point Likert scale (1 = *not at all* to 5 = *very much*).





**FIGURE 3** | Effects of Instructions and Full Item-Level Work Contextualization on the Mean Scores of the Self-Report Personality Measure by Experimental Condition and Personality Trait. *Note:* GIs- (generic instructions, full item-level work contextualization absent), GIs+ (generic instructions, full item-level work contextualization present), PSS- (personnel-selection scenario, full item-level work contextualization absent), PSS+ (personnel-selection scenario, full item-level work contextualization present). The response behavior in the self-report personality measure was measured with a 5-point Likert scale (1 = *not at all* to 5 = *very much*).

Irrespective of the personality trait under investigation and in contrast to our hypothesis, the main effects of instructions and the main effects of full item-level work contextualization were not qualified by significant interactions. That is, the impact of instructions on the scale mean scores of the self-report personality measure was independent of the full item-level work contextualization, and vice versa.<sup>19</sup>

#### 10.2.4 | Exploratory Mediation Analyses

However, our results have also shown that the mechanisms by which instructions and full item-level work contextualization operate are independent. Thus, we added exploratory mediation analyses (PROCESS Model 4; Hayes 2022) to investigate whether, when full item-level work contextualization was absent, the instructions (i.e., independent variable; GIs vs. PSS) impacted the scale mean scores of the self-report personality measure (i.e., dependent variable) independently of the adoption of a work FOR or whether the adoption of a work FOR mediated the effects of instructions on participants' scale mean scores. We used Version 5 (Beta) of the PROCESS macro for R (Hayes 2024) because we had a repeated-measures design, and this version can be used to estimate cluster-robust standard errors, which are an alternative to accounting for clustered observations (McNeish et al. 2017). The results demonstrated that, with the exception of a small partial mediation ( $B = 0.02$ ,  $p = 0.028$ ) of the mean scores in the Conscientiousness scale,<sup>20</sup> there was no mediation and thus no confounding of FOR and faking (see Table 7). With these additional analyses, we can conclude that the internal validity of faking research is not threatened by differences in the extent to which a work FOR is adopted.

#### 10.2.5 | Exploratory Reliability Analyses

As suggested by an anonymous reviewer, we additionally explored the effects of full item-level work contextualization (absent vs. present) in GIs versus the PSS on the reliability (i.e., Cronbach's alpha) of the self-report personality measure. In doing so, we followed the procedures outlined by Lievens et al. (2008).

**Reliability at Different Levels of Between-Person Variability.** We simulated the effects of the degree of between-person variability in responding to items with full item-level work contextualization (absent [−] vs. present [+]) in GIs versus the PSS on Cronbach's alpha by randomly drawing samples from the total sample ( $N = 309$ ) without replacement. We simulated the effects for each personality trait that we investigated (Conscientiousness, Emotionality, and Honesty-Humility). The random samples always consisted of the same 309 participants. However, they differed in the percentages of participants who had responded to the respective personality scale when fully work-contextualized items were used (i.e., +; 10% to 90%) and the percentages of participants who had responded to it when noncontextualized items were used (i.e., −; 90% to 10%), creating nine scenarios. For example, we drew a random sample consisting of the ratings of 10% of the sample when full item-level work contextualization was present and the ratings of 90% of the sample when full item-level work contextualization was absent. For each scenario, 1000 random samples of  $N = 309$  were drawn (i.e., 9000 samples altogether).

Table 8 summarizes the effects of different scenarios of between-person variability in responding to items with full item-level work contextualization (absent vs. present) on

**TABLE 7** | Results of the mediation analyses for each personality trait.

Path	<i>B</i>	<i>SE</i>	<i>t/Z</i>	<i>p</i>	95% CI	
					<i>LL</i>	<i>UL</i>
<i>Conscientiousness</i>						
A path: Instructions (GIs– or PSS–) → adoption of a work FOR	0.08	0.04	2.28	0.023	0.01	0.15
B path: Adoption of a work FOR → scale mean score on Conscientiousness	0.29	0.03	8.88	< 0.001	0.23	0.35
C' path: Instructions (GIs– or PSS–) → scale mean score on Conscientiousness	0.13	0.02	5.30	< 0.001	0.08	0.17
Total effect	0.15	0.02	6.65	< 0.001	0.11	0.19
Indirect effect	0.02	0.02	2.20	0.028	—	—
<i>Emotionality</i>						
A path: Instructions (GIs– or PSS–) → adoption of a work FOR	0.14	0.05	2.97	0.003	0.05	0.23
B path: Adoption of a work FOR → scale mean score on Emotionality	0.02	0.03	0.78	0.435	−0.03	0.08
C' path: Instructions (GIs– or PSS–) → scale mean score on Emotionality	−0.08	0.02	−4.36	< 0.001	−0.12	−0.04
Total effect	−0.08	0.02	−4.27	< 0.001	−0.11	−0.04
Indirect effect	0.00	0.00	0.72	0.472	—	—
<i>Honesty-Humility</i>						
A path: Instructions (GIs– or PSS–) → adoption of a work FOR	0.08	0.04	1.98	0.049	0.00	0.16
B path: Adoption of a work FOR → scale mean score on Honesty-Humility	0.02	0.04	0.50	0.617	−0.05	0.09
C' path: Instructions (GIs– or PSS–) → scale mean score on Honesty-Humility	0.03	0.02	1.56	0.119	−0.01	0.07
Total effect	0.03	0.02	1.66	0.097	−0.01	0.07
Indirect effect	0.00	0.00	0.44	0.663	—	—

Note: FOR = frame of reference; GIs = generic instructions, PSS = personnel-selection scenario. – indicates that the full item-level work contextualization was absent. CI = confidence interval; LL = lower limit; UL = upper limit. For the indirect effects *Z* values from the Sobel tests are reported, *t* values are reported for all other effects.

Cronbach's alpha in GIs versus the PSS concerning the three personality traits that were investigated. As shown by Lievens et al. (2008), the effects of the degree of between-person variability on Cronbach's alpha were quite small but varied across personality traits (i.e., a difference in Cronbach's alpha of 0.00 to 0.06). Extending Lievens et al.'s (2008) results, the pattern was independent of whether GIs or the PSS was used. These results are in line with Shaffer and Postlethwaite's (2012) meta-analysis in showing that the Cronbach's alpha values for non-contextualized and (work-) contextualized measures were virtually identical.

**Reliability at Different Levels of Within-Person Consistency.** To simulate the effects of the degree of within-person consistency in responding to items with full item-level work contextualization (absent [–] vs. present [+]) in GIs versus the PSS on Cronbach's alpha, we followed the procedure described above, but this time, the random samples differed in the number of items that were rated when fully work-contextualized items were used (i.e., +; 1 to 31) and the number of items that were rated when noncontextualized items were used (i.e., –; 31 to 1). For example, we drew a random sample consisting of one item that was rated when full item-level work contextualization was present and 31 items that were rated when full item-level work contextualization was absent. For

each scenario, 1000 random samples of *N* = 309 were drawn (i.e., 31,000 samples altogether).

Table 9 summarizes the reliability effects of different scenarios of within-person consistency in responding to items with full item-level work contextualization (absent vs. present) in GIs versus the PSS on Cronbach's alpha for the three personality traits. As shown by Lievens et al. (2008), the effects of the degree of within-person consistency on Cronbach's alpha were curvilinear and varied across the personality traits. Extending Lievens et al.'s (2008) findings, the pattern was independent of whether GIs or the PSS was used. Also, in extending Lievens et al.'s (2008) findings, Cronbach's alpha was lowest when all participants rated about half of the items when full item-level work contextualization was present and the other half of the items when full item-level work contextualization was absent for Conscientiousness only. For Emotionality and Honesty-Humility, Cronbach's alpha was lowest when all participants rated about three-fourths of the items when full item-level work contextualization was present and the other quarter of the items when full item-level work contextualization was absent. In contrast to Lievens et al. (2008), the effects on Cronbach's alpha were again small and varied somewhat across personality traits (i.e., a difference in Cronbach's alpha of 0.02 to 0.09). Our results are in line with Shaffer and Postlethwaite's

**TABLE 8** | Summary of Cronbach's alphas at different levels of the between-person variability by experimental condition and personality trait.

Scenarios	GIs			PSS		
	Conscientiousness	Emotionality	Honesty-Humility	Conscientiousness	Emotionality	Honesty-Humility
10% participants +, 90% -	0.90	0.85	0.92	0.92	0.87	0.92
20% participants +, 80% -	0.90	0.84	0.91	0.91	0.86	0.91
30% participants +, 70% -	0.90	0.84	0.91	0.91	0.86	0.91
40% participants +, 60% -	0.90	0.83	0.91	0.91	0.86	0.90
50% participants +, 50% -	0.90	0.83	0.91	0.91	0.85	0.90
60% participants +, 40% -	0.90	0.82	0.90	0.91	0.84	0.90
70% participants +, 30% -	0.90	0.81	0.90	0.91	0.83	0.89
80% participants +, 20% -	0.90	0.80	0.90	0.92	0.82	0.89
90% participants +, 10% -	0.90	0.79	0.89	0.92	0.81	0.89
<i>M</i>	0.90	0.82	0.90	0.91	0.85	0.90
<i>SD</i>	0.00	0.02	0.01	0.00	0.02	0.01
Minimum	0.90	0.79	0.89	0.91	0.81	0.89
Maximum	0.90	0.85	0.92	0.92	0.87	0.92

*Note:* For each of the nine scenarios, 1000 random samples were drawn. Summary statistics at the bottom were computed across all 9000 samples. GIs = generic instructions, PSS = personnel-selection scenario. + indicates that the full item-level work contextualization was present. - indicates that the full item-level work contextualization was absent.

**TABLE 9** | Summary of Cronbach's alphas at different levels of within-person inconsistency by experimental condition and personality trait.

Scenarios	GIs			PSS		
	Conscientiousness	Emotionality	Honesty-Humility	Conscientiousness	Emotionality	Honesty-Humility
1 + item, 31 – items	0.90	0.84	0.92	0.91	0.87	0.92
2 + items, 30 – items	0.90	0.84	0.92	0.91	0.86	0.92
3 + items, 29 – items	0.89	0.83	0.91	0.91	0.86	0.91
4 + items, 28 – items	0.89	0.83	0.91	0.90	0.85	0.91
5 + items, 27 – items	0.89	0.82	0.91	0.90	0.85	0.91
6 + items, 26 – items	0.89	0.82	0.91	0.90	0.84	0.91
7 + items, 25 – items	0.88	0.81	0.91	0.90	0.84	0.90
8 + items, 24 – items	0.88	0.81	0.90	0.89	0.83	0.90
9 + items, 23 – items	0.88	0.80	0.90	0.89	0.83	0.90
10 + items, 22 – items	0.88	0.80	0.90	0.89	0.82	0.90
11 + items, 21 – items	0.88	0.79	0.90	0.89	0.82	0.89
12 + items, 20 – items	0.88	0.79	0.90	0.89	0.81	0.89
13 + items, 19 – items	0.88	0.78	0.90	0.89	0.81	0.89
14 + items, 18 – items	0.87	0.78	0.89	0.89	0.81	0.89
15 + items, 17 – items	0.87	0.78	0.89	0.89	0.80	0.89
16 + items, 16 – items	<b>0.87</b>	0.77	0.89	<b>0.89</b>	0.80	0.89
17 + items, 15 – items	0.87	0.77	0.89	0.89	0.80	0.88
18 + items, 14 – items	0.87	0.77	0.89	0.89	0.79	0.88
19 + items, 13 – items	0.87	0.76	0.89	0.89	0.79	0.88
20 + items, 12 – items	0.88	0.76	0.89	0.89	0.79	0.88
21 + items, 11 – items	0.88	0.76	0.89	0.89	0.79	0.88
22 + items, 10 – items	0.88	0.76	0.89	0.89	0.79	0.88
23 + items, 9 – items	0.88	0.76	0.89	0.89	0.78	0.88
24 + items, 8 – items	0.88	<b>0.76</b>	0.89	0.89	0.78	0.88
25 + items, 7 – items	0.88	0.76	0.89	0.90	<b>0.78</b>	<b>0.88</b>
26 + items, 6 – items	0.89	0.76	<b>0.89</b>	0.90	0.78	0.88
27 + items, 5 – items	0.89	0.76	0.89	0.90	0.78	0.88
28 + items, 4 – items	0.89	0.76	0.89	0.90	0.79	0.88
29 + items, 3 – items	0.90	0.77	0.89	0.91	0.79	0.88

(Continues)

TABLE 9 | (Continued)

Scenarios	GIs			PSS		
	Conscientiousness	Emotionality	Honesty-Humility	Conscientiousness	Emotionality	Honesty-Humility
30 + items, 2 – items	0.90	0.77	0.89	0.91	0.79	0.88
31 + items, 1 – item	0.90	0.77	0.89	0.92	0.79	0.88
<i>M</i>	0.88	0.79	0.90	0.90	0.81	0.89
<i>SD</i>	0.01	0.03	0.01	0.01	0.03	0.01
Minimum	0.87	0.76	0.89	0.89	0.78	0.88
Maximum	0.90	0.84	0.92	0.92	0.87	0.92

Note: For each of the 31 scenarios, 1000 random samples were drawn. The summary statistics at the bottom were computed across all 31,000 samples. GIs = generic instructions, PSS = personnel-selection scenario. + indicates that the full item-level work contextualization was present. – indicates that the full item-level work contextualization was absent. Per sample, the particular items rated with + or – were not fixed across participants. Values in bold indicate the minimum value of Cronbach's alpha and the turning point (i.e., after which Cronbach's alpha increased again).

(2012) meta-analysis, which showed that the Cronbach's alpha values for the noncontextualized and (work-)contextualized measures were virtually identical.

## 11 | Discussion

In laboratory settings, faking research typically uses a PSS to increase participants' motivation to fake. However, for more than 20 years, researchers have argued that FOR effects may confound the findings from such studies. This concern has sparked doubts about the internal validity of faking studies and is far from trivial. If FOR and faking are indeed confounded, the findings of decades of faking research may have been biased. For example, changes in participants' scale mean scores could have been falsely attributed to faking only when they had actually been due to both faking and FOR effects. The current study was conducted to disentangle the effects of FOR and faking by varying full item-level work contextualization and thus, to investigate faking while controlling for the FOR. Our findings suggest that no matter which personality trait was investigated (Conscientiousness, Emotionality, or Honesty-Humility), the effects of instructions (GIs vs. PSS) and full item-level work contextualization (absent vs. present) on the adoption of a work FOR and on the scale mean scores of the self-report personality measure were additive and worked independently. Exploratory mediation analyses indicated direct rather than indirect (mediator: adoption of a work FOR) effects of instructions on participants' scale mean scores. The current study thereby demonstrates that the internal validity of faking research is not threatened by confounding FOR effects.

### 11.1 | Theoretical and Practical Implications

#### 11.1.1 | The PSS Promotes the Adoption of a Work FOR

In line with previous suggestions (see Birkeland et al. 2006; Robie 2001; Robie et al. 2001), the current study demonstrated that, compared with GIs, the instructions that are typically used in experimental laboratory faking research (i.e., the PSS) indeed increased the adoption of a work FOR. Moreover, this effect held across the personality traits under investigation (Conscientiousness, Emotionality, and Honesty-Humility). Thus, previous theorizing correctly assumed that the PSS induces more than just faking. It also promotes the adoption of a work FOR (see Anglim et al. 2021; Bing et al. 2004; Birkeland et al. 2006; Robie 2001; Robie et al. 2001). Although researchers should keep this issue in mind, it turned out that this issue does not threaten the internal validity of faking research (see below).

#### 11.1.2 | Full Item-Level Work Contextualization Promotes the Adoption of a Work FOR

In agreement with previous research (e.g., Fisher et al. 2017; Holtrop et al. 2014), the current study demonstrated that full item-level work contextualization increased the adoption of a work FOR. Extending previous research, the current study showed that the effects varied depending on the personality



trait under investigation (i.e., Emotionality and Honesty-Humility: large effects; Conscientiousness: moderate effect). Most likely, the noncontextualized Conscientiousness scale (but not the Emotionality and Honesty-Humility scales) had already elicited a work FOR through a *hidden* work framing of its items (see Schulze et al. 2021), and thus, there was less room for change. In line with this assumption, the adoption of a work FOR was rated highest for the Conscientiousness scale in all conditions, and the differences in ratings between the conditions of full item-level work contextualization (i.e., absent vs. present) were smallest for the Conscientiousness scale (Table 5; Figure 2).

To explore this assumption more deeply, we applied the procedure outlined by Schulze et al. (2021) and searched for the keywords *job*, *work(ing)*, and *task(s)* that may have primed the work FOR in the noncontextualized version of the self-report personality measure. Ten out of the 32 Conscientiousness items (i.e., 31.25%), but only one out of the 32 Emotionality items (i.e., 3.13%) and none of the 32 Honesty-Humility items (i.e., 0.00%), may have already primed the work FOR in the noncontextualized version. These results support the interpretation that hidden work framing impacted how strongly the full item-level work contextualization increased the adoption of a work FOR. Thus, hidden FORs could impact the results of studies investigating the effects of adopted FORs on scale characteristics (e.g., smaller effects when investigating Conscientiousness scales than when using other scales), and researchers should bear this issue in mind.

### 11.1.3 | The Adoption of a Work FOR Is Stronger When Using Full Item-Level Work Contextualization Than When Using the PSS

In line with previous research (e.g., Holtrop et al. 2014; Lievens et al. 2008; Schlotzhauer et al. 2024; Swift and Peterson 2019), the adoption of a work FOR was stronger when the respective contextualization was done via a full item-level contextualization (i.e., moderate to large effects) than via a mere manipulation of instructions (i.e., small effects). These differences were stronger for the Emotionality and Honesty-Humility scales than for the Conscientiousness scale, which again may have been caused by the hidden work framing of Conscientiousness items (Schulze et al. 2021). Thus, hidden framings in items may also have impacted FOR research on different types of contextualization because the choice of traits that were investigated may have differentially restricted the potential to change.

### 11.1.4 | Full Item-Level Work Contextualization and Instructions Operated Independently With Respect to the Adoption of a Work FOR

Importantly and in contrast to previous suggestions, no significant interactions were found for any of the three personality traits that were investigated in this study. The effect of instructions (i.e., GIs vs. PSS) on the adoption of a work FOR was independent of whether or not the items were fully contextualized with a work FOR, and the effect of full item-level

work contextualization (i.e., absent vs. present) on the adoption of a work FOR was independent of whether participants were asked to fake (PSS) or not (GIs). Thus, although the manipulation of the instructions as typically applied in experimental laboratory faking research clearly impacts two processes (i.e., the adoption of a work FOR and the motivation to fake), the results indicate that the mechanisms by which instructions and full item-level work contextualization operate are independent. Each factor exerts its own unique influence on participants' adoption of a work FOR, but these influences do not interact, and the question is why.

When faking is not the focus of a study, instruction-level and item-level contextualization might indeed be redundant, differing only in strength, albeit this claim still has to be demonstrated empirically. However, if the instructions include a request to fake, the situation will most likely become more complex. For example, changes in the adoption of a work FOR via instructions might then not reflect genuine increases in the adoption of a work FOR but rather strategic ones—an increased work framing that is used to respond in the desired manner, possibly because it facilitates faking (e.g., How should I *present* myself at work?). In other words, participants might have reflected on their typical behavior at work but used this schema as a framework to optimize their responses within the given work context. In this case, the work FOR would be used as a guide for faking and its increased adoption would reflect only one step in the larger faking process. This explanation aligns with conceptualizations of faking as role-playing in a certain context (e.g., Levashina and Campion 2006), the adopted schema model (e.g., Furnham 1990; Mahar et al. 1995), with findings that personal experiences serve as anchors for faking (Ellingson and McFarland 2011; Goffin and Boyd 2009), with considerations of self-presentation in selection settings (e.g., Roulin and Krings 2020), and with recently identified faking strategies (Röhner et al. 2025).

### 11.1.5 | Faking Changes Response Behavior and the Changes Vary With the Personality Trait Under Investigation

Replicating the results of previous research, the current study demonstrated that irrespective of the personality trait (i.e., Conscientiousness, Emotionality, or Honesty-Humility), the instructions to fake (PSS) versus to respond honestly (GIs) changed participants' scale mean scores (see Salgado 2016; Ziegler et al. 2012). Thus, participants had faked successfully (i.e., they increased their scores on Conscientiousness and Honesty-Humility, decreased their scores on Emotionality). In line with previous findings (e.g., Birkeland et al. 2006), the effects of faking varied with respect to the personality trait under investigation (i.e., Conscientiousness: large effect vs. Emotionality: moderate effect vs. Honesty-Humility: small effect). Thus, as has previously been recommended (e.g., Birkeland et al. 2006; Röhner et al. 2022), when interpreting the effects of faking, researchers should consider which personality trait was investigated, and when investigating faking, they should ideally compare various traits. This procedure could prevent the risk of overestimating or underestimating the effects

of faking—especially when research is aimed at drawing general conclusions on this topic.

#### **11.1.6 | Full Item-Level Work Contextualization Changes Response Behavior and the Changes Vary With the Personality Trait Under Investigation**

In line with previous research (e.g., Holtrop et al. 2014), the current study demonstrated that full item-level work contextualization changed the scale mean scores on the self-report personality measure (i.e., participants' responses differed, depending on the salience of a work FOR) and that these changes varied depending on which personality trait was investigated (e.g., Conscientiousness and Emotionality: large effects vs. Honesty-Humility: moderate effect).

Compared with when noncontextualized items were used, participants increased (Conscientiousness) or decreased (Emotionality, Honesty-Humility) their scale mean scores on the self-report personality measure when fully work-contextualized items were used. Thus, except for Honesty-Humility, such scores may be considered more positive when fully work-contextualized items are used than when non-contextualized items are used (see Bowling and Burns 2010; Holtrop et al. 2014; Pathki et al. 2022).

Honesty-Humility could be regarded as a specific case because a decrease in this trait is not considered more positive, and previous research on contextualization found only nonsignificant decreases in Honesty-Humility (see Holtrop et al. 2014). From an applied perspective, the decrease is indeed plausible because true high Honesty-Humility could even be detrimental in work settings.<sup>21</sup> This finding supports the idea that the changes in response behavior that are caused by a variation in the adopted FOR reflect “honest” self-presentation that is related to differences in behavior between contexts (here, being less honest and less humble at work). Thus, our research extends previous findings by demonstrating that, for some personality traits, FORs and faking may push participants' scale mean scores in different directions (e.g., faking: higher reported Honesty-Humility vs. FOR: lower reported Honesty-Humility), a finding that researchers should consider.

#### **11.1.7 | Full Item-Level Work Contextualization and Instructions Operated Independently With Respect to Changes in Response Behavior**

In contrast to previous suggestions, there were no significant interactions for any of the three personality traits that we investigated. The effect of instructions (i.e., GIs vs. PSS) on the changes in the scale mean scores was independent of whether the items were fully contextualized with a work FOR or not, and the effect of full item-level work contextualization (i.e., absent vs. present) on the scale mean scores was independent of whether participants were asked to fake (PSS) or not (GIs). Apparently, each factor exerted its own unique influence on participants' changes in response behavior, but these influences did not interact. The current research also showed that self-report

personality measures with full item-level work contextualization can be faked to the same extent as noncontextualized self-report personality measures (see Figure 3).

#### **11.1.8 | The Internal Validity of Faking Research Is Most Likely Not Threatened by Confounded FOR Effects**

On the basis of exploratory mediation analyses, the assumption that the increased adoption of a work FOR (and not faking) causes changes in participants' response behavior could finally be ruled out. In fact, the direct effects of instructions (i.e., GIs vs. PSS) on participants' scale mean scores were stronger than the (predominantly nonsignificant) indirect effects via the adoption of a work FOR. There was a small partial mediation effect on the mean score of the Conscientiousness scale, and future research is needed to determine whether this effect is caused by the strategic adoption of a work FOR in a PSS that might facilitate faking or by the genuine adoption of a work FOR that might instead elicit honest responses (see the argument above). Nevertheless, compared with the direct effects, the findings clearly point to the conclusion that the FOR and faking are not confounded.

### **11.2 | Limitations and Future Research**

The present study has potential limitations. First, the data provided to us included many careless responders. We followed Meade and Craig's (2012) recommendation to detect carelessness by using instructed response items because not only has this approach been demonstrated to have high internal consistency, good convergent validity with other indices of careless responding, and even good discriminant validity with the response style of acquiescence (Kam and Chan 2018), but it has also been demonstrated to detect various types of carelessness (e.g., Gummer et al. 2021). As a consequence, however, this approach is also more conservative (i.e., it flags a comparably large number of participants when compared with, e.g., straightlining indices that focus on only one facet of carelessness). We ensured that, in each wave of measurement, one instructed response item was presented per scale. The presentation of our instructed response items was also randomized between and within participants. By doing so, we aimed to prevent participants from inferring the position of the attention checks (e.g., it was always the last item). This procedure likely increased the number of careless responders that were detected in our study, but as a consequence, we are quite confident that we were able to exclude various types of careless responding before we ran subsequent analyses.

Second, we used a German online community sample (two-thirds male). Future research could investigate the effects in other samples and in other experimental contexts to investigate the generalizability of the current results.

Third, we investigated the effects of instructions and full item-level work contextualization on the adoption of a work FOR and on participants' response behavior in a self-report personality measure assessing only Conscientiousness, Emotionality,

and Honesty-Humility. We investigated these particular traits because they are the three personality dimensions that are most predictive of behavior at work (e.g., De Vries et al. 2014; Holtrop et al. 2014; Lee et al. 2005) and because a self-report personality measure with full item-level work contextualization is available for them (Holtrop et al. 2014). Future research could extend our findings by investigating other personality traits.

Fourth, for response behavior, we focused on changes in the scale mean scores because the effects of both the FOR and faking have typically been documented with respect to these changes (e.g., Shaffer and Postlethwaite 2012; Ziegler et al. 2012). Although the exploratory simulation analyses on between-person variability and within-person consistency in responding to items with full item-level work contextualization (absent vs. present) in GIs versus the PSS on Cronbach's alpha demonstrated rather small effects, changes in reliability and validity are nevertheless interesting avenues for future research because the current study also demonstrated differences across personality traits.

Fifth, on the basis of previous theorizing about the FOR, we had assumed that the effects of instruction-level and full item-level work contextualization were redundant, meaning that the latter alone should elicit the “full FOR-effect” (e.g., Holtrop et al. 2014), whereas the effect of instructions would be weaker. The design of our Registered Report was based on this reasoning. However, our results suggest that the two types of contextualization are independent (at least when the instructions include a request to fake)—a finding that raises many questions and supports recent requests to conduct more systematic comparisons of instruction-level and item-level contextualization (Schlotzhauer et al. 2024). The current study suggests that the specific mechanisms underlying this independence warrant further investigation. Qualitative data (e.g., from think-aloud protocols) could shed light on the cognitive mechanisms behind different types of contextualization under different instructions (e.g., faking vs. responding honestly). Future studies could also add a third set of instructions asking participants to think about their work (as has been done in FOR research before; e.g., Hunthausen et al. 2003) but to respond honestly. By including qualitative data and such instructions, researchers could investigate the respective mechanisms concerning different conditions in more depth.

## 12 | Conclusion

The current study disentangled the effects of FOR and faking to investigate whether the two phenomena may be confounded in faking research. Our findings indicate that, regardless of the personality trait under investigation (i.e., Conscientiousness, Emotionality, or Honesty-Humility), the effects of instructions (GIs vs. PSS) and full item-level work contextualization (absent vs. present) on the adoption of a work FOR and on the changes in response behavior were additive and operated independently. Exploratory mediation analyses indicated direct rather than indirect (mediator: adoption of a work FOR) effects of instructions on participants' scale mean scores. The current study thereby demonstrates that the internal validity of faking research is *not threatened* by confounding FOR effects.

## Author Contributions

**Jessica Röhner:** conceptualization – revisions and extensions, data curation, formal analyses, funding acquisition, methodology – revisions and extensions, project administration, validation, visualization, writing – original draft (stage 2), writing – reviewing and editing (Stages 1 and 2). **Mia Degro:** initial conceptualization and methodology, preparatory work, preliminary analyses, and writing – original draft (Stage 1). **Ronald R. Holden:** writing – reviewing and editing (Stages 1 and 2). **Astrid Schütz:** funding acquisition, supervision, writing – reviewing and editing (Stages 1 and 2). All authors revised and approved the final manuscript.

## Acknowledgments

This study was partly funded by a grant from the Equal Opportunities Office and by a grant from the Internal Research Fund at the University of Bamberg. The funding sources were not involved in designing the study or analyzing the data.

## Ethics Statement

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The study was evaluated by a University Ethics Committee, and approval was granted (approval number: 2022-03/14).

## Consent

All participants provided informed consent to participate in the study and to have their data published in a journal article.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Data Availability Statement

The data that support the findings of this study are openly available in OSF at <https://osf.io/57n4e/>.

## Endnotes

<sup>1</sup>Although Birkeland et al. (2006) used applicants and non-applicants to analyze the effects of faking, the same reasoning can be used to explain differences between faking and control conditions in the laboratory. Note that the approach of comparing applicants with nonapplicants has been criticized because the two samples can differ from each other on a number of relevant aspects (e.g., demographics).

<sup>2</sup>Note that this indicates that the FOR effect elicited by a PSS can be expected to be comparably small.

<sup>3</sup>The items used to assess the adopted FORs were: “When answering these [personality] questions, I considered my typical behavior in educational settings (high school, college, study groups, etc.),” “... I considered my typical behavior in social settings (with friends, family, etc.),” “... I considered my typical behavior in work settings (work assignments, interacting with coworkers, etc.),” and “... I considered my typical behavior in a variety of different settings (educational, social, work, private, etc.)” (see Fisher et al. 2017, p. 23 and p. 25). Responses were given on a 5-point scale ranging from 5 = *strongly agree* to 1 = *strongly disagree* (Fisher et al. 2017). Thus, higher scores indicated a stronger adoption of the respective FOR.

<sup>4</sup>The scores on Conscientiousness increased moderately ( $M_{\text{noncontextualized}}$ : 5.22 vs.  $M_{\text{contextualized:workFOR}}$ : 5.58;  $d=0.57$ , 95% CI [0.60, 0.97]) in

- Bowling and Burns (2010) and small ( $M_{\text{noncontextualized}}$ : 3.57 vs.  $M_{\text{contextualized:workFOR}}$ : 3.72;  $d = 0.47$ , 95% CI [0.31, 0.78]) in Holtrop et al. (2014). In Pathki et al. (2022) the scores on Conscientiousness increased with a medium effect size ( $M_{\text{noncontextualized}}$ : 4.97 vs.  $M_{\text{contextualized:workFOR}}$ : 5.59;  $d = 0.59$ , 95% CI [0.36, 0.83] in Study 4b). The scores on Emotionality decreased strongly ( $M_{\text{noncontextualized}}$ : 3.25 vs.  $M_{\text{contextualized:workFOR}}$ : 2.87;  $d = -2.40$ , 95% CI [-2.67, -2.13]) in Holtrop et al. (2014). The scores on Honesty-Humility decreased nonsignificantly ( $M_{\text{noncontextualized}}$ : 3.97 vs.  $M_{\text{contextualized:workFOR}}$ : 3.93;  $d = -0.08$ , 95% CI [-0.30, 0.18]) in Holtrop et al. (2014).
- <sup>5</sup>In Fisher et al. (2017), the scores on Conscientiousness differed nonsignificantly or to a small extent ( $M_{\text{noncontextualized}}$ : 4.61 vs.  $M_{\text{contextualized:workFOR}}$ : 4.70,  $d = 0.24$ , 95% CI [-0.10, 0.58] in Sample 1;  $M_{\text{noncontextualized}}$ : 4.53 vs.  $M_{\text{contextualized:workFOR}}$ : 4.56,  $d = 0.08$ , 95% CI [-0.27, 0.32] in Sample 2;  $M_{\text{noncontextualized}}$ : 4.46 vs.  $M_{\text{contextualized:workFOR}}$ : 4.59,  $d = 0.36$ , 95% CI [0.04, 0.67] in Sample 3; and  $M_{\text{noncontextualized}}$ : 4.60 vs.  $M_{\text{contextualized:workFOR}}$ : 4.66,  $d = 0.16$ , 95% CI [-0.11, 0.42] in Sample 4).
- <sup>6</sup>Measures with this type of item-level contextualization are still rare.
- <sup>7</sup>The occupations were classified in accordance with the clusters that are used in the Federal Agency of Work registers (Bundesagentur für Arbeit 2020). These registers include all occupations in Germany. For better readability, some longer cluster designations were shortened. The original designations and participants' reports of their occupations are stored on the OSF.
- <sup>8</sup>We deviated from Fisher et al.'s (2017) procedure because we included an item asking for a "faking context." We asked for several FORs to prevent participants from concluding that our focus was on the work FOR.
- <sup>9</sup>The mean scores of the ratings on the adopted FORs, the respective standard deviations, and Cronbach's alphas are presented in the Appendix (see A6).
- <sup>10</sup>The alpha level was adjusted with the Bonferroni correction. For each personality trait under investigation and in each of the two-way repeated-measures ANOVAs, there were four conditions (i.e.,  $k$ ; 2 [instructions: GIs vs. PSS]  $\times$  2 [full item-level contextualization: absent [-] vs. present [+]]), resulting in six possible post hoc comparisons (i.e.,  $j$ ;  $j = [k \times [k - 1]]/2$ ; GIs+ vs. GIs-, PSS+ vs. PSS-, GIs+ vs. PSS+, GIs+ vs. PSS-, GIs- vs. PSS+, and GIs- vs. PSS-). After applying the Bonferroni correction,  $\alpha' = \text{cumulative alpha level}/j$ , the corrected alpha level, was  $0.05/6 = 0.008$ .
- <sup>11</sup>When submitting our Registered Report, we planned to oversample and increase the number of participants necessary for the first wave of measurement, aiming to collect data from at least 700 participants. This value was considered sufficient on the basis of previous research (Davison et al. 2021; Lee et al. 2019; Liu and Zhang 2020; Roess and Roche 2017; Terry et al. 1999). Nevertheless, during the first wave, the data provider informed us of a noticeable reduction in the number of qualified participants. In particular, a substantial number of participants were lost because they did not complete follow-up assessments within the designated time window. Furthermore, nearly the same number of participants were identified as careless participants by our attention checks. Thus, we oversampled even further based on the attrition rate from the first wave of data collection.
- <sup>12</sup>As with the pretest, the occupations were classified in accordance with the clusters used in the Federal Agency of Work registers (Bundesagentur für Arbeit 2020). These registers include all occupations in Germany. For better readability, some longer cluster designations have been shortened. The original designations and participants' reports of their occupations are stored on the OSF.
- <sup>13</sup>Fisher et al. (2017) did not use fully contextualized items but added the words "at work" at the beginning of items (i.e., they used "at work" tags). Furthermore, they did not include a PSS in their study. Nevertheless, we compared our values with Fisher et al.'s (2017) values because their study was most comparable to ours, and they investigated this issue.
- <sup>14</sup>We describe the values as being only *largely* comparable because we obtained larger standard deviations and Cronbach's alphas than were found by Holtrop et al. (2014). However, these differences may be a result of the different samples that were used. Holtrop et al. (2014) restricted their sample to pharmacy assistants, and thus, due to the specific sample they used, their values might be somewhat different from ours (because we included a variety of occupations in our sample).
- <sup>15</sup>These changes differed between the scales and were in line with the main effects of instructions (i.e., GIs vs. PSS and thus, faking; see below; see also Table 6, lower half). The respective effect size of the main effect of instructions was small for Honesty-Humility but large for Conscientiousness and Emotionality.
- <sup>16</sup>To adjust for violations of normality for the adopted work FOR (see A13 to A15 in the Appendix), we used bootstrapping with the bias-corrected and accelerated (BCa) method ( $n = 10,000$  resamples; Berkovits et al. 2000). The requirement of homogeneity of variance was violated (see A16 in the Appendix), and thus, we set  $\alpha$  to 0.010 for all ANOVAs (see Weiner et al. 2003). The requirement of sphericity was not violated, and thus, we did not apply Huynh-Feldt or Greenhouse-Geisser corrections.
- <sup>17</sup>One anonymous reviewer recommended that we reintegrate the participants who failed the predefined markers for nonblinding (see Table 3) and explore whether there were differences in the two-way repeated-measures ANOVAs. This sample consisted of 358 participants (123 women, 234 men, 1 diverse) with an average age of 47.60 years ( $SD = 13.10$ ). The results did not change significantly (see A17 to A23 in the Appendix).
- <sup>18</sup>The requirement of homogeneity of variance was violated (see A16 in the Appendix), and thus, we set  $\alpha$  to 0.010 for all ANOVAs (see Weiner et al. 2003). The requirement of sphericity was not violated, and thus, we did not use Huynh-Feldt or Greenhouse-Geisser corrections. For reasons of comparability, we used bootstrapping with the BCa method ( $n = 10,000$  resamples).
- <sup>19</sup>One anonymous reviewer recommended that we reintegrate the participants who failed the predefined markers for nonblinding (see Table 3) and explore whether there were differences in the two-way repeated-measures ANOVAs. This sample consisted of 358 participants (123 women, 234 men, 1 diverse) with an average age of 47.60 years ( $SD = 13.10$ ). The results did not change significantly (see A17 to A23 in the Appendix).
- <sup>20</sup>In other words, when participants were given a PSS instead of GIs, the indirect effect explained a change in the mean score of the Conscientiousness scale (ranging from 1 to 5) of only 0.02. In addition, the direct effect was still significant and also larger in size ( $b = 0.13$ ; see Table 7).
- <sup>21</sup>By contrast, faking high scores on Honesty-Humility to get the job might be helpful. Note that acting like one is honest and humble will most likely help the participant (and also the employer and organization) make a good impression (i.e., faking). Nevertheless, there may be several reasons to actually be less honest and less humble at work (i.e., actual work-related behavior).
- <sup>22</sup>We used instructed response items as attention checks and chose one item for each scale of the self-report personality measure (Edwards 2019). The presentation of our instructed response items was randomized between and within participants.
- <sup>23</sup>Responses were given on a 5-point Likert scale (1 = *strongly disagree* to 5 = *strongly agree*).
- <sup>24</sup>Job Description Pharmacy Assistant. Translated from Bundesagentur für Arbeit, 2023. Pharmazeutisch-technische/r Assistent/in. Ausbildungsberuf. <https://web.arbeitsagentur.de/berufenet/beruf/8910>



<sup>25</sup> Responses were given on a 5-point Likert scale (1 = *strongly disagree* to 5 = *strongly agree*).

## References

- Anglim, J., F. Lievens, L. Everton, S. L. Grant, and A. Marty. 2018. "HEXACO Personality Predicts Counterproductive Work Behavior and Organizational Citizenship Behavior in Low-Stakes and Job Applicant Contexts." *Journal of Research in Personality* 77: 11–20. <https://doi.org/10.1016/j.jrp.2018.09.003>.
- Anglim, J., K. Molloy, P. D. Dunlop, S. L. Albrecht, F. Lievens, and A. Marty. 2021. "Values Assessment for Personnel-Selection: Comparing Job Applicants to Non-Applicants." *European Journal of Work and Organizational Psychology* 31, no. 4: 524–536. <https://doi.org/10.1080/1359432X.2021.2008911>.
- Anglim, J., G. Morse, R. E. De Vries, C. MacCann, and A. Marty. 2017. "Comparing Job Applicants to Non-Applicants Using an Item-Level Bifactor Model on the HEXACO Personality Inventory." *European Journal of Personality* 31, no. 6: 669–684. <https://doi.org/10.1002/per.2120>.
- Ashton, M. C., X. Xia, and K. Lee. 2020. "Is There a G in HEXACO? Testing for a General Factor in Personality Self-Reports Under Different Conditions of Responding." *Personality and Individual Differences* 156: 109750. <https://doi.org/10.1016/j.paid.2019.109750>.
- Attali, D., and C. Baker. 2023. *ggExtra: Add Marginal Histograms to 'ggplot2', and More 'ggplot2' Enhancements* (Version 0.10.1) [R package]. <https://CRAN.R-project.org/package=ggExtra>.
- Bensch, D., U. Maaß, S. Greiff, K. T. Horstmann, and M. Ziegler. 2019. "The Nature of Faking: A Homogeneous and Predictable Construct?" *Psychological Assessment* 31, no. 4: 532–544. <https://doi.org/10.1037/pas0000619>.
- Ben-Shachar, M., D. Lüdtke, and D. Makowski. 2020. "Effectsize: Estimation of Effect Size Indices and Standardized Parameters." *Journal of Open Source Software* 5, no. 56: 2815. <https://doi.org/10.21105/joss.02815>.
- Berkovits, I., G. R. Hancock, and J. Nevitt. 2000. "Bootstrap Resampling Approaches for Repeated Measure Designs: Relative Robustness to Sphericity and Normality Violations." *Educational and Psychological Measurement* 60, no. 6: 877–892. <https://doi.org/10.1177/00131640021970961>.
- Bing, M. N., J. C. Whanger, H. K. Davison, and J. B. VanHook. 2004. "Incremental Validity of the Frame-of-Reference Effect in Personality Scale Scores: A Replication and Extension." *Journal of Applied Psychology* 89, no. 1: 150–157. <https://doi.org/10.1037/0021-9010.89.1.150>.
- Birkeland, S. A., T. M. Manson, J. L. Kisamore, M. T. Brannick, and M. A. Smith. 2006. "A Meta-Analytic Investigation of Job Applicant Faking on Personality Measures." *International Journal of Selection and Assessment* 14, no. 4: 317–335. <https://doi.org/10.1111/j.1468-2389.2006.00354.x>.
- Blickle, G., T. Momm, P. B. Schneider, D. Gansen, and J. Kramer. 2009. "Does Acquisitive Self-Presentation in Personality Self-Ratings Enhance Validity? Evidence From Two Experimental Field Studies." *International Journal of Selection and Assessment* 17, no. 2: 142–153. <https://doi.org/10.1111/j.1468-2389.2009.00458.x>.
- Bowling, N. A., and G. N. Burns. 2010. "A Comparison of Work-Specific and General Personality Measures as Predictors of Work and Non-Work Criteria." *Personality and Individual Differences* 49, no. 2: 95–101. <https://doi.org/10.1016/j.paid.2010.03.009>.
- Van den Brand, T. 2024. *ggh4x: Hacks for 'ggplot2' (Version 0.2.8)*. [Computer software].
- Brown, A., and U. Böckenholt. 2022. "Intermittent Faking of Personality Profiles in High-Stakes Assessments: A Grade of Membership Analysis." *Psychological Methods* 27, no. 5: 895–916. <https://doi.org/10.1037/met0000295>.
- Bundesagentur für Arbeit. Klassifikation der Berufe 2010 –überarbeitete Fassung 2020: Systematisches Verzeichnis; 2020. Accessed November 16, 2023. <https://statistik.arbeitsagentur.de/DE/Navigation/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/KldB2010-Fassung2020/Systematik-Verzeichnisse/Systematik-Verzeichnisse-Nav.html;jsessionid=5E1E29BA86C52A1456CD4939B1AE1991>.
- Burns, G. N., and N. D. Christiansen. 2006. "Sensitive or Senseless: On the Use of Social Desirability Measures in Selection and Assessment." In *A Closer Examination of Applicant Faking Behavior*, edited by R. L. Griffith and M. H. Peterson, 113–148. <https://doi.org/10.1034/j.1600-0706.2001.950116.x>.
- Canty, A., and B. Ripley. 2024. *Boot: Bootstrap R (S-Plus) Functions (Version 1.3-31)*. [Computer software].
- Connelly, B. S., and L. Chang. 2016. "A Meta-Analytic Multitrait Multirater Separation of Substance and Style in Social Desirability Scales." *Journal of Personality* 84, no. 3: 319–334. <https://doi.org/10.1111/jopy.12161>.
- Corneille, O., and P. Lush. 2023. "Sixty Years After Orne's *American psychologist* Article: A Conceptual Framework for Subjective Experiences Elicited by Demand Characteristics." *Personality and Social Psychology Review* 27, no. 1: 83–101. <https://doi.org/10.1177/10888683221104368>.
- Crowne, D. P., and D. Marlowe. 1960. "A New Scale of Social Desirability Independent of Psychopathology." *Journal of Consulting Psychology* 24, no. 4: 349–354. <https://doi.org/10.1037/h0047358>.
- Dang, J., K. M. King, and M. Inzlicht. 2020. "Why Are Self-Report and Behavioral Measures Weakly Correlated?" *Trends in Cognitive Sciences* 24, no. 4: 267–269. <https://doi.org/10.1016/j.tics.2020.01.007>.
- Davison, H. K., and M. N. Bing. 2009. "Content Validity Does Matter for the Criterion-Related Validity of Personality Tests." *Industrial and Organizational Psychology* 2, no. 4: 501–503. <https://doi.org/10.1111/j.1754-9434.2009.01182.x>.
- Davison, H. K., D. H. Kluemper, S. Tao, D. W. Stewart, and M. Bing. 2021. "The Effects of Faking on the Relationship Between Cognitive Ability and Conscientiousness: A Cautionary Note." *International Journal of Selection and Assessment* 29, no. 2: 203–218. <https://doi.org/10.1111/ijsa.12319>.
- Diekmann, J., and C. J. König. 2015. "Personality Testing in Personnel-Selection: Love it? Leave it? Understand it!" In *Employee Recruitment, Selection, and Assessment: Contemporary Issues for Theory and Practice*, edited by I. Nikolaou and J. K. Oostrom, 117–135. Routledge/Taylor & Francis Group.
- Dragulescu, A., and C. Arendt. 2020. *xlsx: Read, Write, Format Excel 2007 and Excel 97/2000/XP/2003 Files*. R package version 0.6.5, <https://CRAN.R-project.org/package=xlsx>.
- Edwards, J. R. 2019. "Response Invalidity in Empirical Research: Causes, Detection, and Remedies." *Journal of Operations Management* 65, no. 1: 62–76. <https://doi.org/10.1016/j.jom.2018.12.002>.
- Ellingson, J. E., and L. A. McFarland. 2011. "Understanding Faking Behavior Through the Lens of Motivation: An Application of VIE Theory." *Human Performance* 24, no. 4: 322–337. <https://doi.org/10.1080/08959285.2011.597477>.
- Faul, F., E. Erdfelder, A. Buchner, and A.-G. Lang. 2009. "Statistical Power Analyses Using G\*Power 3.1: Tests for Correlation and Regression Analyses." *Behavior Research Methods* 41: 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>.
- Fisher, D. M., S. Cunningham, A. J. Kerr, and S. P. Allscheid. 2017. "Contextualized Personality Measures in Employee Selection: Extending Frame-of-Reference Research With Job Applicant Samples." *International Journal of Selection and Assessment* 25, no. 1: 18–35. <https://doi.org/10.1111/ijsa.12156>.



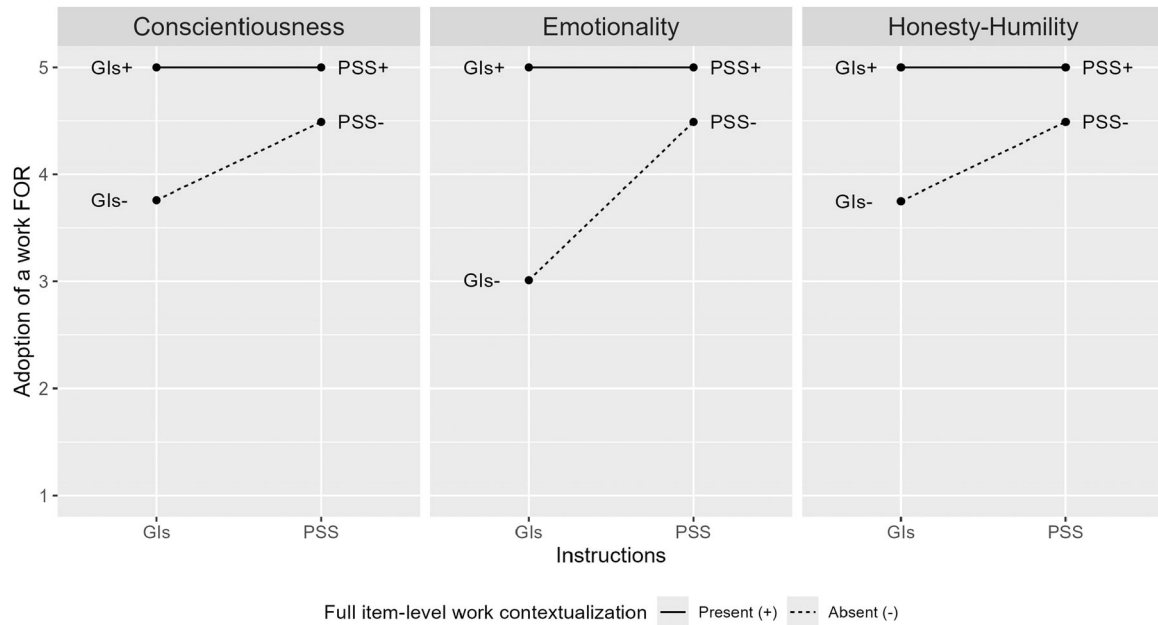
- Fox, J., and S. Weisberg. 2019. *An R Companion to Applied Regression*, 3rd edition. Sage.
- Furnham, A. 1990. "Faking Personality Questionnaires: Fabricating Different Profiles for Different Purposes." *Current Psychology* 9: 46–55. <https://doi.org/10.1007/BF02686767>.
- Geiger, M., R. Bärwaldt, and O. Wilhelm. 2021. "The Good, the Bad, and the Clever: Faking Ability as a Socio-Emotional Ability?" *Journal of Intelligence* 9, no. 1: 13. <https://doi.org/10.3390/jintelligence9010013>.
- Geiger, M., S. Olderbak, R. Sauter, and O. Wilhelm. 2018. "The 'G' in Faking: Doublethink the Validity of Personality Self-Report Measures for Applicant Selection." *Frontiers in Psychology* 9: 1–15. <https://doi.org/10.3389/fpsyg.2018.02153>.
- Goffin, R. D., and A. C. Boyd. 2009. "Faking and Personality Assessment in Personnel Selection: Advancing Models of Faking." *Canadian Psychology/Psychologie Canadienne* 50, no. 3: 151–160. <https://doi.org/10.1037/a0015946>.
- Griffith, R. L., T. Chmielowski, and Y. Yoshita. 2007. "Do Applicants Fake? An Examination of the Frequency of Applicant Faking Behavior." *Personnel Review* 36, no. 3: 341–355. <https://doi.org/10.1108/00483480710731310>.
- Gummer, T., J. Roßmann, and H. Silber. 2021. "Using Instructed Response Items as Attention Checks in Web Surveys: Properties and Implementation." *Sociological Methods & Research* 50, no. 1: 238–264. <https://doi.org/10.1177/0049124118769083>.
- Harris, A. M., J. T. McMillan, and N. T. Carter. 2021. "Test-Taker Reactions to Ideal Point Measures of Personality." *Journal of Business and Psychology* 36: 513–532. <https://doi.org/10.1007/s10869-020-09682-8>.
- Hartman, N. S., and W. L. Grubb. 2011. "Deliberate Faking on Personality and Emotional Intelligence Measures." *Psychological Reports* 108, no. 1: 120–138. <https://doi.org/10.2466/03.09.28.PR0.108.1.120-138>.
- Hayes, A. F. 2022. *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*, 3rd edition. The Guilford Press. <http://www.guilford.com/p/hayes3>.
- Hayes, A. F. 2024. PROCESS macro for SPSS, SAS, and R (Version 5, Beta) [Computer software]. <https://www.processmacro.org>.
- Hester, J., H. Wickham, and G. Csárdi. 2024. *Fs: Cross-Platform File System Operations Based on 'libuv' (Version 1.6.4)*. [Computer software].
- Holtrop, D., M. P. Born, A. De Vries, and R. E. De Vries. 2014. "A Matter of Context: A Comparison of Two Types of Contextualized Personality Measures." *Personality and Individual Differences* 68: 234–240. <https://doi.org/10.1016/j.paid.2014.04.029>.
- Holtrop, D., M. P. Born, and R. E. De Vries. 2014. "Predicting Performance With Contextualized Inventories, No Frame-of-Reference Effect?" *International Journal of Selection and Assessment* 22, no. 2: 219–223. <https://doi.org/10.1111/ijsa.12071>.
- Holtrop, D., J. K. Oostrom, P. D. Dunlop, and C. Runneboom. 2021. "Predictors of Faking Behavior on Personality Inventories in Selection: Do Indicators of the Ability and Motivation to Fake Predict Faking?" *International Journal of Selection and Assessment* 29, no. 2: 185–202. <https://doi.org/10.1111/ijsa.12322>.
- Holtz, B. C., R. E. Ployhart, and A. Dominguez. 2005. "Testing the Rules of Justice: The Effects of Frame-of-Reference and Pre-Test Validity Information on Personality Test Responses and Test Perceptions." *International Journal of Selection and Assessment* 13: 75–86. <https://doi.org/10.1111/j.0965-075X.2005.00301.x>.
- Hu, J., and B. S. Connelly. 2021. "Faking by Actual Applicants on Personality Tests: A Meta-Analysis of Within-Subjects Studies." *International Journal of Selection and Assessment* 29, no. 3–4: 412–426. <https://doi.org/10.1111/ijsa.12338>.
- Huang, J. L., P. G. Curran, J. Keeney, E. M. Poposki, and R. P. DeShon. 2012. "Detecting and Deterring Insufficient Effort Responding to Surveys." *Journal of Business and Psychology* 27: 99–114. <https://doi.org/10.1007/s10869-011-9231-8>.
- Hunthausen, J. M., D. M. Truxillo, T. N. Bauer, and L. B. Hammer. 2003. "A Field Study of Frame-of-Reference Effects on Personality Test Validity." *Journal of Applied Psychology* 88, no. 3: 545–551. <https://doi.org/10.1037/0021-9010.88.3.545>.
- Iannone, R., J. Cheng, B. Schloerke, et al. 2024. *Gt: Easily Create Presentation-Ready Display Tables (Version 0.11.0)*. [Computer software].
- Jeong, Y. R., N. D. Christiansen, C. Robie, M. C. Kung, and T. B. Kinney. 2017. "Comparing Applicants and Incumbents: Effects of Response Distortion on Mean Scores and Validity of Personality Measures." *International Journal of Selection and Assessment* 25, no. 3: 311–315. <https://doi.org/10.1111/ijsa.12182>.
- Kam, C. C. S., and G. H. Chan. 2018. "Examination of the Validity of Instructed Response Items in Identifying Careless Respondents." *Personality and Individual Differences* 129: 83–87. <https://doi.org/10.1016/j.paid.2018.03.022>.
- Kassambara, A. 2023. rstatix: Pipe-friendly Framework for basic statistical tests. R package version 0.7.2. <https://CRAN.R-project.org/package=rstatix>.
- König, C. J., A. S. Merz, and N. Trauffer. 2012. "What is in Applicants' Minds When They Fill out a Personality Test? Insights From a Qualitative Study." *International Journal of Selection and Assessment* 20, no. 4: 442–452. <https://doi.org/10.1111/ijsa.12007>.
- Koo, T. K., and M. Y. Li. 2016. "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research." *Journal of Chiropractic Medicine* 15, no. 2: 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>.
- Lambert, C. E., S. A. Arbuckle, and R. R. Holden. 2016. "The Marlowe-Crowne Social Desirability Scale Outperforms the BIDR Impression Management Scale for Identifying Fakers." *Journal of Research in Personality* 61: 80–86. <https://doi.org/10.1016/j.jrp.2016.02.004>.
- Lanz, L., I. Thielmann, and F. H. Gerpott. 2022. "Are Social Desirability Scales Desirable? A Meta-Analytic Test of the Validity of Social Desirability Scales in the Context of Prosocial Behavior." *Journal of Personality* 90, no. 2: 203–221. <https://doi.org/10.1111/jopy.12662>.
- Lawrence, M. A. 2016. ez: Easy Analysis and Visualization of Factorial Experiments. R package version 4.4-0. <https://CRAN.R-project.org/package=ez>.
- Lee, K., and M. C. Ashton. 2004. "Psychometric Properties of the HEXACO Personality Inventory." *Multivariate Behavioral Research* 39, no. 2: 329–358. [https://doi.org/10.1207/s15327906mbr3902\\_8](https://doi.org/10.1207/s15327906mbr3902_8).
- Lee, K., M. C. Ashton, J. A. Pozzebon, B. A. Visser, J. S. Bourdage, and B. Ogunfowora. 2009. "Similarity and Assumed Similarity in Personality Reports of Well-Acquainted Persons." *Journal of Personality and Social Psychology* 96, no. 2: 460–472. <https://doi.org/10.1037/a0014059>.
- Lee, K., M. C. Ashton, and R. E. De Vries. 2005. "Predicting Workplace Delinquency and Integrity With the HEXACO and Five-Factor Models of Personality Structure." *Human Performance* 18, no. 2: 179–197. [https://doi.org/10.1207/s15327043hup1802\\_4](https://doi.org/10.1207/s15327043hup1802_4).
- Lee, P., S. H. Joo, and S. Lee. 2019. "Examining Stability of Personality Profile Solutions Between Likert-Type and Multidimensional Forced Choice Measure." *Personality and Individual Differences* 142: 13–20. <https://doi.org/10.1016/j.paid.2019.01.022>.
- Levashina, J., and M. A. Campion. 2006. "A Model of Faking Likelihood in the Employment Interview." *International Journal of Selection and Assessment* 14, no. 4: 299–316. <https://doi.org/10.1111/j.1468-2389.2006.00353.x>.
- Lievens, F., W. De Corte, and E. Schollaert. 2008. "A Closer Look at the Frame-of-Reference Effect in Personality Scale Scores and Validity."

- Journal of Applied Psychology* 93, no. 2: 268–279. <https://doi.org/10.1037/0021-9010.93.2.268>.
- Liu, J., and J. Zhang. 2020. “An Item-Level Analysis for Detecting Faking on Personality Tests: Appropriateness of Ideal Point Item Response Theory Models.” *Frontiers in Psychology* 10: 3090. <https://doi.org/10.3389/fpsyg.2019.03090>.
- MacCann, C., N. Pearce, and Y. Jiang. 2017. “The General Factor of Personality is Stronger and More Strongly Correlated With Cognitive Ability Under Instructed Faking.” *Journal of Individual Differences* 38, no. 1: 46–54. <https://doi.org/10.1027/1614-0001/a000221>.
- Mahar, D., J. Cologon, and J. Duck. 1995. “Response Strategies When Faking Personality Questionnaires in a Vocational Selection Setting.” *Personality and Individual Differences* 18, no. 5: 605–609. [https://doi.org/10.1016/0191-8869\(94\)00200-C](https://doi.org/10.1016/0191-8869(94)00200-C).
- 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, no. 4: 417–430. <https://doi.org/10.1111/j.1468-2389.2009.00483.x>.
- McNeish, D., L. M. Stapleton, and R. D. Silverman. 2017. “On the Unnecessary Ubiquity of Hierarchical Linear Modeling.” *Psychological Methods* 22, no. 1: 114–140. <https://doi.org/10.1037/met0000078>.
- Mueller-Hanson, R., E. D. Heggstad, and G. C. Thornton. 2003. “Faking and Selection: Considering the Use of Personality From Select-In and Select-Out Perspectives.” *Journal of Applied Psychology* 88, no. 2: 348–355. <https://doi.org/10.1037/0021-9010.88.2.348>.
- Orne, M. T. 1962. “On the Social Psychology of the Psychological Experiment: With Particular Reference to Demand Characteristics and Their Implications.” *American Psychologist* 17, no. 11: 776–783. <https://doi.org/10.1037/h0043424>.
- Pathki, C. S., D. H. Klumper, J. D. Meuser, and B. D. McLarty. 2022. “The org-B5: Development of a Short Work Frame-of-Reference Measure of the Big Five.” *Journal of Management* 48, no. 5: 1299–1337. <https://doi.org/10.1177/01492063211002627>.
- Pedersen, T. L. 2024. *Patchwork: The Composer of Plots (Version 1.2.0)*. [Computer software].
- R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>.
- Rasheed, S., and C. Robie. 2023. “Faking Resistance of a Quasi-Ipsative RIASEC Occupational Interest Measure.” *International Journal of Selection and Assessment* 31: 321–335. <https://doi.org/10.1111/ijsa.12427>.
- Revelle, W. 2023. *psych: Procedures for Personality and Psychological Research*, Version 2.3.9. Northwestern University, Evanston, IL. <https://CRAN.R-project.org/package=psych>.
- Robie, C. 2001. “Faking and Personality Measurement.” Unpublished Technical Report. Personnel Decisions International.
- Robie, C., M. P. Born, and M. J. Schmit. 2001. “Personal and Situational Determinants of Personality Responses: A Partial Reanalysis and Reinterpretation of the Schmit et al. (1995) Data.” *Journal of Business and Psychology* 16, no. 1: 101–117. <https://doi.org/10.1023/A:1007843906550>.
- Robie, C., S. D. Risavy, D. Holtrop, and M. P. Born. 2017. “Fully Contextualized, Frequency-Based Personality Measurement: A Replication and Extension.” *Journal of Research in Personality* 70: 56–65. <https://doi.org/10.1016/j.jrp.2017.05.005>.
- Robie, C., M. J. Schmit, A. M. Ryan, and M. J. Zickar. 2000. “Effects of Item Context Specificity on the Measurement Equivalence of a Personality Inventory.” *Organizational Research Methods* 3, no. 4: 348–365. <https://doi.org/10.1177/109442810034003>.
- Roess, M., and M. Roche. 2017. “Job Desire and Response Distortion in Personality Assessments.” *E-Journal of Social & Behavioural Research in Business* 8, no. 2: 1–15. Retrieved from <https://hdl.handle.net/10289/11605>.
- Röhner, J., and T. Ewers. 2016. “How to Analyze (Faked) Implicit Association Test Data by Applying Diffusion Model Analyses With the Fast-Dm Software: A Companion to Röhner & Ewers (2016).” *Quantitative Methods in Psychology* 12: 220–231. <https://doi.org/10.20982/tqmp.12.3.p220>.
- Röhner, J., and R. R. Holden. 2022. “Challenging Response Latencies in Faking Detection: The Case of Few Items and No Warnings.” *Behavior Research Methods* 54: 324–333. <https://doi.org/10.3758/s13428-021-01636-z>.
- Röhner, J., R. R. Holden, and A. Schütz. 2023. “IAT Faking Indices Revisited: Aspects of Replicability and Differential Validity.” *Behavior Research Methods* 55: 670–693. <https://doi.org/10.3758/s13428-022-01845-0>.
- Röhner, J., and A. Schütz. 2020. “Verfälschungsverhalten in Psychologischer Diagnostik.” *Report Psychologie* 9: 16–23. <https://www.psychologenverlag.de/Produkte/pid/781/>.
- Röhner, J., and A. Schütz. 2019. “Faking Behavior.” In *Encyclopedia of Personality and Individual Differences*, edited by V. Zeigler-Hill and T. Shackelford, 1553–1558. Springer. [https://doi.org/10.1007/978-3-319-28099-8\\_2341-1](https://doi.org/10.1007/978-3-319-28099-8_2341-1).
- Röhner, J., A. Schütz, and M. Ziegler. 2025. “Faking in Self-Report Personality Scales: A Qualitative Analysis and Taxonomy of the Behaviors That Constitute Faking Strategies.” *International Journal of Selection and Assessment* 33: e12513. <https://doi.org/10.1111/ijsa.12513>.
- Röhner, J., and P. Thoss. 2018. “EZ: An Easy Way to Conduct a More Fine-Grained Analysis of Faked and Nonfaked Implicit Association Test (IAT) Data.” *Quantitative Methods for Psychology* 14, no. 1: 17–37. <https://doi.org/10.20982/tqmp.14.1.p017>.
- Röhner, J., P. Thoss, and A. Schütz. 2022. “Lying on the Dissection Table: Anatomizing Faked Responses.” *Behavior Research Methods* 54: 2878–2904. <https://doi.org/10.3758/s13428-021-01770-8>.
- Röhner, J., P. Thoss, and L. Uziel. 2024. “Can People With Higher Versus Lower Scores on Impression Management or Self-Monitoring be Identified Through Different Traces Under Faking?” *Educational and Psychological Measurement* 84, no. 3: 594–631. <https://doi.org/10.1177/00131644231182598>.
- Roulin, N., and F. Krings. 2020. “Faking to Fit in: Applicants’ Response Strategies to Match Organizational Culture.” *Journal of Applied Psychology* 105, no. 2: 130–145. <https://doi.org/10.1037/apl0000431>.
- Roulin, N., F. Krings, and S. Binggeli. 2016. “A Dynamic Model of Applicant Faking.” *Organizational Psychology Review* 6, no. 2: 145–170. <https://doi.org/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, no. 3: 209–228. <https://doi.org/10.1111/ijsa.12142>.
- Salgado, J. F., and M. Lado. 2018. “Faking Resistance of a Quasi-Ipsative Forced-Choice Personality Inventory Without Algebraic Dependence.” *Revista de Psicología del Trabajo y de las Organizaciones* 34, no. 3: 213–216. <https://doi.org/10.5093/jwop2018a23>.
- Sassenberg, K., and L. Ditrich. 2019. “Research in Social Psychology Changed Between 2011 and 2016: Larger Sample Sizes, More Self-Report Measures, and More Online Studies.” *Advances in Methods and Practices in Psychological Science* 2, no. 2: 107–114. <https://doi.org/10.1177/2515245919838781>.
- Schlotzhauer, A. E., M. A. Ng, and S. Su. 2024. “How to Frame the Frame of Reference: A Comparison of Contextualization Methods.”

- Journal of Business and Psychology* 40, no. 2: 385–403. <https://doi.org/10.1007/s10869-024-09953-8>.
- Schmidt-Atzert, L., and M. Amelang. 2012. “Diagnostische Verfahren.” In *Psychologische Diagnostik*, 5th edition, 175–357. Springer. [https://doi.org/10.1007/978-3-642-17001-0\\_3](https://doi.org/10.1007/978-3-642-17001-0_3).
- Schmidt-Atzert, L., M. Bühner, S. Rischen, and V. Warkentin. 2004. “Erkennen von Simulation und Dissimulation im Test d2.” *Diagnostica* 50, no. 3: 124–133. <https://doi.org/10.1026/0012-1924.50.3.124>.
- Schmit, M. J., A. M. Ryan, S. L. Stierwalt, and A. B. Powell. 1995. “Frame-of-Reference Effects on Personality Scale Scores and Criterion-Related Validity.” *Journal of Applied Psychology* 80, no. 5: 607–620. <https://doi.org/10.1037/0021-9010.80.5.607>.
- Schmitt, N., and J. C. Bradburn. 2018. “An Ideal Student Factor and the Validity of Noncognitive Measures of Student Potential.” *International Journal of Selection and Assessment* 26, no. 2–4: 109–123. <https://doi.org/10.1111/ijsa.12221>.
- Schulze, J., S. G. West, J. P. Freudenstein, et al. 2021. “Hidden Framings and Hidden Asymmetries in the Measurement of Personality—A Combined Lens-Model and Frame-of-Reference Perspective.” *Journal of Personality* 89, no. 2: 357–375. <https://doi.org/10.1111/jopy.12586>.
- Shaffer, J. A., and B. E. Postlethwaite. 2012. “A Matter of Context: A Meta-Analytic Investigation of the Relative Validity of Contextualized and Noncontextualized Personality Measures.” *Personnel Psychology* 65, no. 3: 445–494. <https://doi.org/10.1111/j.1744-6570.2012.01250.x>.
- Shrout, P. E., and J. L. Fleiss. 1979. “Intraclass Correlations: Uses in Assessing Rater Reliability.” *Psychological Bulletin* 86, no. 2: 420–428. <https://doi.org/10.1037/0033-2909.86.2.420>.
- Sisco, H., and R. R. Reilly. 2010. “Effect of Item Level Social Desirability on Factor Structure Stability.” *International Journal of Selection and Assessment* 18, no. 3: 305–309. <https://doi.org/10.1111/j.1468-2389.2010.00513.x>.
- Slowikowski, K. 2024. *ggrepel: Automatically Position Non-Overlapping Text Labels With 'ggplot2'* (Version 0.9.6) [R package]. <https://CRAN.R-project.org/package=ggrepel>.
- Stewart, G. L., T. C. Darnold, R. D. Zimmerman, L. Parks, and S. L. Dustin. 2010. “Exploring How Response Distortion of Personality Measures Affects Individuals.” *Personality and Individual Differences* 49, no. 6: 622–628. <https://doi.org/10.1016/j.paid.2010.05.035>.
- Swift, V., and J. B. Peterson. 2019. “Contextualization as a Means to Improve the Predictive Validity of Personality Models.” *Personality and Individual Differences* 144: 153–163. <https://doi.org/10.1016/j.paid.2019.03.007>.
- Teige-Mocigemba, S., B. Penzl, M. Becker, L. Henn, and K. C. Klauer. 2015. “Controlling the ‘Uncontrollable’: Faking Effects on the Affect Misattribution Procedure.” *Cognition and Emotion* 30, no. 8: 1470–1484. <https://doi.org/10.1080/02699931.2015.1070793>.
- Terry, D. J., M. A. Hogg, and K. M. White. 1999. “The Theory of Planned Behaviour: Self-Identity, Social Identity and Group Norms.” *British Journal of Social Psychology* 38, no. 3: 225–244. <https://doi.org/10.1348/014466699164149>.
- Tett, R., and D. Simonet. 2021. “Applicant Faking on Personality Tests: Good or Bad and Why Should We Care?” *Personnel Assessment and Decisions* 7, no. 1: 6–19. <https://doi.org/10.25035/pad.2021.01.002>.
- Tett, R. P., and D. V. Simonet. 2011. “Faking in Personality Assessment: A ‘Multisaturation’ Perspective on Faking as Performance.” *Human Performance* 24, no. 4: 302–321. <https://doi.org/10.1080/08959285.2011.597472>.
- Tonković, M., Ž. Jernei, and Z. Galić. 2012. “The Effects of Faking on the Construct Validity of Personality Questionnaires: A Direct Faking Measure Approach.” *Psychological Topics* 21, no. 3: 455–472.
- Vaughan, D., and M. Dancho. 2022. *Furrr: Apply Mapping Functions in Parallel Using Futures (Version 0.3.1)*. [Computer software].
- Viswesvaran, C., and D. S. Ones. 1999. “Meta-Analyses of Fakability Estimates: Implications for Personality Measurement.” *Educational and Psychological Measurement* 59, no. 2: 197–210. <https://doi.org/10.1177/00131649921969802>.
- De Vries, R. E., I. Zettler, and B. E. Hilbig. 2014. “Rethinking Trait Conceptions of Social Desirability Scales: Impression Management Aas an Expression of Honesty-Humility.” *Assessment* 21, no. 3: 286–299. <https://doi.org/10.1177/1073191113504619>.
- Waring, E., M. Quinn, A. McNamara, E. Arino de la Rubia, H. Zhu, and S. Ellis. 2022. *Skimr: Compact and Flexible Summaries of Data (Version 2.1.5)*. [Computer software].
- Weiner, I. B., D. K. Freedheim, J. A. Schinka, and W. F. Velicer. 2003. *Handbook of Psychology: Research Methods in Psychology*. Wiley.
- Wickham, H., M. Averick, J. Bryan, et al. 2019. “Welcome to the Tidyverse.” *Journal of Open Source Software* 4, no. 43: 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, H., E. Miller, and D. Smith. 2023. *Haven: Import and Export ‘SPSS’, ‘Stata’ and ‘SAS’ files (Version 2.5.4)*. [Computer software].
- Wickham, H., T. L. Pedersen, and D. Seidel. 2023. *Scales: Scale Functions for Visualization (Version 1.3.0)*. [Computer software].
- Ziegler, M., and D. Bensch. 2015. “Personality Assessment, faking and.” In *International Encyclopedia of the Social & Behavioral Sciences*, edited by J. D. Wright, 17, 2nd ed., 834–839. <https://doi.org/10.1016/B978-0-08-097086-8.25083-6>.
- Ziegler, M., C. MacCann, and R. D. Roberts. 2012. “Faking. Knowns, Unknowns, and Points of Contention.” In *New Perspectives on Faking in Personality Assessment*, edited by M. Ziegler, C. MacCann, and R. Roberts, 3–16. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195387476.003.0011>.

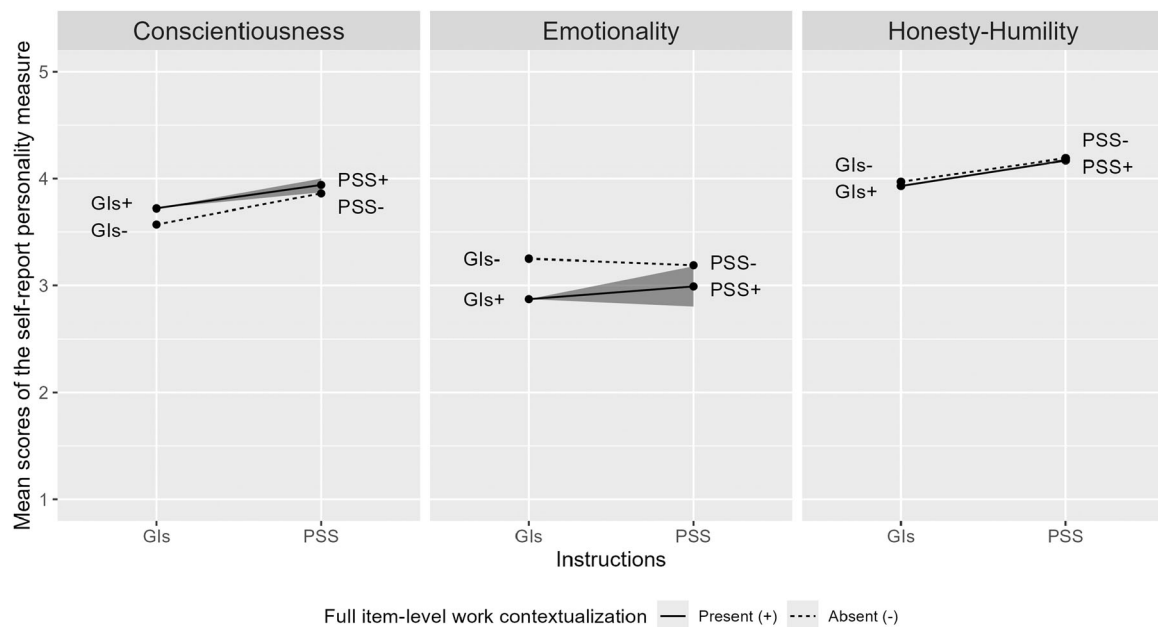
## Appendix

### A1: Expected Effects of Instructions and Full Item-Level Work Contextualization on the Adoption of a Work FOR by Experimental Condition and Personality Trait



Note: GlS- (generic instructions, full item-level work contextualization absent), GlS+ (generic instructions, full item-level work contextualization present), PSS- (personnel-selection scenario, full item-level work contextualization absent), PSS+ (personnel-selection scenario, full item-level work contextualization present). The adoption of a work FOR was measured with a 5-point Likert scale (1 = *not at all* to 5 = *very much*). We determined the values for GlS+ and PSS+, assuming that full item-level work contextualization would have effects on all items by eliciting a stable, clearly pronounced work FOR. For PSS-, we took the values from Fisher et al. (2017, Table 5) and averaged them across samples. Because those authors did not differentiate between traits, we used the same values for Conscientiousness, Emotionality, and Honesty-Humility. For GlS-, we used the work FOR mean scores of Conscientiousness, Emotionality, and Honesty-Humility from our pretest.

### A2: Expected Effects of Instructions and Full Item-Level Work Contextualization on the Mean Scores of the Self-Report Personality Measure by Experimental Condition and Personality Trait





*Note:* GIs– (generic instructions, full item-level work contextualization absent), GIs+ (generic instructions, full item-level work contextualization present), PSS– (personnel-selection scenario, full item-level work contextualization absent), PSS+ (personnel-selection scenario, full item-level work contextualization present). The response behavior in the self-report personality measure was measured with a 5-point Likert scale (1 = *not at all* to 5 = *very much*). As rough estimates of GIs– and GIs+, we took the values from Holtrop et al. (2014). To determine PSS–, we interpolated score differences from GIs– to PSS– from previous research and found that Conscientiousness and Honesty-Humility increased by about 8% and Emotionality decreased by about 2% (Anglim et al. 2018; Anglim et al. 2017; Ashton et al. 2020; Holtrop et al. 2021; MacCann et al. 2017). To determine PSS+, we derived approximations from theory and assumed that (a) work contextualization on the item level should be stronger than via instructions (see Holtrop et al. 2014; Lievens et al. 2008; Swift and Peterson 2019) and (b) the PSS elicits faking and a work FOR and thus partly accounts for the adoption of a work FOR. We thus estimated that PSS+ would be relatively higher than PSS– for Conscientiousness and that PSS+ would be relatively lower than PSS– for Emotionality and Honesty-Humility but possibly with smaller differences than for the comparison between GIs– and GIs+ for Conscientiousness, Emotionality, and Honesty-Humility (the area of possible outcomes is shaded gray).

### A3: Demographics

Which gender do you identify with? (Female/Male/Diverse/Prefer not to say)

How old are you?

What is your primary occupation?

### A4: Measure of the Adopted FORs

When giving my responses to these statements...

... I considered my typical behavior in educational settings (high school, college, study groups, etc.).

... I considered my typical behavior in social settings (with friends, family, etc.).

... I considered my typical behavior in work settings (work assignments, interacting with coworkers, etc.).

... I was thinking of my typical behavior in situations where you have to embellish your responses if necessary to achieve your goal (recruitment tests, assessment center, job interview, etc.).

*Note:* Responses were given (pretest and main study) after each page of the self-report personality measure (i.e., after every four items out of a total of 32 items that measured each personality trait) on a 5-point Likert scale (1 = *strongly disagree* to 5 = *strongly agree*).

### A5: Noncontextualized Self-Report Personality Measure

*Note:* The noncontextualized self-report personality measure included the Conscientiousness, Emotionality, and Honesty-Humility scales from the HEXACO-PI-R (Lee and Ashton 2004; Lee et al. 2009). The Dutch, German, and English versions can be retrieved from those authors upon request. Responses were given (pretest and main study) on a 5-point Likert scale (1 = *strongly disagree* to 5 = *strongly agree*).

### A6: Mean Scores, Standard Deviations, and Cronbach's Alphas for the Adopted FORs for GIs– and Each Personality Trait

Scale	Adopted FOR							
	Education		Social		Work		Faking	
	<i>M</i> ( <i>SD</i> )	Cronbach's $\alpha$ [95% CI]	<i>M</i> ( <i>SD</i> )	Cronbach's $\alpha$ [95% CI]	<i>M</i> ( <i>SD</i> )	Cronbach's $\alpha$ [95% CI]	<i>M</i> ( <i>SD</i> )	Cronbach's $\alpha$ [95% CI]
Conscientiousness	2.84 (1.33)	0.99 [0.98, 0.99]	3.85 (0.75)	0.95 [0.92, 0.97]	3.81 (0.88)	0.97 [0.95, 0.98]	2.69 (1.23)	0.98 [0.97, 0.99]
Emotionality	2.66 (1.26)	0.98 [0.97, 0.99]	4.01 (0.77)	0.97 [0.95, 0.98]	3.34 (1.09)	0.97 [0.95, 0.98]	2.62 (1.26)	0.99 [0.98, 0.99]
Honesty-Humility	2.73 (1.25)	0.99 [0.98, 0.99]	3.78 (0.82)	0.95 [0.92, 0.97]	3.50 (0.97)	0.95 [0.92, 0.97]	2.72 (1.27)	0.99 [0.98, 0.99]

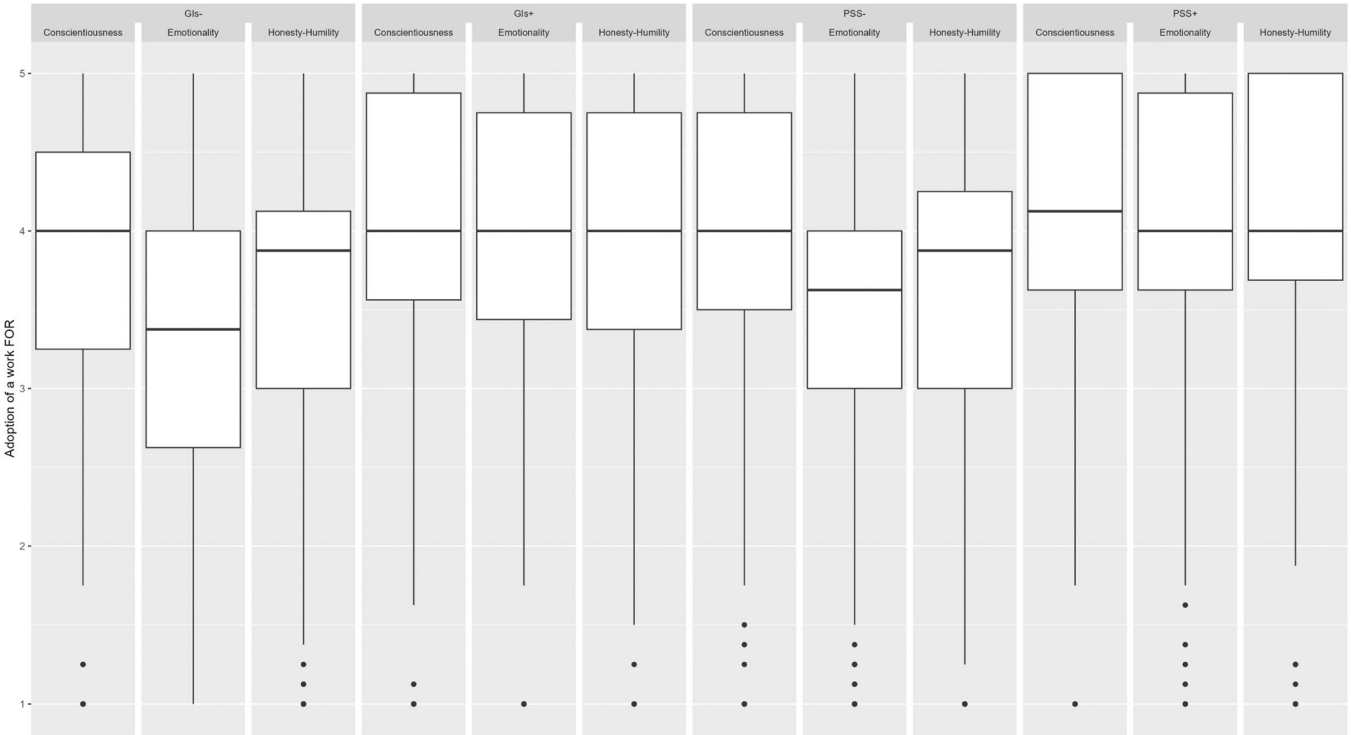
*Note:* The adoption of a work FOR was measured with a 5-point Likert scale (1 = *not at all* to 5 = *very much*). *N* = 51. CI = confidence interval.



A7: Inclusion Criteria

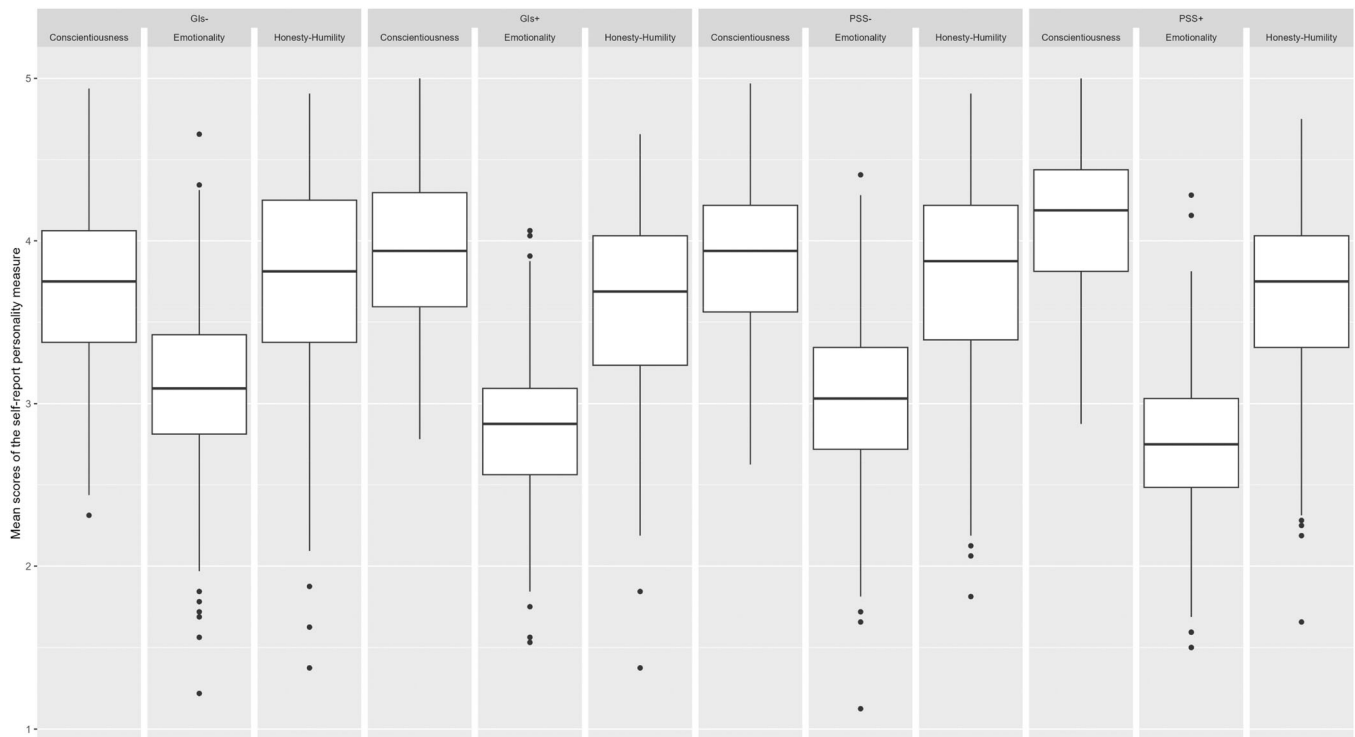
- a. Completeness of data: Participants took part in all four waves of measurement (i.e., there were data for all four conditions)
- b. Careful responding: Participants responded correctly to all attention items (i.e., “I am being attentive and choosing ‘strongly disagree’ as my response to this item,” “I am responding to this item with ‘strongly agree’ because I am reading this item carefully,” “I am responding to this item by choosing the middle category,” “I am reading the instructions carefully, and I am responding to this item with ‘strongly agree,’” “I am choosing ‘strongly disagree’ because this item instructs me to do so”).<sup>22, 23</sup>
- c. Motivation: Participants indicated that they were motivated (i.e., “I was motivated to follow the instructions in every wave of measurement”).<sup>23</sup>
- d. Appropriate testing environment: Individuals indicated that they participated in a quiet environment (i.e., “I completed the questionnaires in a quiet environment”).<sup>23</sup>
- e. Identification with the PSS: Individuals indicated that they could identify with the PSS (i.e., “I could identify with the personnel-selection scenario well”).<sup>23</sup>
- f. Blind to the purpose of the study: Individuals passed the predefined markers for nonblinding (see Table 3; i.e., “Do you have an assumption about the purpose of the study?”).

A8: Box Plots for the Adoption of a Work FOR by Experimental Condition and Personality Trait



Note: GIs- (generic instructions, full item-level work contextualization absent), GIs+ (generic instructions, full item-level work contextualization present), PSS- (personnel-selection scenario, full item-level work contextualization absent), PSS+ (personnel-selection scenario, full item-level work contextualization present). *N* = 309.

#### A9: Box Plots on the Mean Scores of the Self-Report Personality Measure by Experimental Condition and Personality Trait



Note: GIs- (generic instructions, full item-level work contextualization absent), GIs+ (generic instructions, full item-level work contextualization present), PSS- (personnel-selection scenario, full item-level work contextualization absent), PSS+ (personnel-selection scenario, full item-level work contextualization present).  $N = 309$ .

#### A10: Job Description Pharmacy Assistant

Pharmaceutical assistants dispense prescription medicines to customers in accordance with a prescription and under the supervision of a pharmacist. They sell over-the-counter medicines and other pharmaceutical goods. When doing so, they explain the mode of action and application of the product and provide information about possible side effects and storage. Depending on their professional experience, pharmaceutical assistants may also be able to carry out activities without pharmacist supervision. Furthermore, they assist, for example, in the production of ointments and solutions, check the substances according to the specifications in the pharmacopoeia and dose them accurately. They also carry out, for example, simple chemical and physical drug analyses and measurements of customers' cholesterol, urine, or blood values. They monitor product supplies and organize orders. They conduct spot checks as required by law as well as conduct inventory checks and record the issuing and stock of poisons and narcotics. They also compile information for special advisory campaigns (e.g., on the subject of allergies) and participate in social events.<sup>24</sup>

#### A11: Translations of the Fully Work-Contextualized Self-Report Personality Measure

German version	English version
<b>Conscientiousness</b>	
Ich mag es, wenn alle Packungen an der richtigen Stelle einsortiert sind.	I like it when all of the packaging is in the right place.
Ich möchte nie, dass eine Kundin oder ein Kunde unzufrieden nach Hause gehen.	I never want a customer to go home unhappy.
Oft überprüfe ich ein Rezept mehrfach, um Fehler zu vermeiden.	I often check a prescription repeatedly to avoid mistakes.
Ich verliere im Verkaufsraum der Apotheke nie den Überblick, egal wie voll es ist.	I never lose control over the public area of the pharmacy, no matter how busy it gets.
Wenn ich an einem Arbeitsplatz gearbeitet habe, Sorge ich dafür, dass ich alles für die nächste Person sauber und ordentlich hinterlasse.	After working in a certain area, I make sure to leave everything clean and tidy for the next person.
Wenn ich einen Auftrag abwickle, setze ich mich voll ein.	When I process an order, I put my all into it.
Wenn ich eine Bestellung abwickle, ist es mir egal, wie ordentlich ich die Produkte einräume.	When I process an order, I don't care how neatly I put the products away.
Wenn eine Kundin oder ein Kunde ein Problem hat, entscheide ich eher aus dem Bauch heraus als durch sorgfältiges Nachdenken.	If a client has a problem, I make a decision based on a gut feeling rather than careful consideration.

(Continues)

German version	English version
Ich versuche, meine Aufgaben zu Beginn des Tages aufzulisten, damit ich nicht in letzter Minute noch etwas erledigen muss.	I try to list my tasks at the beginning of the day so that nothing is left to the last minute.
Meine Arbeit ist von hoher Qualität, weil ich mich mehr anstrengte als meine Kolleg*innen.	My work is of high quality because I try harder than my colleagues.
Ich versuche immer, alle Kund*innen zufrieden zu stellen, auch wenn es mich zusätzliche Zeit kostet.	I always try to satisfy all customers, even if it takes extra time.
Ich mache viele Fehler, weil ich auf Kund*innen reagiere, ohne nachzudenken.	I make a lot of mistakes due to responding to customers without thinking.
Ich würde den Arbeitsplatz lieber am Ende des Tages aufräumen als zwischendurch.	I would prefer to clean up the workplace at the end of the day, rather than in between.
Meine Kolleg*innen bezeichnen mich manchmal als „Workaholic“.	My colleagues sometimes call me a “workaholic.”
Ich will, dass jedes Rezept perfekt erledigt wird.	I want every prescription done perfectly.
Ich werde niemals ein Medikament ohne Rücksprache herausgeben, wenn ich Zweifel habe.	I will never dispense a medicine without consultation if I have doubts.
Wenn ich mir einen Stift von einer Kollegin oder einem Kollegen ausleihe, gebe ich ihn immer zurück.	When I borrow a pen from a colleague, I always give it back.
Bei der Konsultation mit Ärzt*innen kommt es manchmal vor, dass ich die Diskussion aufgebe.	If there is a disagreement when consulting with a doctor, I sometimes concede.
Ich mag es nicht, Verschreibungen doppelt zu überprüfen, wenn sie richtig zu sein scheinen.	I don't like to double-check prescriptions if they seem right.
Ich denke sorgfältig nach, bevor ich mit gefährlichen Substanzen arbeite.	I think carefully before working with dangerous substances.
Meine Kolleg*innen machen manchmal Witze über meine Unachtsamkeit.	My colleagues sometimes joke about my carelessness.
Ich arbeite nur so viel wie nötig, um nicht von meiner oder meinem Vorgesetzten in der Apotheke auf meine Arbeit angesprochen zu werden.	I work just enough to make sure that my pharmacist doesn't talk to me about my work.
Es ist mir egal, ob das, was ich im Apothekeninformationssystem schreibe, Rechtschreibungs- oder Zeichensetzungsfehler hat.	I don't mind if the notes I type into the pharmacy information system have spelling or punctuation errors.
Normalerweise schaffe ich es, mich zurückzuhalten, bevor ich wütend auf eine Kollegin oder einen Kollegen werde.	I usually manage to hold back before getting angry at a colleague.
Ich vergesse manchmal, Informationen über eine Kundin oder einen Kunden in das Informationssystem einzugeben.	I sometimes forget to enter information about a customer into the information system.
Wenn ein Rezept kompliziert ist, übergebe ich es lieber an eine Kollegin oder einen Kollegen.	If a prescription is very difficult, I prefer to hand it over to a colleague.
Wenn ich die Dosis für ein Medikament berechnen muss, überprüfe ich alles sorgfältig, um sicherzugehen, dass ich keine Fehler mache.	When calculating the dose for a medicine, I check carefully to make sure that there are no mistakes.
Manchmal frage ich eine Kundin oder einen Kunden Dinge, die mich nichts angehen.	Sometimes I ask clients about things that are none of my business.
Manchmal mache ich Fehler bei Rezepten, weil ich schlampig bin.	Sometimes I make mistakes with prescriptions because I'm sloppy.
Ich neige dazu, Fortbildungen lange aufzuschieben, bevor ich sie angehe.	I tend to put off refresher courses as long as possible before starting them.
Selbst wenn ich eine Notiz für meine Kolleg*innen schreibe, lese ich sie noch mal durch, um sicherzugehen, dass darin keine Fehler sind.	Even when writing a note for my colleagues, I read it over to make sure that there are no errors.
Ich folge lieber meinem Instinkt als dem Qualitätshandbuch.	I prefer to follow my instincts rather than the quality manual.
<b>Emotionality</b>	
Ich hätte Angst, wenn es donnert, während ich bei der Arbeit bin.	I would feel afraid if I heard thunder while at work.
Manchmal mache ich mir Sorgen über Kleinigkeiten, die meine Kolleg*innen sagen.	I sometimes can't help but worry about little things my colleagues say.
Ich verlasse mich stark auf meine Kolleg*innen und Vorgesetzte, wenn ich deprimiert bin.	I rely a great deal on my colleagues and supervisors when I feel depressed.

(Continues)

German version	English version
Ich könnte weinen, wenn ich Kund*innen sehe, die weinen.	I feel like crying when I see customers crying.
Selbst wenn ich verletzt bin, gehe ich zur Arbeit.	Even when I'm hurt, I go to work.
Auch in den Pausen denke ich oft an meine Arbeit.	During my break, I continue to think about my work.
Ohne die emotionale Unterstützung durch meine Kolleg*innen und Vorgesetzte fühle ich mich manchmal hilflos.	Without the emotional support of my colleagues and supervisors, I sometimes feel helpless.
Wenn meine Kolleg*innen unglücklich sind, kann ich ihren Schmerz selber spüren.	When my colleagues are unhappy, I can almost feel their pain myself.
Bei der Arbeit habe ich keine Angst vor Infektionen.	I am not afraid of infections at work.
Wäre ich für die ganze Apotheke verantwortlich, würde ich mir vermutlich häufig Sorgen machen.	If I were responsible for the whole pharmacy, I would probably tend to worry a lot.
Wenn ich eine schwierige Kundin oder einen schwierigen Kunden hatte, brauche ich danach eine Person, mit der ich mich darüber unterhalten kann.	When I have had a difficult customer, I need someone to talk with afterwards.
Ich erlebe starke Emotionen, wenn eine von mir geschätzte Kollegin oder ein von mir geschätzter Kollege uns wegen eines Umzugs verlässt.	I feel strong emotions when a colleague I value leaves us because they are moving away.
Es stört mich nicht, mit gefährlichen Substanzen zu arbeiten.	I don't mind working with hazardous substances.
Ich mache mir viel weniger Sorgen als meine Kolleg*innen.	I worry a lot less than my colleagues.
Wenn mir meine Arbeit eine Zeit lang nicht gefällt, werde ich das alleine durchstehen.	If I'm unsatisfied with my work for a while, I will deal with it on my own.
Ich verstehe nicht, warum manche Kolleg*innen bei kranken Kindern so emotional werden.	I don't understand why some colleagues get so emotional about sick children.
Meine Kolleg*innen sagen, dass ich eine furchtlose Person bin.	My colleagues say that I am a fearless person.
Manchmal zweifle ich an meiner Arbeitsleistung, ohne zu wissen, warum.	Sometimes I doubt my work performance without really knowing why.
Wenn ich Meinungsverschiedenheiten mit meiner oder meinem Vorgesetzten habe, mag ich es, Ratschläge von anderen zu erhalten.	When I have disagreements with my superior, I like to get advice from others.
Wenn Kolleg*innen um etwas besorgt sind, bin ich auch besorgt.	When colleagues are concerned about something, I feel concerned too.
Ich würde Kund*innen meiden, wenn ich befürchte, dass sie mir physisch gefährlich werden könnten.	I would avoid clients if I fear that they could become physically dangerous.
Ich mache bei der Arbeit selten, wenn überhaupt, Fehler aus Stress oder Angst vor dem Scheitern.	I rarely, if ever, make mistakes at work due to stress or fear of failure.
Ich kann mit schwierigen Kund*innen umgehen, ohne dass ich emotionale Unterstützung von meinen Kolleg*innen brauche.	I can handle difficult customers without needing emotional support from my colleagues.
Andere sagen manchmal, ich könne mich nicht in Kund*innen einfühlen.	Others sometimes say that it is difficult for me to empathize with customers.
Ich versuche Arbeitsaufgaben zu vermeiden, bei denen ein körperliches oder gesundheitliches Risiko besteht.	As much as possible, I avoid tasks at work that come with a physical or health risk.
Ich neige dazu, ruhig zu bleiben, auch wenn meine Kolleg*innen gestresst sind.	I tend to remain calm even when my colleagues get stressed out.
Wenn ich Streit mit einer Kollegin oder einem Kollegen habe, möchte ich mit jemandem darüber reden.	When I have an argument with a colleague, I would like to talk with someone about it.
Ich werde nicht leicht emotional, selbst in Situationen, in denen meine Kolleg*innen sehr sentimental werden.	I don't become emotional easily, even in situations where my colleagues get very sentimental.
Ich würde auch angesichts eines Raubüberfalls ruhig bleiben.	I would remain calm even in the face of a robbery.
Ich bin sehr nervös, wenn ich mich mit einer Ärztin oder einem Arzt beraten muss.	I get very nervous when I have to consult a doctor.
Ich spreche bei der Arbeit nicht über meine Probleme.	I don't talk about my problems at work.
Wenn ich an ehemalige Arbeitgeber*innen und Kolleg*innen denke, werde ich manchmal ziemlich emotional.	When I think about former employers and colleagues, I sometimes get quite emotional.

#### ***Honesty-Humility***

(Continues)

German version	English version
Wenn ich von Kolleg*innen, die ich nicht mag, etwas will, verhalte ich mich diesen Kolleg*innen gegenüber sehr nett, um es zu bekommen.	If I want something from colleagues who I don't like, I am very nice to them to get it.
Wenn ich wüsste, dass ich niemals erwischt werde, wäre ich bereit, Geld aus der Kasse zu nehmen.	If I knew that I would never get caught, I would be willing to take money from the register.
Es ist mir egal, ob meine Kolleg*innen zu mir aufschauen.	I don't care whether or not my colleagues look up to me.
Ich verdiene mehr Ansehen und Einfluss als meine Kolleg*innen.	I deserve more prestige and influence than my colleagues.
Ich finde überhaupt nichts Verwerfliches daran, Schmeicheleien gegenüber Vorgesetzten einzusetzen, um besser beurteilt zu werden.	I don't see anything wrong with using flattery with superiors so as to gain esteem.
Ich würde niemals eine Ärztin oder einen Arzt falsch informieren, auch wenn es sehr leicht wäre.	I would never lie to a doctor, even if it were very easy.
Viel Geld zu verdienen, ist nicht besonders wichtig für mich.	Earning a lot of money is not especially important to me.
Ich bin wie gewöhnliche Angestellte, nicht besser als andere.	I am an ordinary employee who is no better than others.
Ich versuche manchmal, bei Ärzt*innen oder Assistent*innen Schuldgefühle auszulösen, damit sie tun, was ich will.	I sometimes try to make doctors or assistants feel guilty so that they will do what I want.
Ich hätte keine Probleme damit, Kund*innen etwas zu verkaufen, das sie nicht brauchen.	I wouldn't feel bad about selling customers something that they don't need.
Ich mag es, mit meinen Apotheker*innen und Ärzt*innen befreundet zu sein.	I like being friends with my pharmacist and the doctors.
Ich will nicht, dass meine Apothekerin oder mein Apotheker mich besser behandelt als andere Mitarbeiter*innen.	I don't want my pharmacist to treat me better than other employees.
Ich würde gegenüber Vorgesetzten keine Schmeicheleien benutzen, um eine Gehaltserhöhung zu bekommen oder befördert zu werden, auch wenn ich wüsste, dass es erfolgreich wäre.	I wouldn't use flattery with superiors to get a raise or promotion, even if I thought it would work.
Ich würde in Versuchung geraten, am Arbeitsplatz Dinge zu stehlen, wenn ich knapp bei Kasse wäre.	I would be tempted to steal things at work if I were in a financially tight situation.
Ich helfe gern anständigen, reichen Leuten.	I like to help respectable, well-off people.
Ich bin in allen Aspekten meiner Arbeit besser als meine Kolleg*innen.	I am better than my colleagues in all aspects of my work.
Wenn ich Vorgesetzte bei Laune halten will, lache ich auch noch über deren schlechtesten Witze.	If I want to keep superiors happy, I laugh at their worst jokes.
Ich würde mich auch dann an der Finanzierung von Betriebsausflügen beteiligen, wenn ich wüsste, dass es nicht auffallen würde, wenn ich es ließe.	I would pay my share for company outings even if I knew no one would notice if I didn't.
Ich würde gerne von meinen Kolleg*innen dabei gesehen werden, wie ich mit einem sehr teuren Auto herumfahre.	I would like to be seen by my colleagues driving around in a very expensive car.
Manchmal habe ich das Gefühl, dass ich die Anweisungen meiner Apothekerin bzw. meines Apothekers nicht befolgen muss.	Sometimes I feel like I don't need to follow my pharmacist's instructions.
Ich würde nicht vortäuschen, eine Kollegin oder einen Kollegen zu mögen, nur um diese/n dazu zu bringen, mir Gefälligkeiten zu erweisen.	I wouldn't pretend to like a colleague just to get them to do favors for me.
Ich würde nie jemandem ohne Erlaubnis Betäubungsmittel geben, selbst wenn ich dafür viel Geld bekäme.	I would never give someone narcotics without permission, even if I were paid a lot of money.
Ich würde es genießen, von Vorgesetzten eine Sonderbehandlung zu bekommen.	I would enjoy getting special treatment from supervisors.
Ich denke, dass ich mehr Respekt verdiene als die meisten anderen in meinem Job.	I think I deserve more respect than most other people at my job.
Wenn ich zu einem ungünstigen Zeitpunkt frei haben möchte, frage ich direkt, anstatt zu versuchen, meine Vorgesetzte bzw. meinen Vorgesetzten zu manipulieren.	If I want time off at an inconvenient moment, I ask directly, instead of trying to manipulate my supervisor.
Ich würde gerne wissen, wie ich Informationen über die Medikamente von Personen erhalten kann, ohne dass es jemand erfährt.	I would like to know how I can retrieve information about people's medication without anyone finding out.

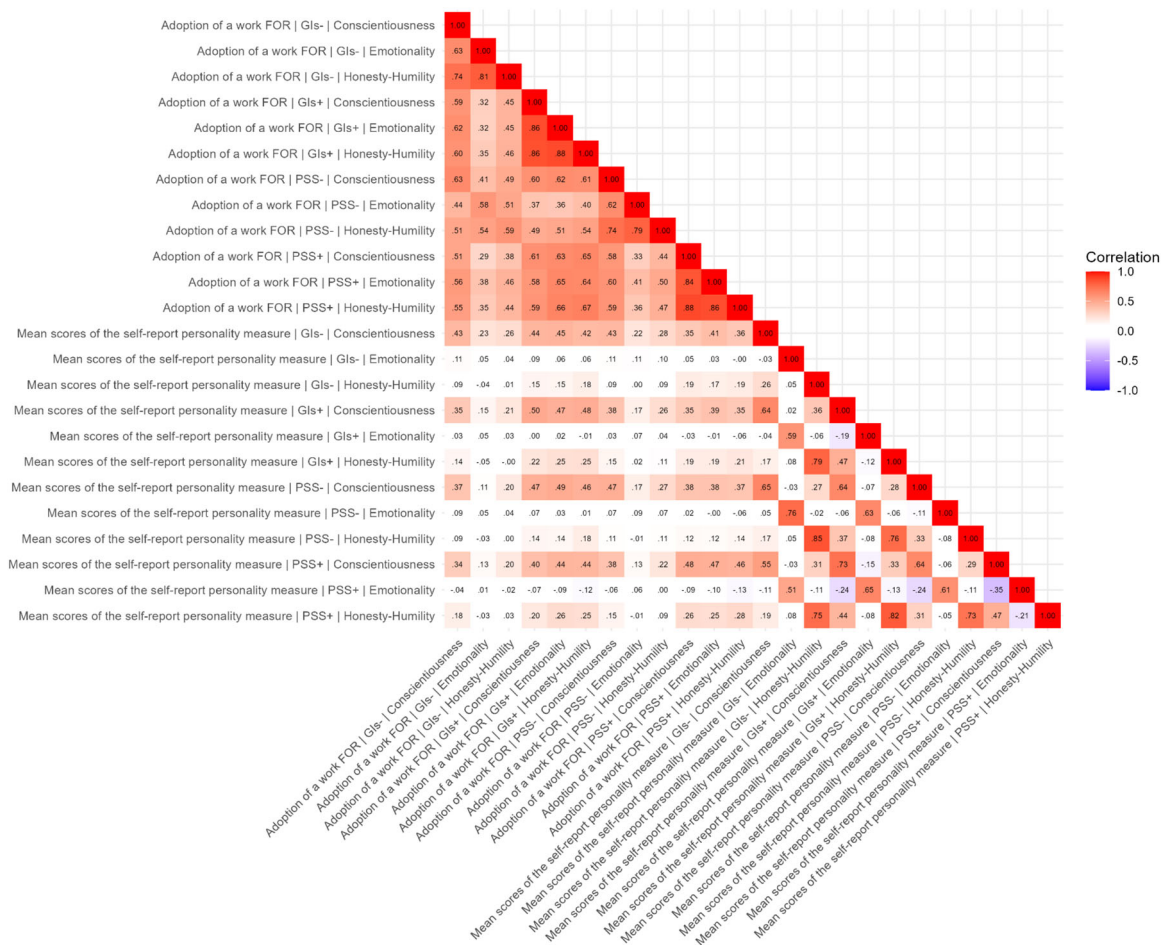
(Continues)



German version	English version
Ich arbeite gerne mit schönen, modernen Geräten.	I like to work with nice, modern equipment.
Manche Kolleg*innen würden sagen, dass ich ein überhöhtes Ego habe.	Some colleagues would say that I have an over-inflated ego.
Ich bringe Kolleg*innen oft dazu, mir einen Gefallen zu tun, indem ich ihnen das Gefühl gebe, sie schulden mir etwas.	I often get colleagues to do favors for me by making them feel that they owe me.
Ich würde in Versuchung geraten, mich öfter krank zu melden, wenn ich sicher sein könnte, damit durchzukommen.	I'd be tempted to call in sick more often if I were sure I could get away with it.
Wenn ich die Chance habe, meine Kolleg*innen dazu zu bringen, zu mir aufzuschauen, nehme ich dafür hohe Risiken in Kauf.	If I have the chance to do something that would get my colleagues to look up to me, I am willing to take big risks.
Ich will, dass die Kund*innen sehen, wie wichtig ich für die Apotheke bin.	I want the customers to see how important I am to the pharmacy.

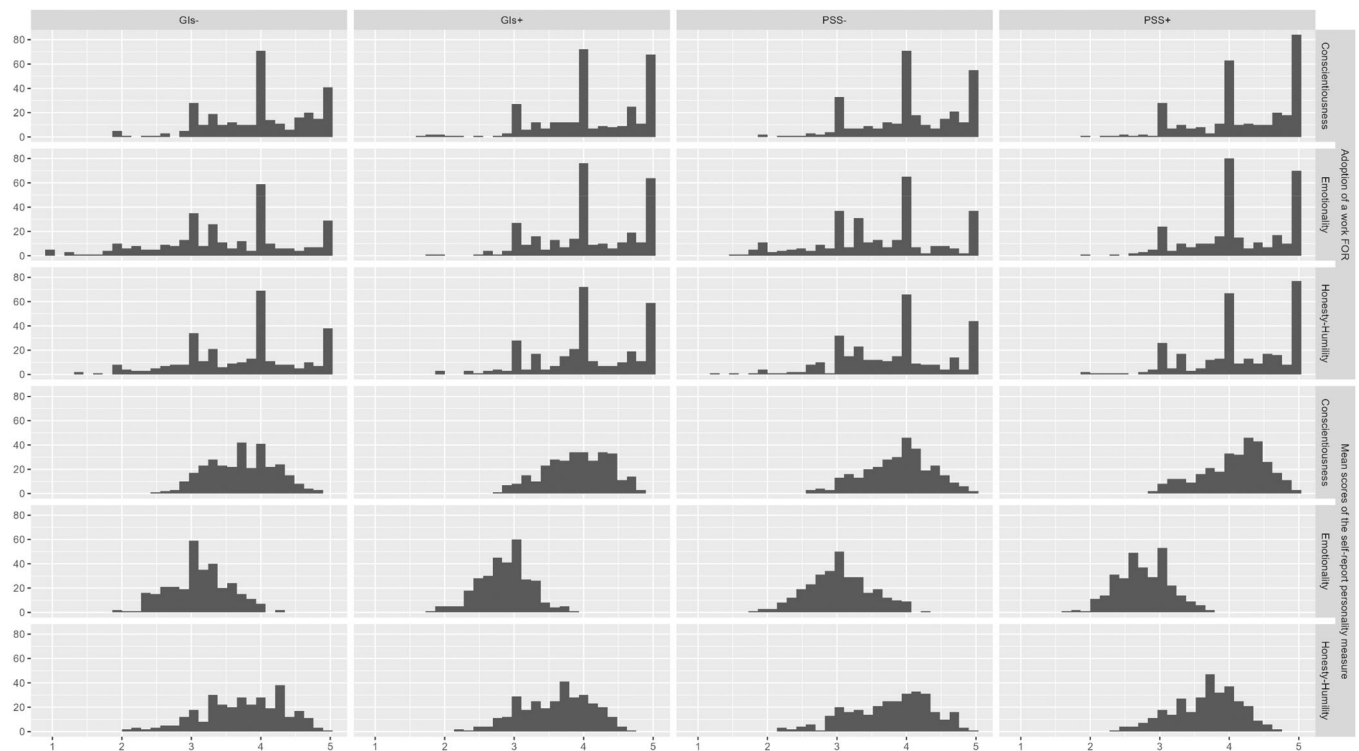
*Note:* The fully work-contextualized self-report personality measure included the adapted Conscientiousness, Emotionality, and Honesty-Humility scales of the HEXACO-PI-R (Lee and Ashton 2004; Lee et al. 2009). The Dutch version of this measure can be requested from its authors (i.e., Holtrop et al. 2014). The English and German versions, translated by the authors of the current manuscript, are presented here. Responses were given on a 5-point Likert scale (1 = *strongly disagree* to 5 = *strongly agree*).

#### A12: Correlations Between the Adoption of a Work FOR and the Mean Scores of the Self-Report Personality Measure by Experimental Condition and Personality Trait



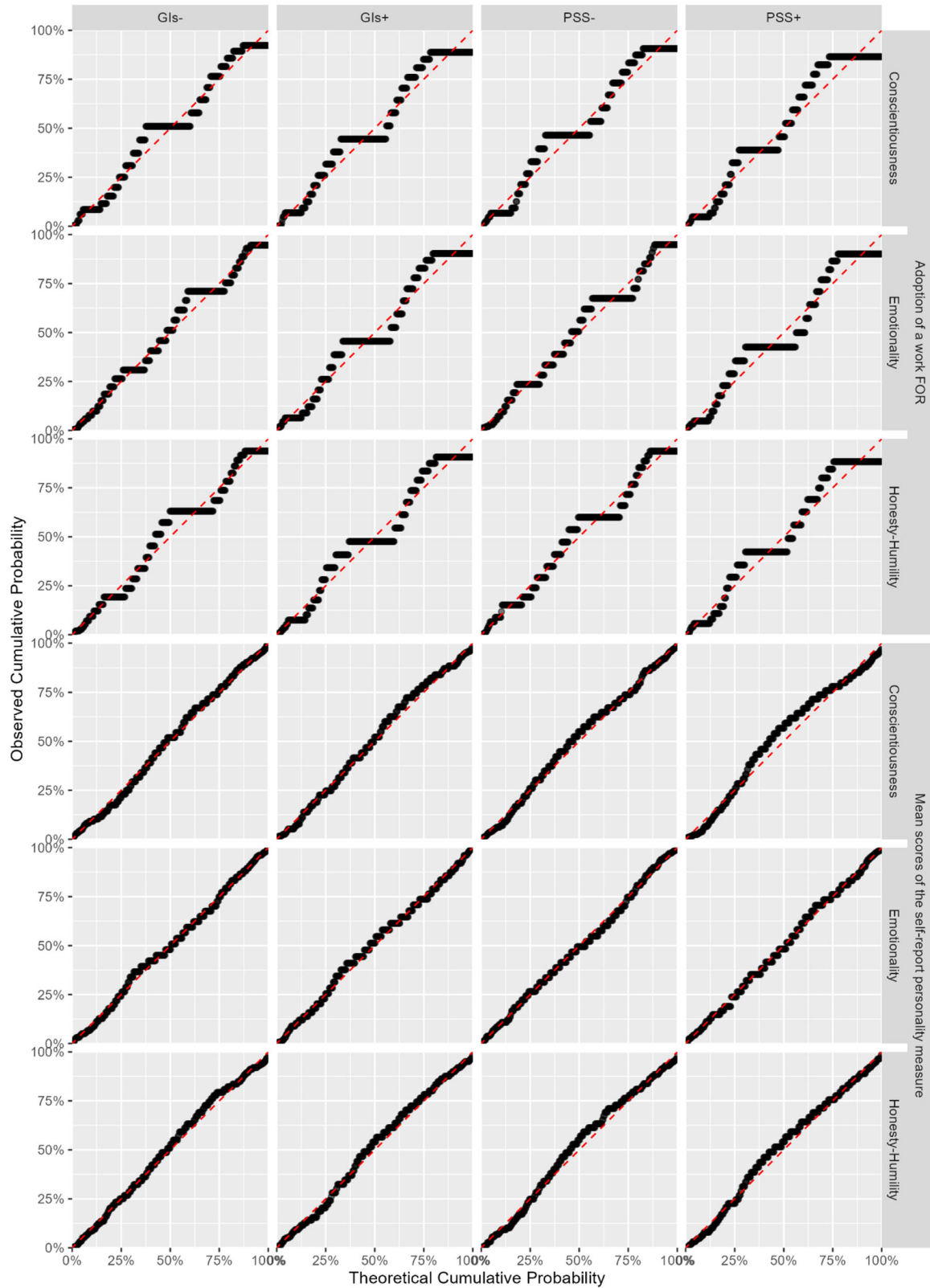
*Note:* GIs– (generic instructions, full item-level work contextualization absent), GIs+ (generic instructions, full item-level work contextualization present), PSS– (personnel-selection scenario, full item-level work contextualization absent), PSS+ (personnel-selection scenario, full item-level work contextualization present).  $N = 309$ .

**A13: Histograms for the Adoption of a Work FOR and the Mean Scores of the Self-Report Personality Measure by Experimental Condition and Personality Trait**



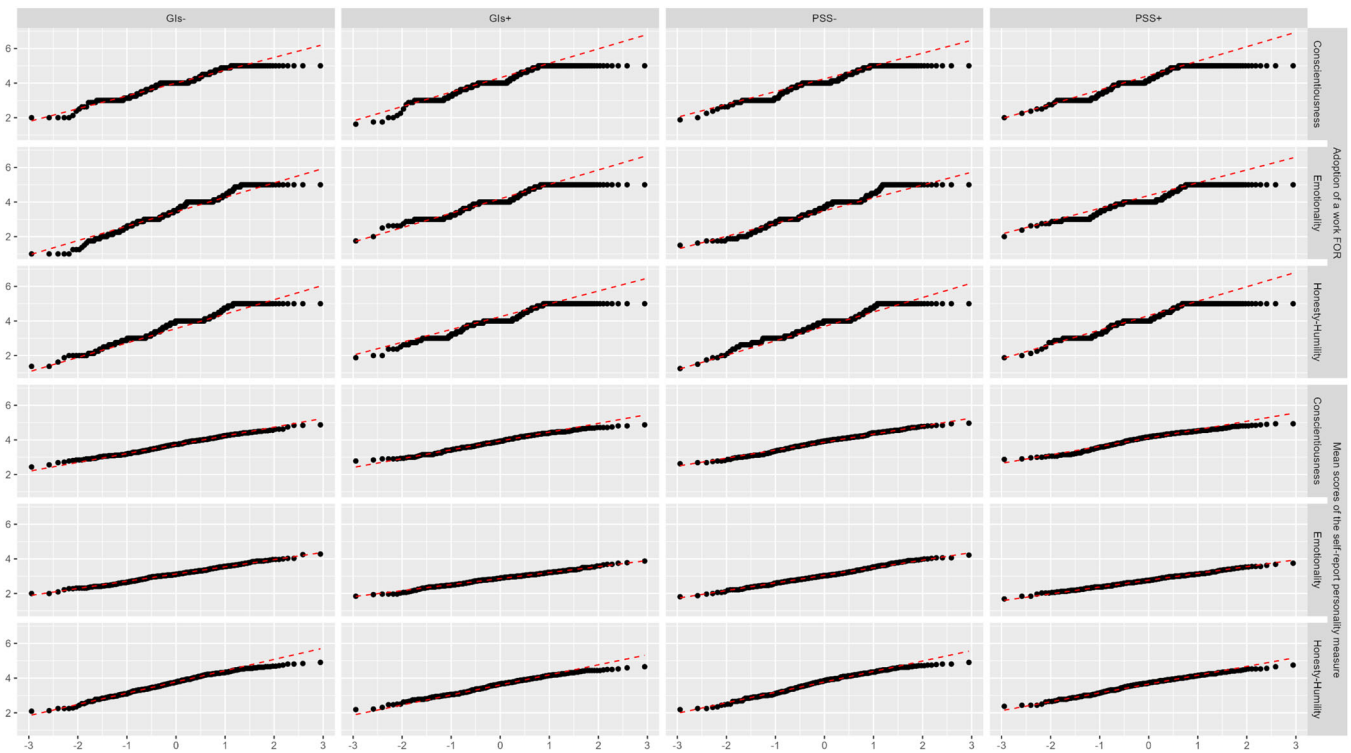
*Note:* GIs- (generic instructions, full item-level work contextualization absent), GIs+ (generic instructions, full item-level work contextualization present), PSS- (personnel-selection scenario, full item-level work contextualization absent), PSS+ (personnel-selection scenario, full item-level work contextualization present).  $N = 309$ .

**A14: PP-Plots for the Adoption of a Work FOR and the Mean Scores of the Self-Report Personality Measure by Experimental Condition and Personality Trait**



*Note:* GIs- (generic instructions, full item-level work contextualization absent), GIs+ (generic instructions, full item-level work contextualization present), PSS- (personnel-selection scenario, full item-level work contextualization absent), PSS+ (personnel-selection scenario, full item-level work contextualization present).  $N = 309$ .

A15: QQ-Plots for the Adoption of a Work FOR and the Mean Scores of the Self-Report Personality Measure by Experimental Condition and Personality Trait



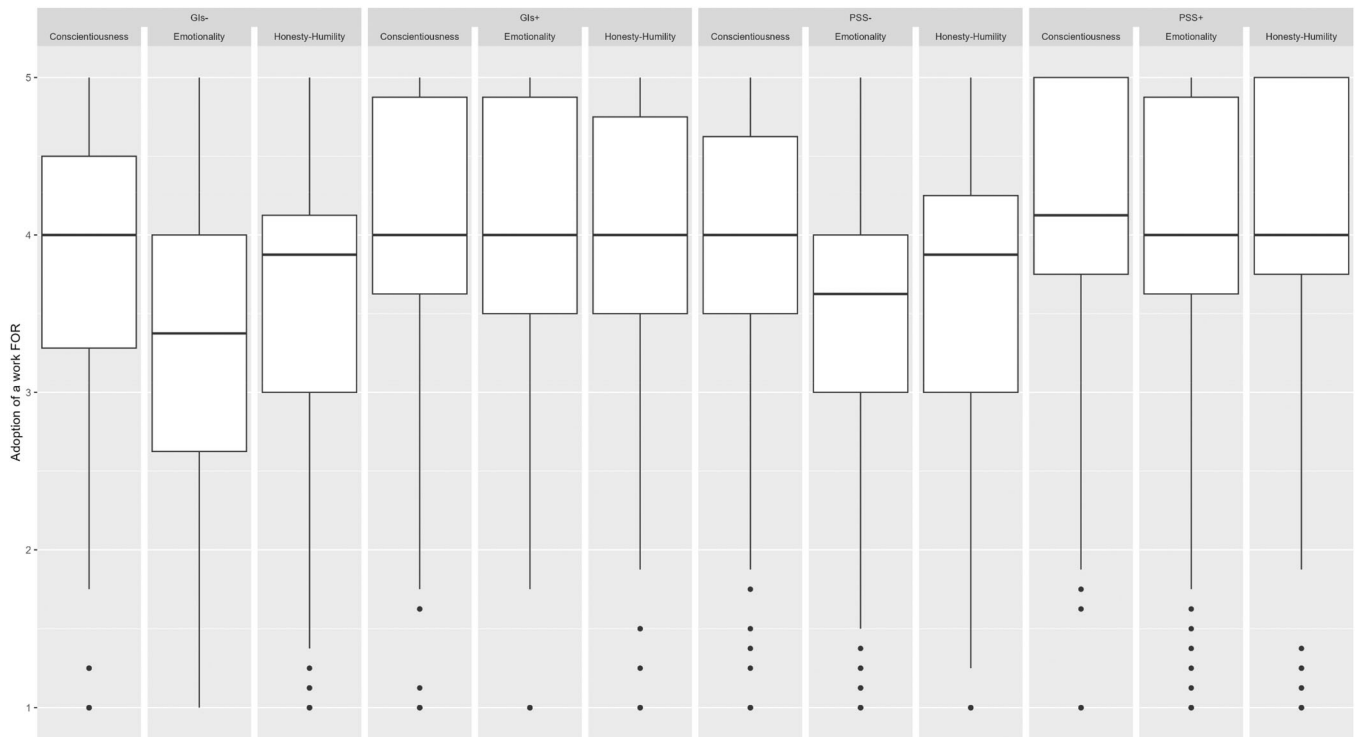
Note: GIs- (generic instructions, full item-level work contextualization absent), GIs+ (generic instructions, full item-level work contextualization present), PSS- (personnel-selection scenario, full item-level work contextualization absent), PSS+ (personnel-selection scenario, full item-level work contextualization present). N = 309.

A16: Results of Levene's Tests in the Two-Way Repeated-Measures ANOVAs for the Adoption of a Work FOR and the Mean Scores of the Self-Report Personality Measure for Each Personality Trait

Adoption of a work FOR				
Scale	df1	df2	F	p
Conscientiousness	3	1232	0.98	0.403
Emotionality	3	1232	18.03	< 0.001
Honesty-Humility	3	1232	2.80	0.039
Mean scores of the self-report personality measure				
Scale	df1	df2	F	p
Conscientiousness	3	1232	0.32	0.813
Emotionality	3	1232	6.21	< 0.001
Honesty-Humility	3	1232	7.15	< 0.001

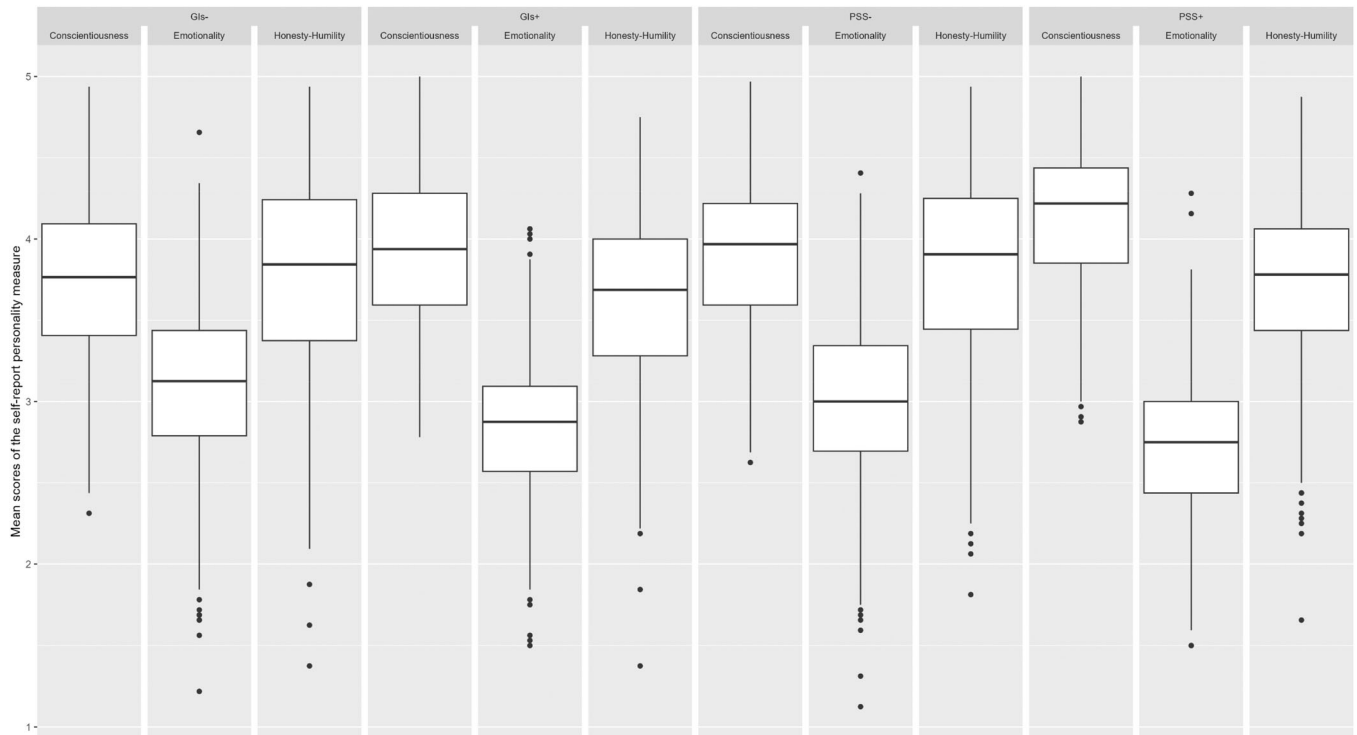
Note: N = 309.

**A17: Box Plots for the Adoption of a Work FOR by Experimental Condition and Personality Trait (Including Participants Who Failed the Predefined Markers for Nonblinding)**



Note: GIs- (generic instructions, full item-level work contextualization absent), GIs+ (generic instructions, full item-level work contextualization present), PSS- (personnel-selection scenario, full item-level work contextualization absent), PSS+ (personnel-selection scenario, full item-level work contextualization present).  $N = 358$ .

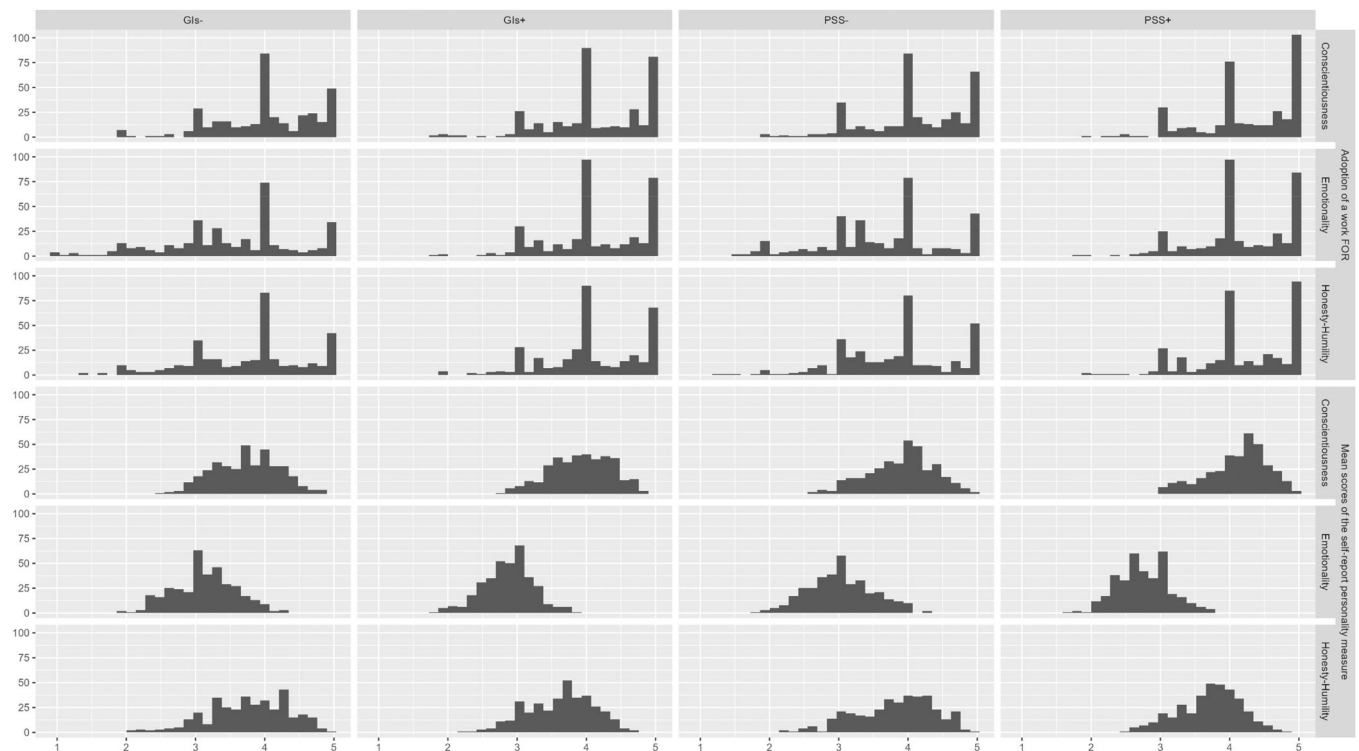
**A18: Box Plots on the Mean Scores of the Self-Report Personality Measure by Experimental Condition and Personality Trait (Including Participants Who Failed the Predefined Markers for Nonblinding)**



Note: GIs- (generic instructions, full item-level work contextualization absent), GIs+ (generic instructions, full item-level work contextualization present), PSS- (personnel-selection scenario, full item-level work contextualization absent), PSS+ (personnel-selection scenario, full item-level work contextualization present).  $N = 358$ .

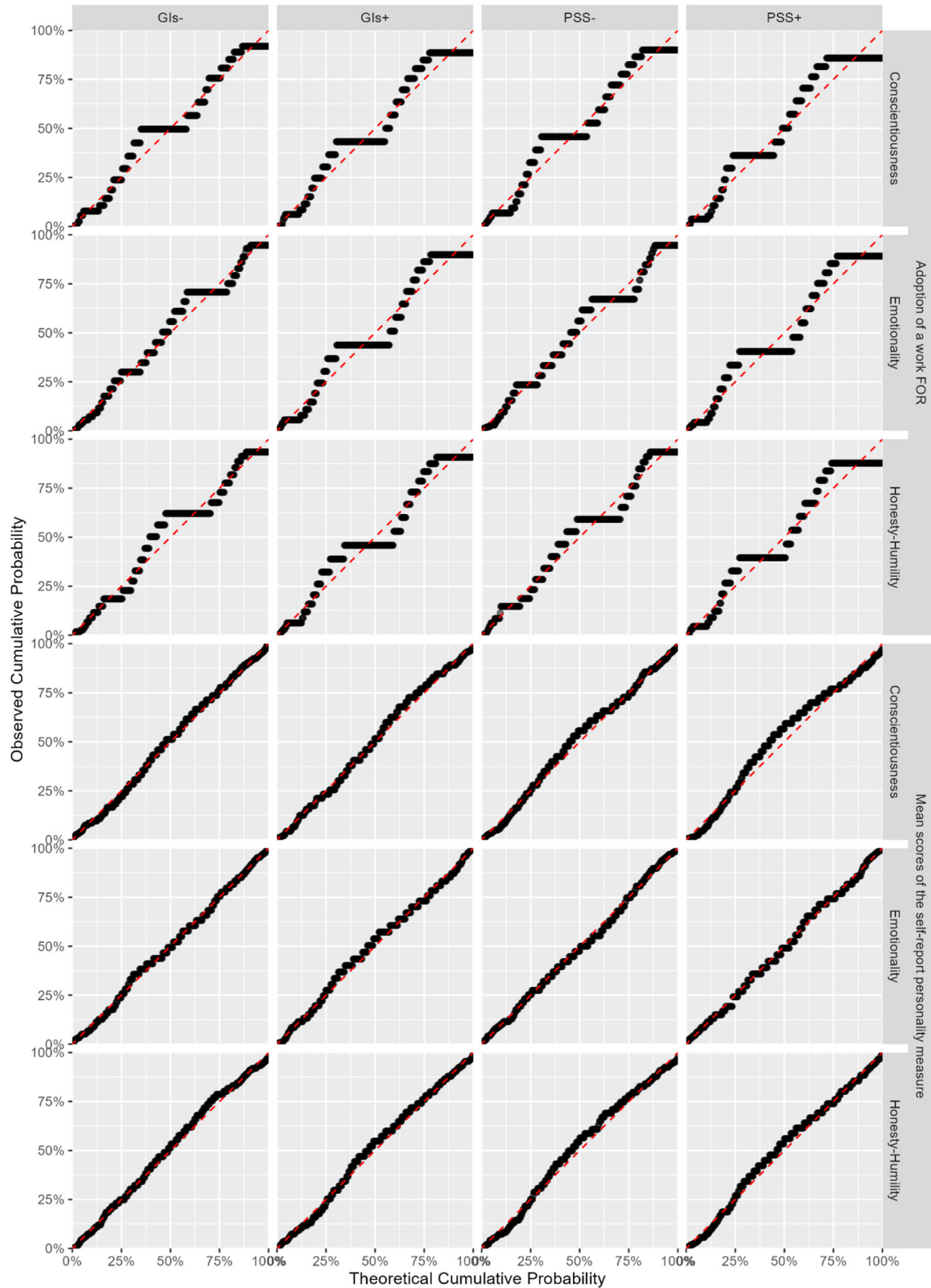


**A19: Histograms for the Adoption of a Work FOR and the Mean Scores of the Self-Report Personality Measure by Experimental Condition and Personality Trait (Including Participants Who Failed the Predefined Markers for Nonblinding)**



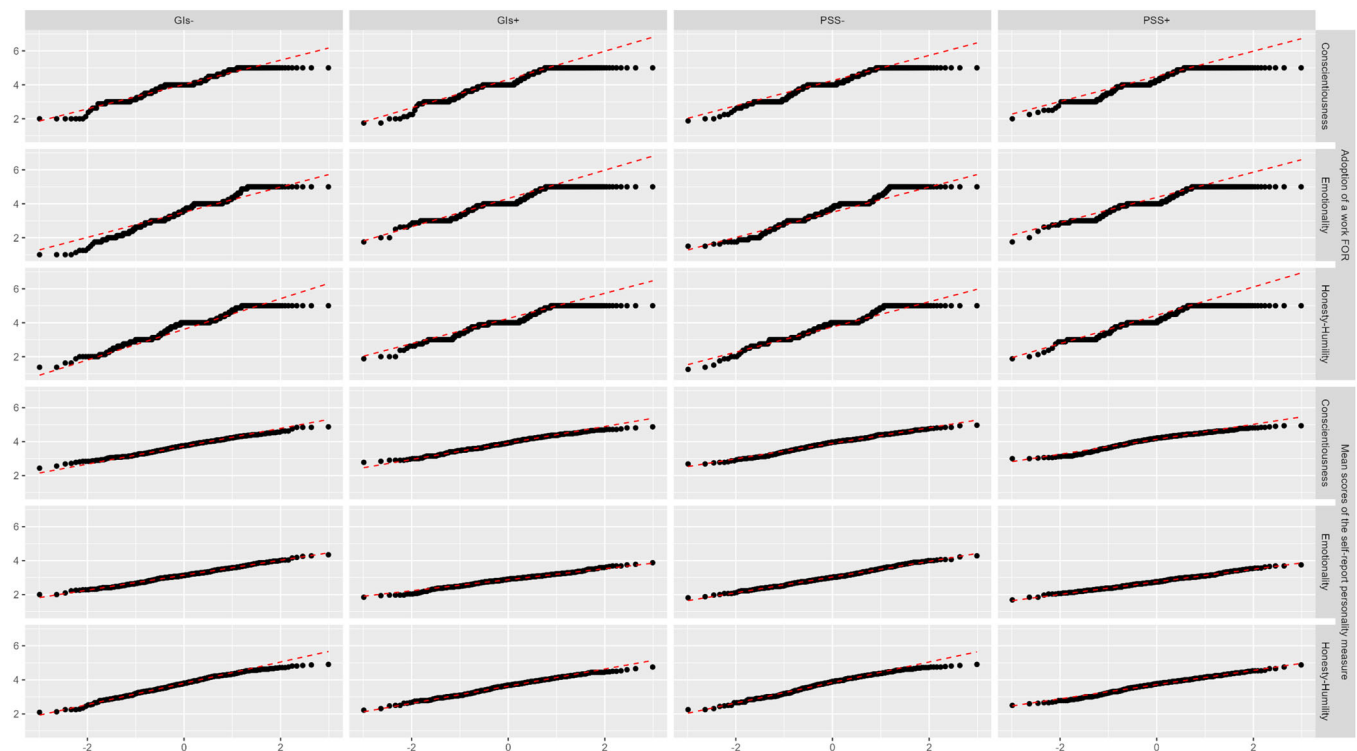
*Note:* GIs– (generic instructions, full item-level work contextualization absent), GIs+ (generic instructions, full item-level work contextualization present), PSS– (personnel-selection scenario, full item-level work contextualization absent), PSS+ (personnel-selection scenario, full item-level work contextualization present).  $N = 358$ .

**A20: PP-Plots for the Adoption of a Work FOR and the Mean Scores of the Self-Report Personality Measure by Experimental Condition and Personality Trait (Including Participants Who Failed the Predefined Markers for Nonblinding)**



*Note:* GIs- (generic instructions, full item-level work contextualization absent), GIs+ (generic instructions, full item-level work contextualization present), PSS- (personnel-selection scenario, full item-level work contextualization absent), PSS+ (personnel-selection scenario, full item-level work contextualization present).  $N = 358$ .

**A21: QQ-Plots for the Adoption of a Work FOR and the Mean Scores of the Self-Report Personality Measure by Experimental Condition and Personality Trait (Including Participants Who Failed the Predefined Markers for Nonblinding)**



Note: GIs– (generic instructions, full item-level work contextualization absent), GIs+ (generic instructions, full item-level work contextualization present), PSS– (personnel-selection scenario, full item-level work contextualization absent), PSS+ (personnel-selection scenario, full item-level work contextualization present).  $N = 358$ .

**A22: Results of Levene's Tests in the Two-Way Repeated-Measures ANOVAs for the Adoption of a Work FOR and the Mean Scores of the Self-Report Personality Measure for Each Personality Trait (Including Participants Who Failed the Predefined Markers for Nonblinding)**

Adoption of a work FOR				
Scale	df1	df2	F	p
Conscientiousness	3	1428	0.63	0.595
Emotionality	3	1428	18.58	< 0.001
Honesty-Humility	3	1428	3.81	0.009
Mean scores of the self-report personality measure				
Scale	df2	F	p	p
Conscientiousness	3	1428	1.38	0.247
Emotionality	3	1428	10.46	< 0.001
Honesty-Humility	3	1428	10.38	< 0.001

Note:  $N = 358$ .

**A23: Results of the Two-Way Repeated-Measures ANOVAs on the Adoption of a Work FOR and the Mean Scores of the Self-Report Personality Measure for Each Personality Trait (Including Participants Who Failed the Predefined Markers for Nonblinding)**

Scale	Effect	<i>F</i> (1, 357)	<i>p</i>	Partial $\eta^2$	95% CI (Bootstrapped)	
					LL	UL
Adoption of a work FOR						
Conscientiousness	I	14.54	0.001	0.04	0.01	0.09
	C	28.34	< 0.001	0.07	0.03	0.13
	I $\times$ C	1.31	0.254	0.00	0.00	0.03
Emotionality	I	10.63	0.001	0.03	0.01	0.07
	C	213.84	< 0.001	0.37	0.32	0.43
	I $\times$ C	1.79	0.181	0.01	0.00	0.03
Honesty-Humility	I	14.39	< 0.001	0.04	0.01	0.09
	C	111.54	< 0.001	0.24	0.18	0.30
	I $\times$ C	0.52	0.470	0.00	0.00	0.02
Mean scores of the self-report personality measure						
Conscientiousness	I	153.03	< 0.001	0.30	0.23	0.36
	C	134.22	< 0.001	0.27	0.19	0.34
	I $\times$ C	1.36	0.245	0.00	0.00	0.03
Emotionality	I	63.61	< 0.001	0.15	0.10	0.21
	C	260.00	< 0.001	0.42	0.36	0.48
	I $\times$ C	0.26	0.610	0.00	0.00	0.01
Honesty-Humility	I	28.93	< 0.001	0.07	0.03	0.13
	C	53.72	< 0.001	0.13	0.07	0.20
	I $\times$ C	1.75	0.186	0.00	0.00	0.03

Note: N = 358. I = main effect of instructions (GIs vs. PSS). C = main effect of full item-level work contextualization (absent vs. present). I  $\times$  C = interaction effect. CI = confidence interval; LL = lower limit; UL = upper limit.