Examining the Effectiveness of the Identify and Removal Faking Correction Strategy that Developed on the Response Pattern Based Faking Detection Methods in Selection Settings

Weiwen Nie

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Psychology

Roseanne J. Foti

Jorge I. Hernendez

Neil M. Hauenstein

Louis Tay

July 20, 2021

Blacksburg, VA

Keywords: Faking, Personality, Machine Learning, Correction

Examining the Effectiveness of the Identify and Removal Faking Correction Strategy that Developed on the Response Pattern Based Faking Detection Methods in Selection Settings

Weiwen Nie

ABSTRACT

In selection research, many studies have attempted to correct for the impact of faking on personality assessment. One of the often-used correction strategies is to identify and remove faking job applicants from the job applicant pool from consideration for employment. In the current study, a simulation study was conducted to demonstrate the identify and removal strategies developed upon two response pattern based faking detection (machine learning method and covariance index method) methods could reduce the propotion of fakers being selected but could not improve the group mean criterion score in the selected group that vary across faker to honest respondent ratios (FHRs) and selection ratios (SRs).

The simulation conditions that used the identify and removal strategies developed based on either the machine learning method or the covariance index method hire a significantly lower the proportion of fakers than the no correction strategy condition across all FHRs and SRs.

The simulation condition that used the identify and removal correction strategy developed based on the machine learning method had a higher correlation coefficient

between emotional stability and the criterion than the no correction condition as long as the FHRs were equal or bigger to 0.15. The simulation condition that used the identify and removal correction strategy developed based on the covariance index method had a higher correlation coefficient between emotional stability and the criterion than the no correction condition as long as the FHRs were equal or larger than to 0.3. On the other hand, the simulation condition that used the identify and removal correction strategy developed based on the machine learning method had a higher correlation coefficient between conscientiousness and the criterion than the no correction condition only when the FHR was equal to 0.5. The simulation condition that used the identify and removal correction strategy developed based on the covariance index method did not have a higher correlation coefficient between conscientiousness and the criterion than the no correction condition across FHRs.

The simulation conditions that used the identify and removal correction strategies developed based on either the machine learning method or the covariance index method did not improve group mean criterion score in the final selected group than the no correction condition across all FHRs and SRs.

These findings suggest that if the main objective is to reduce the proportion of fakers being selected, the use of the identify and removal strategies developed based on the two response pattern based faking detection methods is effective. On the other hand, if the main objective is to improve the group mean criterion score in the selected group by maximizing the criterion related validity, the use of this correction strategy will produce a negative effect.

Examining the Effectiveness of the Identify and Removal Faking Correction Strategy that Developed on the Response Pattern Based Faking Detection Methods in Selection Settings

Weiwen Nie

GENERAL AUDIENCE ABSTRACT

In selection research, many studies have attempted to correct for the impact of faking on personality assessment. One of the often-used correction strategies is to identify and remove faking job applicants from the job applicant pool from consideration for employment. In the current study, a simulation study was conducted to demonstrate the identify and removal strategies developed upon two response pattern based faking detection (machine learning method and covariance index method) methods could reduce the proportion of fakers being selected but could not improve the group mean criterion score in the selected group that vary across faker to honest respondent ratios (FHRs) and selection ratios (SRs).

Acknowledgements

I would first like to thank my dissertation chairs Dr. Ivan Hernandez and Dr. Roseanne Foti. Dr. Hernandez patiently works with me to organize my dissertation ideas and coding related matters. Dr. Foti patiently help me improve my academic writing skill. This dissertation paper would never be completed without both Dr. Hernandez and Dr. Foti's guidance. I would also like to express my appreciation to my committee members, Dr. Louis Tay and Dr. Neil Hauenstein for serving as my committee members. I want to thank you for your brilliant comments and suggestions on the design of this study.

TABLE OF CONTENTS

List of Tables

List of Figures

1. Introduction

Meta-analyses (Barrick & Mount, 1991; Schmidt & Hunter, 1998) demonstrate that the criterion-related validity of some personality traits and job performance generalize across different jobs. Organizations increasingly use personality measurements in personnel selection contexts (Hough & Oswald, 2008; Rothstein & Goffin, 2006). However, as this practice increases, the concern of response distortion, also called “faking,” in personality tests increases as well (Burns & Christiansen, 2006; Hogan, Hogan & Murtha, 1992). Evidence suggests that individuals are capable of faking when instructed in a lab setting (Viswesvaran & Ones, 1999; Ziegler, Schmidt-Atzert, Buehner, & Krumm, 2007) and nearly one third (about 28% -31%) of job applicants do fake in actual selection context (Peterson, Griffith, O’Connell, & Isaacson, 2008; Donovan, Dwight, & Schneider, 2008; Griffith, Chmielowski, & Yoshita, 2007).

Research investigating the impact of faking on personality assessment in selection focuses on two perspectives, the proportion of fakers being hired and the group mean criterion score in the selected group. Research that concerns the impact of faking on the proportion of fakers being hired finds that faking has a strong impact on the rank order of the job applicant sample (e.g., Rosse, Stecher, Miller, & Levin, 1998). As a result, fakers tend to rise to the top of the selection ranking and are likely to be hired compared with honest job applicants who did not fake their personality assessments. Research that concerns the impact of faking on the effectivness of the selection suggested faking has a significant impact on the group mean criterion score in the selected group (e.g., Donovan, Stephen & Dwight, 2013; Mueller-Hanson, Heggestad, & Thornton, 2003). As a result, the organization will hire less qualified employees.

To correct for faking in personality measurement in the selection context, one often-used strategy is the identify and removal correction. The identify and removal strategy first uses a faking detection method to identify potential fakers in the job applicant group and then removes them from progressing further in the selection procedure. Hough (1998) found when the identify and removal strategy was implemented, the mean score difference between the job applicant group and the incumbent group was reduced and led to a significant change in the final selected job applicant group. Still, it did not improve the criterion-related validity of the job applicant group (e.g., Hough, 1998).

When implementing the identify and removal correction strategy, researchers often use the social desirability responding (SDR) scale as the faking detection method (e.g., Hough, 1998). The SDR scale was developed to identify individuals who are intentionally faking in self-report surveys (e.g., Wiggins Social Desirability Scale, Wiggins, 1959). Individuals who score high on the SDR scale are intentionally faking, and they are removed from the job applicant pool. However, research suggests that the SDR scale is a blunt instrument. In a meta-analysis, Viswesvaran and Ones (1999) found that the difference between the SDR scores of the participants who were under the fake good instruction and honest response instruction was approximately only one standard deviation. In other words, the Pearson's r correlation between faking and the SDR scores was approximately r=.45, which means that only 20% variance of the SDR scores was explained by faking (Christiansen, Robie, Burns &, Speer, 2017). The SDR scores' explained variance to faking means using the SDR scores to identify fakers can only achieve low classification accuracy to classify fakers and honest respondents in a job applicant sample. This low predictive ability of commonly used identification methods cast serious doubt on the effectiveness of the identify and removal strategy because this strategy's effectiveness heavily depends on the classification accuracy of the faking detection method.

To address the low predictability issue of the SDR scale in faking detection tasks, researchers developed multiple response-patten-based faking detection methods. The response-pattern-based faking detection methods do not rely on external scales to identify potential fakers. Instead, they rely on the internal item response pattern, which provides a wealth of data for large personality inventories. Three response-pattern-based faking detection methods—idiosyncratic item method (Kuncel & Borenman, 2007), covariance index method (Christiansen et al., 2017), and machine learning model method (Calanna, Lauriola, Saggino, Tommasi & Furlan, 2019)—have found that these faking detection methods can outperform the faking detection method based on SDR scales in terms of classification accuracy. Recently, Nie, Hernandez, Zhang, Yang, and Tay (under review) found that the machine learning model method outperformed the idiosyncratic item method and covariance index method across all conditions, regardless of the different ratios between fake and honest respondents on all common classification quality metrics (i.e., area under the curve, accuracy, precision, and recall).

For the current study, I plan to further test the effectiveness of the identify and removal correction strategy based on two different response-pattern-based faking detection methods with a pseudo simulation study. I structure my dissertation with the following sections: first, I present the potential impact of faking on personality measurement in selection context; second, I detail the advantages of the identify and removal correction strategy to correct for faking in personnel selection context; third, I review the studies that suggest machine learning methods have the potential to provide the best classification accuracy in a real-world job applicant sample; fourth, I lay out my research questions on how to assess the effectiveness of the identify and removal correction strategy that is based on the machine learning faking detection method; finally, I propose the research method to investigate these research questions.

2. Literature review

2.1 The Definition of Faking in Selection Context

Faking in personality assessment is typically considered a form of response bias. Response bias refers to any systematic variance in the test score that is unrelated to the true score of the attribute of interest (Paulhus, 1991). Jackson and Messick (1958) argued there are two different types of response bias, response style and response set. Response style refers to the biases that are consistent across time and questionnaires. Response set refers to the short-lived response biases attributable to some temporary distraction or motivation.

Researchers who study faking have been tried to distinguish response style and response set with different terms. Sackeim and Gur (1978) used the terms “self-deception” and “other-deception.” Damarin and Messick (1965) used the terms “self-regard” and “propagandistic bias.” Kusyszyn and Jackson (1968) used the terms “desirability” and “defensiveness.” Paulhus (1984) used the terms “self-deception” and “impression management.” Regardless of how these researchers name these two dimensions of response bias when respondents believe their self-report result regardless of whether the self-report result reflects the ultimate truth of the respondent or not, this response distortion should be considered a response style because these responses are stable across situation and time. On the other hand, when respondents consciously distort their response because of situational demand, this should be considered as a response set because this response distortion will change across situations.

When considering faking in the selection, we should consider faking as a response set in which individuals intentionally distort their response in the personality assessment because of situational demands. Faking in the selection context is planned behavior that is activated by the high-stake situational demand of getting hired (Ziegler, Maccann & Robert, 2012). Ajzen (1991) argued that behavior intention is the best way to predict the planned behavior. He suggested behavior intention is determined by three factors, the attitude of the behavior (the self-evaluation of the behavior), the subjective norm of the behavior (the evaluation of whether others will perform the behavior), and perceived behavioral control (self-evaluation of the difficulty to perform the behavior). Based on Ajzen's theory of planned behavior, McFarland and Ryan (2006) proposed three situational-based cognitive variables, attitude toward faking (the degree to an individual perceive faking as favorable or unfavorable), subject norm of faking (whether an individual thinks others fake), and perceived behavior control of faking (perception of whether one can successfully fake) that should impact the intention to fake in a selection context. They found supports for their model in two studies. In study one, they instructed participants to fill out a survey to measure these three variables and their intention of faking personality assessment in the future on selection test. They found attitude toward faking ($b = .55$, $p \leq .01$), perceived behavior control of faking ($b = .17$, $p \leq .01$), and subjective norm ($b = .13$, $p \leq .01$) of faking had significant coefficients predicting the intention to fake. In study two, they randomly assign participants to the incentive condition (participants were told that those who scored in the top 15% would receive $20) and the non-incentive condition (no incentive was offered). They found individuals with a strong intention to fake in the incentive condition significantly distorted their personality trait scores. Using a similar theoretical framework, Mueller-Hanson, Heggestad, and Thornton (2006) proposed both the situational-based cognitive variables (e.g., belief in the importance of faking, perceived behavior control, subjective norms, and the trait-like variable willingness to fake variables (e.g., Machiavellianism, lack of rule-consciousness, and self-monitoring) should have positive correlations with the intention to fake in a selection context. They found that only perception of the situation but not willingness to fake had positive correlations with the intention to fake. In summary, these findings suggest whether an individual decides to engage in faking in a selection context is largely determined by the cognitive perception of the situational demand.

While situational demands influence whether a person decides to engage in faking, the degree to which an individual can distort the personality trait score in a desirable direction is determined by their abilities to fake. To achieve the intended goal in the selection context, individuals who engage in faking need to create cognitive schemas consistent with the situational role demand (Jansen, KöNIG, Kleinmann, & Melchers, 2012). Social psychology studies suggested individuals create impressions that are consistent with the role demands of different situations (e.g., Leary & Kowalski, 1990; Leary, Robertson, Barnes, & Miller, 1986). In a job application context, job applicants expect they will be evaluated based on the job role requirements. Therefore, the job applicants who decide to engage in faking (fakers) create cognitive schemas that they think best exemplify the job role that they are applying for. This process was named "role-playing" in the literature that studied response distortion (Cofer, Chance, & Judson, 1949; Wiggins, 1959). After fakers create the cognitive schema for the position they apply for, they try to decipher the underlying behavior constructs measured by the personality items and match the cognitive schemas they created specifically for the job to the behavior constructs (Holden, Kroner, Fekken &, Popham,1992; Fekken & Holden, 1992). Therefore, fakers respond to the items in the personality assessments based on the cognitive schemas they created specifically for the job instead of the own self-schema.

To successfully create a cognitive schema that matches the situational demand (e.g., the job posting) to fake the personality assessment in a selection context, survey respondents need to have the necessary abilities to fake (Ellingson & McFarland, 2011; Snell, Sydell, & Lueke, 1999). The necessary abilities refer to dispositional factors (e.g., cognitive ability and emotional intelligence) and experiential factors (e.g., the knowledge of what is required for the job and the knowledge of what the test is being measured) (Snell et al., 1999). This is because having these necessary abilities to fake will help the individuals who intend to fake to create a cognitive schema that fits better to the job (Pelt, van der Linden, & Born, 2018; Raymark & Tafero, 2009; Roulin, Krings, & Binggeli, 2016) and interpret the underlying behavior constructs that being measures by items more accurately (Cunningham, Wong, & Barbee, 1994; Dwight & Alliger, 1997; Holden & Jackson, 1981; Napier, 1979). Therefore, individuals who have a high level of the necessary abilities that are required for faking can distort their scores in more favorable directions and to a higher degree.

In summary, faking in the job application context should be considered as a response set that differs from one's initial response. Therefore, the chance of an individual decides to engage in faking is determined by the individual's cognitive perception of the situational demand. However, the degree of an individual distorts his/her response in a personality assessment is determined by the faking abilities of this individual.

2.2 The Impact of Faking on Personality Measurement in The Personnel Selection Context

Research suggests most individuals are capable of distorting their response when they are instructed to fake in lab settings (Smith, Hanges, & Dickson, 2001), and a significant proportion of job applicants are willing to fake in job application settings (Birkeland, Manson, Kisamore, Brannick, & Smith, 2006). In a meta-analysis, Viswesvaran and Ones (1999) found, when individuals were instructed to fake in laboratory settings, their personality trait scores increased in both between-subject designed studies and within-subject designed studies. In between-subject designed studies, the subjects increased more than half a standard deviation (Cohen's $d \approx .6$) for all big five personality traits. In within-subject designed studies, the subjects increased more than a two-third quarter of standard deviation (Cohen's $d \approx .8$) for emotional stability, openness to new experiences, and conscientiousness. Additionally, subjects increased roughly half a standard deviation (Cohen's $d \approx .5$) for extraversion and agreeableness. In another meta-analysis, Birkeland et al. (2006) found that test-takers inflated their personality trait scores in real-world applications as well. They found test-takers inflated the trait scores of conscientiousness (Cohen's $d = .45$) and emotional stability (Cohen's $d = .44$) to a greater degree and the trait scores of extraversion (Cohen's $d = .11$) and openness to new experience (Cohen's $d = .13$) to a lesser degree.

Given that faking distorts overall personality scores, the impact of faking on personality measurement can be described in terms of the proportion of faking being hired and the group mean criterion score in the selected group. Faking has a strong impact on the rank order of the job applicant samples (Holden, 2008). When a top-down selection method is implemented (e.g., strict top-down referral in order of test score), the rank order of the job applicants in terms of test scores directly determines whether they are selected (Cascio, Outtz, Zedeck & Goldstein, 1991; Sackett & Roth, 1991). Therefore, job applicants who fake to inflate their trait scores are more likely to rise to the top of the rank and be hired. Zickar, Rosse, Levin, and Hulin (1997) found that including 5% of fake stimulated-job -applicants that only slightly inflated their test scores, 15% of these fake stimulated-job- applicants would end up in the upper end of the ranking order. When including 25% fake stimulated-job-applicants that moderately inflated their test scores, 63% of these fake stimulated-job-applicants would end up in the upper end of the ranking order. In a field study, Rosse et al. (1998) supported Zickar et al.'s (1996) finding with an actual job applicant sample. They found that as the selection ratio (SR) decreases, the extent of faking (measured by the response distortion score in the Balanced Inventory of Desirable Responding scale (Paulhus, 1991) increases for the selected job applicant group. When the SR is equal to or smaller than 60%, job applicants who have response distortion scores of 3 standard deviations higher than the incumbent's response distortion score will always be selected. These findings are concerning because they suggest when using personality assessment as a selection tool, the main factor that determines whether job applicants are hired was not the latent trait score that being accessed by the personality assessment tool. Rather it is the willingness and capability of faking. Given that only a proportion of job applicants are willing to engage in faking (Griffith, Chmielowski, & Yoshita, 2007), these fakers will have an unfair advantage at being hired. It is ethically problematic for employers to punish job applicants who are honest and reward the job applicants who are willing to lie.

However, the impact that faking has on the validity of the personality assessment is not as direct as its impact on the proportion of fakers being selected. Research findings on the impact of faking on the criterion-related validity of personality data are mixed. While findings from lab research suggest that when participants were instructed to fake, the criterion-related validity of the personality measurement was significantly damaged (Caldwell-Andrews et al., 2000; Holden, 2008; Topping & O’Gorman, 1997), research findings in real-world assessment suggest that faking had minimal to no impact on criterion-related validity (Hough, Eaton, Dunnette, Kamp, and McCloy, 1990; Barrick & Mount, 1997). However, there is one weakness in the design of the studies that suggest faking has minimal to no impact on criterion-related validity. These studies assumed faking occurred in the applicant sample without knowing the proportion of subjects who faked and the degree of faking in the sample. In a simulation study, Zickar, Rosse, Miller, and Levin (1996) illustrated both (1) the proportion of subjects who fake and (2) the degree of faking reduce criterion-related validity. They found that manipulating the proportion of individuals who faked and the extent of score inflation could reduce the criterion-related validity by up to 20%. Specifically, this occurred when the simulated faking subjects inflated their scores to an extreme degree (by adding 1 SD to the score) and 50% of fakers were included in the sample. For the rest of the simulation conditions, faking had a weaker impact on criterion-related validity.

However, two studies that directly investigated the impact of faking on the group mean criterion score in the selected group suggested faking is detrimental to the performance of the selected group. In a lab experiment, Mueller-Hanson et al. (2003) randomly assigned 444 participants into two conditions—the "honest response" condition and the "incentivized to fake" condition—to fill out the achievement motivation questionnaire. In the "honest response" condition, participants were instructed to fill out the questionnaire honestly. In the "incentivized to fake" condition, participants were told that people who were hard-working, motivated, and conscientious would be selected to participate in a second part of the study, which would have a chance to win $20. At the end of the study, the participants from both conditions were invited to participate in the second stage of the study to perform 50 boring tasks that consisted primarily of matching tasks, finding discrepancies, and simple problem-solving tasks. The performance on these 50 boring tasks served as a criterion of the study. After combining the achievement motivation questionnaire from both the honest and the incentive conditions, they found that as the SR decreased, more participants from the incentive condition were selected than the participants from the honest condition. More importantly, the average performance of the participants selected from the honest condition always exceeded the participants selected from the incentive groups across different SRs. In a field study, Donovan, Dwight, and Schneider

(2013) investigated the impact of faking on both the training performance difference and the job performance difference (operationalized by calculating the change in market share percentage) of 162 sales representatives in a pharmaceutical company. Participants in this study were required to complete the dispositional goal orientation survey (VandeWalle, 1997) two times, the first time when the participants were job applicants and completed the survey for hiring purposes, and the second time when the participants were hired for developmental purpose. Donovan et al. (2013) used the standard error of measurement to calculate the 95% confidence interval (CI) for the trait score with the posted-hired sample. Participants whose trait scores in the pre-hired job applicant sample were outside the 95% CI on the higher end are identified as fakers. They then conducted a similar simulation as the Mueller-Hanson et al. (2003) study. They found a similar result; as the SR decreased, more and more identified fakers were hired. More importantly, when the SR was 20%, the training performance of the participants who were classified as fakers was significantly lower than the participants who were classified as honest respondents identified as honest respondent(t (46) = 2.48, p<05, d = .76). Even though the difference in terms of actual job performance was not statistically significant at traditional levels (t (46) = 1.34, p=10, one-tailed), the mean standardized difference was nonetheless moderate in magnitude (d = .39). These two studies suggest faking lowers the overall quality of the selection outcome when applicants are selected from a truncated top-tier. In turn, faking damages the group mean criterion score in the selected group.

2.3 Identify and Removal as a Correction Strategy for Faking

Multiple strategies were developed to address the negative effect faking has on the selection process. These strategies include the suppression strategy, the moderation strategy, and the identify and removal strategy (Reeder, & Ryan, 2012). Each strategy uses a different approach to use the identification of "faker" to correct the criterion-related validity between the applicants' scores and their outcomes.

With the suppression strategy, researchers assume that faking only introduces construct irrelevant variance to the personality trait but that the tendency to fake is uncorrelated with the criterion (Ones, Viswesvaran & Reis, 1996). Therefore, including a suppressor term (the measurement of faking) in the regression equation regressing the criterion on the personality trait can correct the criterion-related validity mitigated by faking. However, there is a significant weakness in the suppression strategy: it needs to rely on the SDR scales to measure the level of faking. The meta-analysis found SDR scores correlated with three of the Big Five personality traits, emotional stability ($\rho = .37$), conscientiousness ($\rho = .20$), agreeableness ($\rho = .14$) a small to moderate degree (Ones et al., 1996). Based on this finding, individuals who are highly emotionally stable, conscientious, and agreeable will be more likely to receive a lower post-correction score even if they did not fake in the personality survey when the suppression correction strategy is applied. This is because the sign of the SDR score term is always negative in the regression equation when a suppression strategy is applied. Therefore, honest respondents who have higher trait scores on emotional stability, conscientiousness, and agreeableness will have lower post-correction scores of these personality traits. This is the reason why prior studies using the suppression strategy to correct for faking fail to reduce the proportion of faker being hired (Ellingson, Sackett & Hough, 1999) and the criterion-related validity (Hough, 1998; Hough, Eaton, Dunnette, Kamp, & McCloy, 1990)

With the moderation strategy, researchers assume that faking moderates the relationship between the personality trait and the criterion in which the relationship between the personality trait scores and the criterion under faking condition is lower than the between the personality trait scores and the criterion under honest condition (Ganster, Harry, & Hennessey, 1983). Based on this assumption, the actual criterion-related validity is restored by adding an interactional term between the personality trait and the measurement of faking in the regression equation that predicts the criterion. Even though no study has investigated whether the moderation strategy can directly reduce the proportion of fakers being hired, Holden (2007) found evidence suggesting that the moderation strategy can reduce some degree of score distortion. In a lab study, Holden (2007) had all participants honestly respond to the first personality assessment (NEO-FFI, Holden, Wood, & Tomashewski (2001), and then randomly assigned them to two groups (the group that response honestly and the group that instructed to fake good) to complete the second personality assessment (NEO-FFI; Costa & McCrae, 1991) and the Impression Management scale of the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1998). They found that the R-squared coefficient increases significantly when including the actual group assignment as a moderator in the regression models that regressing the personality trait scores of the first assessment onto the same personality trait scores of the second assessment compared with the regression models that did not include the moderator. However, the impression management score as a moderator did not have any effect on the regression model. On the other hand, Holden's (2008) findings further support that using the moderation strategy can improve criterion-related validity. In a lab study, Holden (2008) randomly assigned participants into the group with standard honest response instruction and the group that instructed individuals to fake good. Participants in both groups answer the Holden Applicant Reliability Measure (HARM; Holden, 2000) and five other SDR scales according to their assigned group instructions and answering the Levenson Self-Report Psychopathy Scale (LSRP; Levenson, Kiehl, & Fitzpatrick, 1995) honestly regardless of the group assignment. Holden found when regressing the HARM onto the LSRP, including the group assignment as a moderator results in a significantly higher R-squared coefficient compared with the regression model with no moderator. However, he also found, including any of the five SDR scores assessed by the SDR scales as the moderator did not result in a significantly higher R square compare with the regression model with no moderator. The fact that only the actual group assignment but not the SDR score could effectively correct for faking with the moderation strategy was not surprising. Multiple lab studies provided evidence suggesting when participants were instructed to fake, the criterion-related validity of the personality measurement was significantly lower than when participants respond honestly (Caldwell-Andrews, et al., 2000; Holden, 2007; Topping & O'Gorman, 1997). Given that the correlation between the predictor and the criterion is significantly lower in the faking group compared to the honest group, including the actual group assignment as a moderator should result in a significantly higher R-squared coefficient in the regression model. On the other hand, as the prior paragraph mentioned, the SDR score has a significant moderate correlation with personality measurement. As a result, the signs of both the SDR score term and the multiplicative term of the SDR score and the personality trait score are always in the opposite direction of the predictor in the regression equation when a moderation strategy is applied. Therefore, honest respondents who have a higher score on the predictor will result in lower post-correction scores of these personality traits because of the correlations between these trait scores and the SDR score as well. This is the reason why including SDR score and as a moderator in the regression model does not increase the R-squared coefficient compared with no moderator included.

The identify and removal strategy is an extension of the moderation strategy. Given the correlation between the predictor and the criterion for the faking group is significantly weaker than the honest group, by identifying and removing the suspected fakers in the job applicant sample, fewer fakers will be included in the job applicants sample and the correlation between the predictors and the criterion should improve. However, for the identify and removal strategy to be effective, another assumption besides the assumption of the moderation strategy must be met: a proportion of job applicants must be honest. If all job applicants intentionally fake in the selection context, all job applicants need to be removed from the job applicant sample. Findings from multiple studies provided evidence to support this assumption. In a study with an actual job applicant sample, Griffith et al.(2007) collected data on a conscientiousness measurement from the temporary workers at the application stage and then had the same subjects completed the same items one month later for a research purpose. They detected faking by forming a confidence interval by multiplying the standard deviation of the measurement by 1.96 with the data collected for a research purpose. This confidence interval implied if the subjects responded to the conscientiousness measurement honestly in the job application stage, 95% of the time, they would fall within this confidence interval. With this detection method, they found 31% of job applicants were categorized as fakers. Peterson, Griffith, O'Connell, and Isaacson (2008) replicated this study with a manufacturing position sample and found similar results. Donovan et al., (2008) conducted a similar study with a sample of job applicants for a pharmaceutical sales position. They collected personality data for the job applicants in the application stage and then administrated the same personality measurement to the subjects one month after they were hired for a development purpose. Used the same method in Griffith et al.'s (2007) study, they found 27 % of applicants were classified as fakers. Even though empirical findings support the assumption of the identify and removal correction strategy, past research did not have success using the identify and removal correction strategy to improve the criterion-related validity in both simulation study and field study.

Schmitt and Oswald (2006) conducted a simulation study to investigate the impact of the identify and removal strategy on the group mean criterion score of the selected group. They simulated job applicant samples by varying six conditions, sample size, SR, removal ratio(proportion removed for suspected faking), the correlation between the faking measurement score and the trait score, the correlation between the faking measurement score and the criterion score, and the correlation between the trait score and the criterion score. They estimated the impact of the identify and removal strategy with four steps: first, creating selected groups with the SRs in the simulated job applicant samples; second, estimating the group mean criterion score in the selected group; third, removing the simulated individuals based on the faking measurement score by removal ratios in the simulated job applicants samples and recreate the selected groups with the post-correction simulated job applicant samples; fourth, re-estimating the group mean criterion score of the newly form selected groups. They found that the identify and removal strategy only have a negligible impact on the group mean criterion score

However, there was a fundamental problem with the simulation study conducted by Schmitt and Oswald (2006). Schmitt and Oswald simulated faking with a continuous faking measurement score. This means they assumed all job applicants who take the personality assessment intentionally fake the assessment but just to a different degree. However, as the previous paragraph discussed, one of the assumptions of identify and removal strategy is there must be honest respondents in the job applicant samples. Therefore, the simulation method Schmitt and Oswald (2006) used in the study did not match with one of the assumptions of the identify and removal strategy.

In a field study, Hough (1998) administrated the unlikely virtue scale (a form of SDR scales) to both the incumbents and the job applicants. Job applicants who scored higher or equal to the top 5 percentile of the incumbents' unlikely virtual scores were identified as potential fakers. Hough found that implementing the identify and removal strategy resulted in more similar mean scores between the mean trait scores of interest of the incumbent group and the job applicant group, which led to a very different set of job applicants would be selected if this strategy was used. However, Hough also found that adapting the identify and removal strategy had no impact on criterion-related validity. The major reason for the failure to improve criterion-related validity in Hough's study might be due to the method used to identify potential fakers had low accuracy. In Hough's study, she used the unlikely virtue scale (a form of SDR scale) to identify potential fakers. Job applicants were removed from the sample if their unlikely virtue scores were higher than 5 percentiles of the incumbent sample. In a lab study, Ellingson et al. (1999) found the Cohen's *d* of the unlikely virtue score between the faking and honest respondents was only .86 with the mean score of honest respondents = 14.4 (*SD* = 2.8) and mean score of the faking respondents =18 (*SD* = 5.2). In other words, a significant proportion of the unlikely virtue score range between the honest and faking condition overlapped. As a result, it significantly lowers the faking classification accuracy of the unlikely virtue scale.

To use the identify and removal correction strategy to effectively correct for faking, the method that is used to identify potential fakers must have high accuracy. The faking detection method developed using a machine learning model should help achieve this goal. In the following sections, I will describe faking detection methods developed based on external measurement scales and the faking detection method developed based on the internal item response pattern. Then I will detail why the faking detection approach developed based on a machine learning model (i.e., a random forest) can outperform these current faking detection methods.

2.4 A Review of the Current Faking Detection Methods

To better identify who fakes in personality measurement, personality researchers developed four different types of faking detection methods: response latency method, eye-tracking method, external measurement scale method and, response patterns based method.

Since Dunn, Lushene, and O'Neil (1972) suggested that the amount of time elapsed between the presentation of an item and a respondent's subsequent response could be used to measure response deviance, researchers started to utilize this response latency to detect faking in personality assessment. However, researchers could not agree on how and why the response latency would be in effect in honest respondents and fakers. One perspective proposed that fakers should have shorter response latency compare with honest respondents (Holden et al., 1992). Researchers who supported this perspective argued that fakers only use semantic evaluation to respond to an item while honest respondents need to refer to self-referencing autobiographic information to respond to an item (Dunn et al., 1972). Given fakers do not need to recall their daily episodes that associate with the item to respond to an item but only focus on the semantic meaning of the items, their response time should be shorter than the honest respondent (Hsu, Santelli &, Hsu, 1989). However, the other perspective proposed that fakers should have longer response latency compare with honest respondents (Zukerman, DePaulo &, Rosenthal, 1981). Researchers who supported this perspective argued that fakers require more cognitive processing and editing than honest respondents. While honest respondents only need to respond to the item that is consistent with their self-schemas, fakers need to respond to the item that is inconsistent with their self-schemas (McDaniel & Timm, 1990). Therefore, fakers need to inhibit the information from their self-schema and replace this information from the cognitive schemas that are created by the fakers when responding to an item (Vasilopoulos, Reilly, & Leaman, 2000). Given that the cognitive processes for the fakers are more complex and no immediate ready for recall, fakers should take longer to respond to an item compare with honest respondents (Walczyk, Roper, Seemann &, Humphrey, 2003). Empirical findings also were mixed. While some studies found evidence suggesting response latency is longer for honest respondent (e.g., Brunetti, Schlottmann, Scott, & Hollrah, 1998; Holden et al., 1992), some other studies found evidence suggesting response latency is longer for fakers (e.g. Fine & Pirak, 2016; McDaniel & Timm, 1990; Walczyk, Schwartz, Clifton, Adams, Wei, & Zha, 2005).

The theories behind the eye-tracking faking detection method are similar to the theories behind the response latency faking detection method but operationalized differently. With the eye-tracking method, researchers use the numbers eye fixation as a metric to detect faking (e.g. van Hooft & Born, 2012; Simmering, 2012). Our eyes make rapid movements or remain relatively still during fixations on specific regions of interest when looking at a picture or reading text (Rayner, 1998). We obtain information when our eyes fixate on a specific region (Rayner, 1998). Faking research that studies the eye-tracking method focus on the number of eye fixations (e.g., van Hooft & Born, 2012; Simmering, 2012). Given that semantic evaluation requires a less complex cognitive process than self-referencing(Hsu et al., 1989), if fakers do use semantic evaluation and the honest respondent use self-referencing cognitive process to respond to an item as Dunn et al. (1992) proposed, the number of eye fixation should be large for the honest respondent and smaller for fakers. This is because after the honest respondents interpret the semantic meaning of the items, they need to match the semantic meaning of the items with their semantic meaning with their self-schema to respond to the item, but the fakers only respond based on the positive or negative of the sematic meaning of the items without matching their self-schema (van Hooft & Born, 2012). On the other hand, if the fakers need to inhibit the self-schema information and replace the information with an ideal schema and the honest respondent just quickly retrieve information from their self-schema as Vasilopoulos et al. (2000) proposed, the number of eye fixations should be large for fakers and smaller for the honest respondent because faker requires a more complex cognitive process to respond to an item (van Hooft & Born, 2012). Like the response latency method, findings from empirical research do not agree that the number of eye fixations should be large or smaller when comparing the number of eye fixations between fakers and honest respondents. While van Hooft and Born (2012) found the number of eye fixation was larger for honest respondents than the fakers, Simmering (2012) found the number of eye fixation was smaller for honest respondents than the fakers.

Research on both the response latency and eye-tracking faking detection methods does not offer any consensus on how researchers and practitioners detect faking in research and application settings. On the other hand, faking research on the external scale and the response pattern-based faking detection method does have consensus on how to detect faking, but they are simply different in terms of effect size. In the rest of this section, I will detail these two faking detection methods.

Two types of external scales, the social desirability response (SDR) scale and the overclaiming questionnaire (OCQ), were developed to detect faking. SDR scales were developed using two approaches. In the first approach, researchers collect a large variety of items and have participants rate these items' social desirability. The final version of the SDR scale includes items that have extremely high or low ratings of social desirability (e.g., Edwards Social Desirability Scale, Edwards, 1953). The second approach is to use a between-group design.

Researchers have one group of participants responds to a set of items honestly and another group of participants responds to the same set of items with fake-good instruction. The items that best discriminate against these two groups of participants are selected to form the SDR measurement scale (e.g., Wiggins Social Desirability Scale, Wiggins, 1959). Individuals who score high on these two types of scales are considered to be faking on the personality test. However, as the prior section mentions, the SDR scale score has a moderate to small correlation with three of the Big-Five personality traits. Therefore, it will lead to a high false-positive rate and low accuracy to identify potential fakers. Meta-analysis finds that Cohen's *d* of the SDR score is 1.06 (Viswesvaran & Ones, 1999).

The other type of faking detection method that relies on external measurement attempts to detect faking with an over-claiming technique. Phillips and Clancy (1972) found survey takers claimed familiarity with a set of non-existent consumer-knowledge-related items. Based on this finding, Paulhus, Harms, Bruce, and Lysy (2003) developed an overclaiming questionnaire (OCQ) to measure test-takers' tendency to claim knowledge about the nonexistent topics. In the OCQ, a group of genuine topic items (e.g., Manhattan Project), embedded with a small number of nonexistent topics items (e.g., ultra-lipid), are presented to the test takers. Test takers are instructed to rate their familiarity with each of these topic items. The extent of test-takers' endorsement of the knowledge of the nonexistent items is used as a proxy to indicate faking. This method assumes that overclaiming knowledge on nonexistent items reflects the test-takers' SDR scores in other tests that were administered in the same setting (Burns & Christiansen, 2011).

Findings regarding the effect of OCQ to detect faking are mixed. Bing, Kluemper, Davison, Taylor, and Novicevic (2011) found the OCQ bias score (calculated based on the degree of nondiscriminatory endorsement of all the topic items) was significantly higher when undergraduate students were instructed to pretend to apply for a university admission than the honest response condition (Cohen's *d* = .57). However, Feeney and Goffin (2015) found the positive correlation between the OCQ bias score and the personality score inflation in a simulated job application condition was very weak. They found the OCQ bias score only weakly predicted residualized individual change scores (*r* = .17), which was calculated by regressing participants' item scores, collected in the simulated job application condition, on the item scores collected in the honest condition, and then by averaging the residuals and near-zero relations with admissions of purposefully faking in the simulated job application condition. Dunlop, Bourdage, De Vries, McNeill, Jorritsma, and Choe (2019) suggested that the mixed finding of OCQ might be due to the relevancy of the OCQ in the measurement context. They further argued OCQ needs to be tailored to be relevant to the job content to be effective. In a lab study, they found that the effect size of the job-relevant OCQ (Cohen's *d* = .75) was significantly stronger than the job-irrelevant OCQ (Cohen's *d* = .43). However, even with the job-relevant OCQ, the effect size (Cohen's *d*) of the job-relevant OCQ is still lower than the effect size of the SDR scales.

To improve the effectiveness of the faking detection method, three faking detection methods that rely on the internal item response pattern were developed. They are the idiosyncratic item method (Kuncel & Borenman, 2007), covariance index method (Christiansen et al., 2017), and the machine learning method (Calanna et al., 2019). All these methods utilize the property that test-takers have different response patterns when responding honestly compared and when they respond to fake good to make a good impression. These faking detection methods develop a numeric index of unusualness based on the response patterns expected under honest conditions. By comparing the faking index generated to a threshold value, researchers can classify test-takers into either “honest” or “fake” respondents. The difference between these three methods is the rationale developed and the threshold value used.

The covariance index method was developed based on the observational findings that the correlations between some item-pairs are higher under high-stakes testing conditions (Schmitt & Ryan, 1993; Burns & Christiansen, 2006; Douglas, McDaniel, & Snell, 1996; Schmitt, Ryan, Steirwalt, & Powell, 1995). Given this pattern of increased correlations, Christiansen et al. (2017) developed a metric that indexes the increase in covariance due to faking between items. Christiansen and colleagues identified item pairs from the Goldberg Adjective Markers (Goldberg, 1992) with extremely low to no observed correlations (ranging from -.03 to +.03) in low-stakes testing conditions. The covariance index is computed by cross-multiplying standardized item scores with unrelated paired items and taking the average of those products. Using a decision threshold determined by logistic regression, the covariance index approach can achieve 61.5% classification accuracy with a Cohen’s *d* equal of 1.18 between the honest and faking respondents.

The idiosyncratic item method is based on the empirical finding that, for some items, test takers cannot agree on what endorsement level reflects maximal social desirability (Kuncel & Tellegen, 2001). Therefore, item response distributions on these items will have roughly normal distributions under honest conditions but will have multimodal distributions (e.g., bi-modal or tri-model) under faking conditions (Kuncel & Borenman, 2007). The authors used the Goldberg Adjective Markers (Goldberg, 1992) to identify 11 such items. They constructed the empirical keying scale using a within-group design that all participants answered the personality test honestly and then were instructed to answer the same personality test as if they were applying for a desirable job. They implemented a 2-fold cross-validation process that used half of the honest response sample and half of the job application response sample to create an empirical keying scale. Using the other half of the respondents, they then cross-validated this empirical keying scale. They then repeated this process by using the cross-validation sample in the prior step to develop a new empirical keying scale and used the sample that created the empirical keying scale in the prior step to cross-validate the new empirical keying scale. The scoring scheme of the empirical keying scale depended on the response pattern difference between the honest condition sample and the faking condition sample. Each response anchor represented a score ranging from negative 1 to positive 1 depending on the endorsement frequency difference between the honest condition sample and the faking condition sample. A response anchor would receive a positive 1 if the endorsement frequency in the honest condition sample were higher than in the job application condition sample. A response anchor would receive a negative 1 if the endorsement frequency in the honest condition sample were lower than in the job application condition sample. The use of a response anchor will receive a zero if the endorsement frequency in the honest condition sample is roughly the same as the endorsement frequency in the job application condition sample. Under this scoring scheme, if the sum score of the 11 items was less than zero, participants were identified as honest respondents. On the other hand, if the sum score of the 11 items was larger than zero, participants were identified as fakers. In the two cross-validation samples, the classification accuracy was 71.7% and 82%, and the effect size (Cohen's *d*) was 1.4 and 1.7, respectively.

The machine learning method takes a different approach to classify fakers and honest respondents. Instead of making any preliminary assumptions about the item-level response pattern difference between fakers and honest respondents, the machine learning method makes a more general assumption: the item-level response pattern of fakers and honest respondents is described by a series of item-level mean differences and interactions, and this function can be modeled by using the observed data. Calanna et al. (2019) viewed classifying fakers and honest respondents as a supervised learning task (predicting objective labels in a dataset) for machine learning models. They used ensemble tree models, random forest, and extreme boosted trees (XGB) to learn a collection of the item–level respondent rules/patterns that differentiate between the faking respondents and honest respondents. They found both the random forest and XGB models can achieve classification accuracies of 74% and 76% with effect size (Cohen's *d*) equal to 1.3 and 1.41, respectively.

Both the random forest and the XGB model use the classification and regression tree (CART) method as their base learner, which is then expanded upon by the random forest and XGB algorithms. CART uses a recursive partitioning method to systematically build a set of extensive hierarchical rules to construct a decision tree to classify a datapoint's category (e.g., fake/honest) based on the value of the predictors in a sample. After developing a binary rule based on a cutoff-value of the predictor variable that best partitions the outcome, the tree finds a subsequent rule to partition the outcomes further until a stopping criterion is reached or the outcomes are perfectly separated (Putka, Beatty, & Reeder, 2018). CART is used as the learner rather than other algorithms (e.g., linear regression) because decision trees are a universal function approximator. This property means that they can approximately represent any arbitrary continuous function (rules/patterns) among a set of predictors and outcomes (Gopalan, Nisan, Servedio, Talwar & Wigderson, 2016). Thus, regardless of the level of complexity of the relationships between predictors and outcomes are, CART can reproduce the observed linear and non-linear relationships between outcomes and predictors if given a large enough sample size

(James, Witten, Hastie, & Tibshirani, 2013; Kuhn, & Johnson, 2013). Because CART can reproduce any complex function, given enough data, it is likely to provide improved predictive ability compared to other internal detection methods that use a linear decision function to make predictions.

Although CART is flexible at learning complex relationships, this strength means that it tends to overfit the training data by treating the unsystematic exceptions in the data as systematic rules, which leads to low generalizability. That is, CART readily approximates the relationship between predictors and outcomes, which will incorporate random noise from the training data into its learned rules (Putka et al., 2018). To address this limitation, researchers use random forests and XGBs to maintain low bias when creating rules but also provide better generalizability and avoid learning unsystematic relationships. Each uses a unique approach to construct decision trees. The random forest model attains high generalizability and low bias by using an ensemble of small decision trees. Each decision tree is constructed only on a random subset of predictors and a random sample of the participants. Specifically, each decision tree learns its rules using only a smaller (e.g., 20% of the predictors and 50% of the participants) bootstrapped dataset sampled from the larger training sample. The final output is computed based on averaging all the prediction outputs from all the individual small decision trees. Averaging the outputs of all decision trees allows the random forests to learn the systematic data pattern in the training dataset because the trees are not allowed to be large enough to make unnecessary exceptions, focusing instead on the most important predictors. Complex relationships can be learned because the trees each use different combinations of the predictors when making rules. Therefore, random forests have a low bias, learning complex relationships found within training data while maintaining high generalizability in its predictions of new data points (Breiman, 2001). On the other hand, the XGB model addresses this limitation with a sequential way of building decision trees, conditional on the types of errors made by previous decision trees. The XGB model builds a small decision tree to fit the data one after another (Chen & Guestrin, 2016). Each new tree aims to correct its predecessors' mistakes, and a small learning rate is multiplied with each tree to limit how fast the mistakes are corrected (Géron, 2019). Like the random forest, every tree only utilizes a small proportion of predictors in the tree building process. Unlike random forests, in XGB, the trees are built in a sequential order, where the subsequent tree gives greater weight to the mistakes the prior trees made when constructing rules that minimize misclassifications. With this stepwise process, less noise is included in the XGB model. As a result, higher generalizability can be achieved in the predictions of new data points (Chen & Guestrin, 2016).

In comparison to the faking detection methods that rely on external scales, the faking detection methods that rely on the internal item response pattern generally achieve higher classification accuracy. Therefore, they are better candidates to be used along with the identify and remove correction strategy.

3. The Imbalanced Sample Problem in Real-World Application

Even though empirical evidence suggested the faking detection methods that rely on the internal item response pattern achieve higher classification accuracy than the traditional faking detection methods that rely on external scales, a substantial limitation in these previous studies precludes researchers from drawing a definitive conclusion regarding the effectiveness of these strategies in actual selection contexts. This is because all prior research evaluating these faking detection methods was based on test samples containing roughly equal proportions of the faking and the honest respondents. However, these are often not the case in real-world selection contexts. Past research consistently suggests that about 23%-31% of job applicants faked personality assessments in the selection context (Peterson et al., 2008; Donovan et al., 2008; Griffith et al., 2007). In other words, the fake-to-honest ratio (FHR) in a real-world job applicant sample is often imbalanced. Detection methods often fail by underpredicting the lower base rate outcome in an imbalanced sample (Japkowicz, 2000). When the test sample is imbalanced, for example, only 20% of the individuals in the sample are fakers, a faking detection method classifying all the test takers as honest respondents can still achieve 80% classification accuracy. Therefore, to evaluate whether a faking detection method is effective in an imbalanced sample, precision (the true positive rate of a specific class) and recall (the fraction of the total amount of relevant class that was retrieved) are also needed to calculate beyond the overall classification accuracy. Because job applicant samples are more likely to have imbalanced samples than balanced samples, the reported performances in the literature are likely optimistic estimates. They may be far lower when detecting under realistic proportions of imbalance. Additionally, some methods may be more robust to faking imbalances.

To address this problem, Nie et al. (under review) conducted a pseudo simulation study to investigate whether the faking detection methods (covariance index method, idiosyncratic item response method, and machine learning method) that rely on internal item response pattern can effectively classify fakers and honest respondents with a wide range of faker to honest respondent ratios. They first used the same training sample to develop the covariance index method, the idiosyncratic item response method, the XGB model, and the random forest model. They then created new simulated samples with FHR range from .1 to .9 based on a new test sample. After applying all these faking detection methods to the simulated samples, they found all three faking detection methods could effectively (the precisions of both the honest respondents and fakers were higher than 50%) classify fakers and honest respondents across a wide range of fakers to honest respondents ratios (random forest model, XGB model, and idiosyncratic item response method effectively classified fakers and honest respondents when the FHRs are between 0.2 to 0.8, and the covariance index method effectively classified fakers and honest respondents when the FHRs are between 0.2 to 0.7). Among these faking detection methods, XGB and random forest model achieved almost identical overall classification accuracy across all fakers to honest respondent ratios; both the XGB and random forest models outperformed the other two faking detection methods in terms of the overall classification accuracy and effect size and matched or exceeded the performance of the next best faking detection method in terms of precision and recall both honest respondents and fakers classes in all fakers to honest respondent ratios. This study suggested even though all four faking detection methods can effectively classify honest and faking respondents across a wide range of fakers to honest respondent ratios, the machine learning faking detection method offers the best outcome to classify honest respondents and fakers in samples that resemble real-world selection context.

4. Study Purpose and Research Questions

As suggested in section 2.3, the effect of the identify and removal strategy heavily relies on the classification accuracy of the faking detection method. The purpose of the current study is to test the effectiveness of the identify and removal correction strategy based on pattern-based faking detection methods that have relatively high classification accuracy (machine learning method) and another pattern-based faking detection method that has relatively low classification accuracy (covariance index method). More specifically, I will answer four research question that concerns the impact of faking on personality assessment in selection context from the perspectives of the proportion of fakers being selected and the group mean criterion score in the selected group.

I chose the random forest model over XGB to develop the machine learning method because all the evaluation metrics were identical between random forest and XGB in Nie et al. (under review). However, compared to XGB, the random forest model requires significantly lesser computational power and time to train. Therefore, the random forest has more applied value when practitioners only have limited resources and under a time constraint.

Even though research suggests all current faking correction strategies have little to no impact on correcting for the criterion-related validity and on the true job applicant ranking, a significant number of researchers and practitioners continue to use them to accompany personality measurement (Goffin & Christiansen, 2003). This finding suggests that both researchers and practitioners are deeply concerned about the impact of faking on personality measurement leading them to use correction strategies that have no empirical evidence support. The finding of the current study might provide a useful tool for both practitioners and researchers to alleviate their concern on the impact of faking on personality measurement.

As discussed in the prior section, the evidence on whether faking impacts the criterion-related validity is mixed. However, in these studies, researchers often ignored the impact of fakers to honest respondent ratios on criterion-related validity. Zickar et al.'s (1996) finding suggests the faker to honest respondent ratios play a significant role in determining the impact of faking on criterion-related validity. To evaluate the effectiveness of a correction strategy, we need to test whether this correction strategy can improve the criterion-related validity in different faker inclusion ratios. Therefore, for the current study, I explore the following research question that concerns the group mean criterion score in the selected group:

1. Can the identify and removal correction strategy based on the random forest faking detection method and the covariance index method improve the criterion-related validity compared to no correction across different faker inclusion ratios?

Studies that investigated the impact of faking on job applicant ranking found that, as the SR decreases, the portion of fakers being selected in the sample increases (Ellingson et al., 1999; Rosse et al., 1998). These findings were not surprising, given that fakers tend to inflate their trait scores to get to the top of the job applicant ranking. To evaluate the effectiveness of a correction strategy, we need to test whether the correction strategy can reduce the proportion of fakers being selected in different SRs and different faker inclusion ratios in the sample. Therefore, for the current study, I explore the following question that concerns the proportion of fakers being selected in selection context:

2. Can the identify and removal correction strategy based on the random forest faking detection method and the covariance index method reduce the proportion of fakers being selected across different FHRs and SRs?

Even though the findings on whether faking reduces criterion-related validity were mixed, Mueller-Hanson, et al. (2003) and Donovan et al. (2013) found when implemented a top-down selection strategy, the group mean criterion score of the selected test-takers who were motivated to fake was significantly lower than the selected test-takers who were not motivated to fake. Therefore, to evaluate the effectiveness of a correction strategy, we need to further test whether the implementation of the faking correction strategy can improve the mean level of the criterion in the selected group. Therefore, for the current study, I explore the following question that concerns the group mean criterion score in the selected group:

3. Can the identify and removal correction strategy based on random forest faking detection method and the covariance index method improve the mean level of the criterion in the selected group across different FHRs and SRs?

Lastly, the classification accuracy of the faking detection method should have a strong impact on the effectiveness of the identify and removal strategy. This is because when the correlation between the predictor and the criterion is weaker in the faking group than the honest group, correctly removing as many fakers as possible while keeping as many honest respondents in the job applicant sample as possible can reduce the negative impact of faking on both the proportion of fakers being selected and the group mean criterion score in the selected group. However, no study has investigated how much impact the classification accuracy has on the criterion-related validity in the job applicant sample, and the faker inclusion ratios, and the group mean criterion score in the selected group. Therefore, for the current study, I explore the following question that concerns both the proportion of fakers being selected and the group mean criterion score in the selected group:

4. How much does the identify and removal correction strategy based on random forest faking detection method and the covariance index method differ in terms of post-correction performance metrics (validity coefficient in the job applicant sample, faker inclusion ratios, and the criterion group mean score in the selected group) across different FHRs and SRs?

5. Method

Given that the faking detection method based on a random forest model has been developed in a prior study (Nie et al., under review), for the current study, I will apply this faking detection method along with the identify and removal correction strategy on a new data set to test the effectiveness of the correction strategy.

5.1 Participants and Procedures for Developing the Random Forest Faking Detection Method

The random forest faking detection method was developed based on the dataset provided by Zhang et al. (2019). Zhang et al. (2019) recruited 1,066 participants through Amazon Mechanical Turk (Mturk) in 2012. Each participant received 55 cents (USD) for their participation. All participants were randomly assigned to either faking or honest conditions. Participants in the faking condition were instructed to complete the personality assessment as if they were applying for a job that they desired and select the response that would make them look like the best job candidate. Participants in the honest condition were instructed to complete the personality assessment honestly. Two quality control items (e.g., "Please click 'agree' for this question") were embedded to control for careless responses. Participants who incorrectly respond to either of these quality control items were excluded from the analysis. 947 participants (458 in the faking condition and 489 in the honest condition) were included in the dataset. 58.9% of the participants were female. The average participant age was 31.96 years (SD = 11.85). 79.5% of the participants were Caucasian, followed by 6.5% of African American, 6.1 % of Asian, 4.4% of Latino/Hispanic, and 2.9% were other ethnicities not specified. There was no significant difference in any of the demographic variables between the two groups.

5.2 Development of the Random Forest and Covariance Index Faking Detection Method

To develop a random forest model that can detect faking, I used the item-level data to train the random forest model to classify the respondent as either “honest” or “fake.” A random forest model has many potential hyperparameters, which adjust the complexity of learning and criteria used to develop rules. I varied the hyperparameters of how deep an individual tree could be (from 3 to 10 or infinity), the number of predictors considered at each split (from 10% to 90%), whether cases were bootstrap re-sampled with replacement (yes or no), and what objective criterion was optimized when splitting cases (gini or entropy). The hyperparameters were chosen through 10-fold cross-validation on the training data. 10 folds were chosen following the recommendation of Kuhn & Johnson (2013) to balance bias and variance trade-off. All possible combinations of hyperparameters were evaluated, and the combination that produced the best cross-validated receiver operating characteristic area under the curve (ROC AUC) score was used to fit a final random forest model using the entire training dataset. The receiver operating characteristic curve (ROC) describes the trade-off between a true positive rate and a false positive rate of all classification thresholds (Murphy, 1996). The area under the ROC is often used to evaluate the efficiency of a machine learning model in a classification task because it has a clear reference point. The ROC AUC has a range of 0 to 1. A ROC AUC score of 1 indicates the classification machine learning model has 100 % classification accuracy. A ROC AUC score of 0 indicates the classification machine learning model has 0 % classification accuracy. When ROC AUC = .5, the machine learning model’s predictions are equivalent to guessing (i.e., either randomly predicting an outcome, or predicting the majority category every time). When ROC AUC < 0 .5, the machine learning model is worse than guessing, and when ROC AUC > .5, the machine learning model is better than guessing. The hyperparameters were selected based on the best cross-validated ROC AUC score using only the training sample.

I calculated the covariance index using the guidelines described by Christiansen et al. (2017). I used the entire honest response condition sample and calculated the pairwise correlation between all possible pairs of the 50 items (n = 1,225). Any items whose correlation was between -.03 and +.03 were considered uncorrelated, resulting in 127 uncorrelated item pairs. I then separated each item pair into two groups: first item and second item. I then calculated the mean and standard deviation of each group to use when standardizing the values. The covariance index for a participant is calculated by standardizing the values for an item pair using the mean and standard deviation of the items within its group. I then multiplied those standardized values and repeated the process for all other item pairs. A participant's covariance index score is the average of these products. Higher values indicate greater covariance inflation, and therefore, a greater likelihood of faking. To classify participants as either "fake" or "honest" and provide probabilities for the predictions, I fit a logistic regression during the training phase using the covariance index score as the predictor and faking status as the outcome.

5.3 Participants and procedures for the current study

For the current study, I used another archival dataset provided by Zhang et al. (2019) in their second part of the same study. Zhang et al. recruited another 499 participants from Mturk. 53% of the participants were female. Their average age was 36.92 years old (SD = 11.08). The sample was predominantly White (81.53%), followed by African American (10.55%), Asian (7.91%), Native American Indian or Alaska Native (2.20%), and Other (2.20%).

Participants were randomly assigned to one of the two conditions with a counterbalancing procedure: 1) respond honestly at time 1 and fake at time 2 (the honest-fake condition), and 2) fake at time 1 and respond honestly at time 2 (the fake-honest condition). Participants in two conditions had a 14-day interval between time 1 and time 2. This introduction of the time interval was to minimize the potential carry-over effects. Participants were paid 2 dollars (USD) for the completion of time point 1 study and another 3 dollars if they completed the time point 2 study. The higher compensation provided to the participants in the second time point of the study was to incentivize the higher retention rate of the study. Given that participants were randomly assigned to two conditions, the difference in two points of compensation should not create a confounding variable for the research outcome. Participants who were assigned to the faking condition were instructed to imagine that they were "applying for a job that you want very much," so they should "select the response that will make you look like the best job applicant." Participants who were assigned to the honest condition were told that their "answers will be used for research purposes only," so they should "answer the questions as honestly as possible." The instructions for time point 1 and time point 2 were the same as the study that developed the random forest faking detection method. At each time point, participants were instructed to first fill out the demographic measurement and complete the personality measurement based on either faking or responding honestly. Then they were instructed to complete the criterion measurements honestly in both conditions for all time points. There is a concern that participants were primed by their assigned condition to perhaps also respond honestly or fake when answering the criterion variables. However, Zhang et al. (2019) conducted t-tests on the mean difference scores of all the criterion variables, respectively. They found no evidence that the mean scores were significantly different (CWB: t = -.89, Cohen's d = -.05; OCB: t = 3.27, Cohen's d = .18; job satisfaction: t = .29, Cohen's d = .02; job stress: t = -.92, Cohen's d = .05; turnover intention: t = -.16, Cohen's d = -.01; life satisfaction: t = .11, Cohen's d = -.01; work status: t = .83, Cohen's d = .05). This finding suggests that participants' responses on all the criterion variables were not affected by the conditions they were assigned to, refuting the possibility that a priming effect occurred.

To ensure high data quality, Zhang et al. (2019) embedded five quality control items and two manipulation check items in the survey. Participants who missed more than one quality control item or any manipulation check items were excluded from the dataset. A total of 457 participants were included in time 1 of the data collection, with 238 participants in the honest condition and 219 participants in the faking condition. 388 participants were included in time 2 of the data collection time, with 197 participants in the honest condition and 191 participants in the faking condition. No significant demographic differences were found between the two conditions.

Given that participants were paid $.55 in the data collection procedure for the training dataset used to develop the random forest faking detection method, and the participants were paid $2 in the data collection procedure for the testing dataset for the current study, participants who received $2 might put into more effort to complete the survey compared to the participants who received $.55, because of stronger incentives. This difference in incentives might further lead to a systematic difference in the distributions of the two datasets, which might artificially improve the classification accuracy of the random forest faking detection model. To rule out this concern, I constructed 95% confidence intervals (CIs) for the mean trait scores in both datasets. These are presented in Table 1. All the 95% CIs of the Big 5 trait scores between the training and testing dataset in both honest and faking condition are overlapping except the mean trait score of conscientiousness in the honest condition($\text{Conscientious}_{\text{train}}$ = 2.83, (95% CI = 2.79, 2.87); $\text{Conscientious}_{\text{test}}$ = 3.08, 95% CI = (3.02, 3.148). However, it is the only difference in all 10 pairs of comparisons of mean trait scores. Nie et al. (under review) found, the overall classification accuracies of the random forest faking detection method range from .839 to .875 across 9 faker-to-honest respondents (range from .1 to .9) with the same training and testing datasets. This finding suggested random forests can achieve high classification accuracy even with a minor distribution difference in training and testing datasets.

5.4 Measures

Both the Comprehensive Personality Scale (CPS; Wang, 2013) and the International Personality Item Pool (IPIP; Goldberg, 1999) were included in the dataset used to develop the random forest faking detection method and the current study. For simplicity, I only developed the random forest and the covariance index faking detection method with the IPIP data. Therefore, I will only test the effectiveness of the identify and removal strategy based on the random forest and the covariance index faking detection method using the IPIP data.

In the IPIP scale, there were 10 items for each of the Big Five personality scales. All items were presented in random order. Participants were asked to respond to the item on a 4-point Likert-type scale, where 1 = “Strongly Disagree”, 2 = “Disagree”, 3 = “Agree”, and 4 = “Strongly Agree”.

Seven criteria, Counterproductive Work Behaviors (CWB), Organizational Citizenship Behaviors (OCB), Job Satisfaction (Jsat), Job Stress (Jstress), Turnover Intention (TO), Life Satisfaction (Lsat), and Workplace Status (WS) were included in the dataset for the current study. Among these criteria, I used turnover intention as the criteria measurement. I detail the reason in section 5.5 simulation design.

There might be a concern of using a self-report measure as a criterion because participants might distort their response due to social desirability (Krumpal, 2013). However, this is likely not a significant concern for the current study. Even if some participants distorted their response to the turnover intention items unconsciously, they were randomly assigned to the faking and honest conditions. Therefore, their impact on the current study is minimized. The findings mentioned in the prior paragraph that the mean scores of the turnover intention in the faking and honest conditions were not significantly different corroborate this argument. Moreover, the group mean turnover intention of the participants in both the faking condition and the honest condition is not significantly different. If the responses of the turnover intention items are distorted, it will reduce the variance of the outcome (because people are using the high extreme of the scale less often). Therefore, the results would be attenuated due to restriction of range and would only be expected to be stronger if turnover intention could be measured without distortion.

5.5 Simulation Design

A pseudo-simulation design was created for the current study. Four simulation conditions: FHRs, sample size, predictor-and-criterion, and SRs were introduced in the simulation process. I created a new sample by combining the honest and faking conditions in the second part of the dataset; this resulted in 417 participants in the honest condition and 410 participants in the faking condition.

FHRs. For the FHR conditions, multiple simulation samples were created based on randomly selecting participant data from the honest and faking conditions of the sample. The FHRs had four levels: 50%, 30%, 15% and, 0%. Given the majority of the faking research in the lab studies had a FHR roughly equal to 50%, I chose 50% percent as one of the FHRs as a comparison to the prior faking research. The 30% and 15% FHRs were chosen based on the empirical findings suggesting the proportion of fakers in the job applicant sample is around 32%-

23%. The 0% FHR provided a benchmark on the true criterion-related validity when no faking occurs.

Sample Size. I used two sample sizes, 50 and 200 to create the simulation samples. I chose these two numbers to emulate the situations in the selection context that have a relatively large number of job applicants and a relatively small number of job applicants. As Schmitt and Oswald (2007) suggested, the sample size should not systematically influence the findings of random sampling simulation, but it would inversely influence the amount of the sampling error variance in the result. Given the current study aims to compare the performance of the identify and removal strategy with two faking detection methods with the performance of no correction, I included two sample sizes as one of the simulation conditions to exam the robustness of the results.

Predictors-and-Criterion. I only used the conscientiousness and emotional stability trait scores in the current study. Meta-analysis on criterion-related validity found conscientiousness and emotional stability have the highest criterion-related among big 5 trait scores (Schmidt, Oh,& Shaffer, 2016) in organizational settings. This finding suggests making selection decisions with conscientiousness and emotional stability should result in high group mean criterion score in the selcted group. However, when individuals engage in faking, they also distort the trait scores of conscientiousness and emotional stability to the largest degree among the Big 5 traits. In a meta-analysis, Viswesvaran and Ones (1999) found conscientiousness and emotional stability have the strongest effect size among Big-Five personality traits in faking studies that use a within-person design(Viswesvaran & Ones, 1999). This might due to the personality items that assess conscientiousness and emotional stability are the most transparent in terms of social desirability among other personality items (Ones, Viswesvaran & Reiss, 1996). Therefore, I used the conscientiousness and emotional stability trait scores as the simulation conditions.

I used only turnover intention among the seven criteria measured in the dataset. Research suggests turnover intention as a strong predictor of voluntary turnover. Both Chao and Lewis (2012) and Cohen, Blake, and Goodman (2016) found a significantly strong correlation (r = .7 or higher) between turnover intention and probability of voluntary turnover using individuals as the units of analysis for most demographic groups. Voluntary turnover often relates to the significant cost of an organization. This is because other than the replacement fees; there are hidden costs such as productivity loss, workplace safety issues, and moral damage(O'Connell & Kung, 2007). Even though both OCB and CWB also have a strong dollar impact on an organization (OCB: Podsakoff & MacKenzie, 1994, Podsakoff, Ahearne, & MacKenzie, 1997; CWB: Camara & Schneider, 1994; Vardi & Weitz, 2004), the correlations between personality trait score (conscientiousness and emotional stability) with OCB and CWB in the dataset used in the current study did not meet one of the assumptions of the identify and removal strategy. To effectively correct faking, the identify and removal strategy requires the criterion-related validity for the honest respondent to be higher than the faking respondent. The correlations between conscientiousness and emotional stability with OCB were 0.185 and 0.184 in the faking condition and 0.172 and 0.19 in the honest condition. The correlations between conscientiousness and emotional stability with CWB were -0.409 and -0.428 in the faking condition and -0.222 and -0.269 in the honest condition. On the other hand, the correlations between conscientiousness and emotional stability with turnover intention were -0.034 and -0.083 in the faking condition and -0.289 and -0.423 in the honest condition. Therefore, only turnover intentions in the sample met the assumption. This finding implies that when using a faking detection method, we would not expect all measures of potential interest to improve via the method.

SRs. I used five levels of SRs (10%, 20%, 30%, 40%, 50%) to create selective groups. Schmitt and Oswald (2008) found data patterns with only three levels of SRs in their simulation study that investigated the effect of identify and removal strategy. Therefore, five levels of SRs should be enough to identify a pattern of findings that would be able to indicate the values at other SRs.

To answer research questions 1 and 4, only the first three simulation conditions but not the SRs were required. The following three steps were implemented. Step 1, I created simulation samples based on each of the FHRs and sample size. Step 2, I calculated the simple correlations between emotional stability, turnover intention, conscientiousness, and turnover intention. Step 3, I then applied the identify and removal correction strategy based on the random forest faking detection method to each simulation dataset to remove all the predicted fakers inside each simulation dataset to create a new simulated sample. After the predicted fakers were removed from all the simulation datasets, I recalculated the same prediction quality metrics for each new simulated sample again. I then repeated step 3 for the application of the identify and removal correction strategy based on the faking detection methods. To illustrate this simulation process better, please refer to Figure 2 as an example for the simulation condition, when faking ratio is 30%. I repeated this process multiple times to create multiple simulation samples to create a 95 percent confidence interval (95% CI) for each prediction performance metric. Given it is not immediately known how many repetitions were needed to have stable estimations of the metrics, I determined the number of repetitions needed for appropriate precision by estimating multiple numbers of repetitions until I reached an asymptotically low range of the largest 95% CI. I started at 1000 repetitions and then increased the number of repetitions by 1000. The 95% CIs of the correlations between emotional stability and turnover intention in the faking condition among all the calculated metrics were overlapped. This suggested all the calculated metrics were stabilized after 1000 repetitions. Therefore, I used 1000 repetitions for the current study. After the repetition was completed, I calculated the 95% confidence intervals for all the metrics based on the average value of the simulated samples for the metrics computed based on the simulation datasets before the correction and the simulation samples after the correction at each level. If the correlations on the before-correction-simulated samples were significantly lower than the same metrics calculated based on the after-correction-simulated samples, I would have found evidence to support the use of the identify and removal strategy that is based on response pattern based faking detection methods to improve criterion-related validity. If the correlations between the predictors and the criterion of the posted-correction sample created by the random forest faking detection method were stronger than the same correlations of the posted-correction sample created by the covariance index faking detection method, I would have found evidence to support the use of the faking detection method that has higher classification accuracy would result in higher criterion-related validity.

The following seven steps were implemented in the simulation process to answer research questions 2 and 4. Step 1, I created simulation samples based on each of the FHRs and sample size. Step 2, participants in the simulation samples were selected into the selected group based on the SRs. Step 3, faker-inclusion in the selected group was calculated for the post-correction selected groups. Step 4, the identity and removal strategy were applied to the simulated sample to remove predicted fakers from the simulated sample. Step 5, participants in the simulation samples after the correction were selected into the selected group following the same procedure in the pre-correction process. Step 6, faker-inclusion in the selected group was calculated for the post-correction selected groups. To illustrate the simulation process better, please refer to Figure 2 as an example for the simulation conditions, the faking ratio is 30%, and the SR is 10%. I calculated the 95% confidence interval and the mean level of the ratios of the faking and honest participants in the selected group within each cross-over condition after 1000 repetitions. If the proportion of faking participants in the selected group of applicants was consistently lower after the correction strategy than the proportion before the correction strategy was applied, I would have found evidence to support the use of the identity and removal correction strategy. This decrease in selected fakers would ideally hold across all SRs. That is, the identity and removal correction strategy is an effective method to reduce the proportion of fakers being selected. Moreover, if the faker inclusion ratios in the selective groups of the posted-correction sample created by the random forest faking detection method was lower than the faker inclusion ratios of the posted-correction sample created by the covariance index faking detection method, I would have found evidence to support the use of the faking detection method that has higher classification accuracy would result in lower faker inclusion ratios in the selective groups.

To answer research question 3, the simulation procedure was the same as the simulation procedure which was implemented to answer research question 2 with one difference. Instead of calculating the faker inclusion ratios for the selected groups, I calculated the group mean scores for the turnover intention in the selected group. I calculated the 95% confidence interval and the group mean level of turnover intention in the selected group within each cross-over condition after 1000 repetitions. If the group mean level of the turnover intention of the applicants were consistently lower after the correction strategy than the proportion before the correction strategy was applied, I would have found evidence to support the use of identify and removal correction strategy. This decrease in the group mean level of turnover intention would ideally hold across all SRs. That is, the identify and removal strategy is an effective method to improve the group mean criterion score in the selected group. Moreover, if the group mean turnover intention score in the selective groups of the posted-correction sample created by the random forest faking detection method was lower than the faker group mean turnover intention score of the posted-correction sample created by the covariance index faking detection method, I would have found evidence to support the use of the faking detection method that has higher classification accuracy would result in lower group mean turnover intention score in the selective groups.

6. Results

All the descriptive statistics and the correlation table for the honest and faking sample are reported in Tables 2 and 3. The correlation between the two predictors conscientiousness and emotional stability was significantly stronger in the faking condition compare with the honest condition. This finding was alight with the finding from prior studies (e.g., Burns & Christiansen, 2006; Douglas, McDaniel, & Snell, 1996). The honest condition included 417 participants and 410 participants were in the faking sample. The effect size (*Cohen's d*) of the mean difference between the honest sample and the faking sample for conscientiousness and emotional stability were 1.2 and 1.3, respectively. This finding is stronger than the sample size weighted mean effect size estimated in the meta-analysis findings (conscientiousness = .89; emotional stability = .93) but within the range of the 90% CIs (conscientiousness = .14 to 1.63; emotional stability = .31 to 1.54) (Viswesvaran & Ones, 1999). To evaluate the impact of identify and removal strategy on the two faking detection methods of interest before the simulation began, I also calculated the descriptive statistics and the correlations after removing the suspected fakers identified by the two faking detection strategies respectively. The descriptive statistics and the correlation table for the honest and faking samples that were corrected by the covariance index faking detection method are reported in Figures 4 and 5. With the samples that were corrected by the covariance index faking detection method, 374 participants remained in the honest sample and 252 participants remained in the faking sample. This resulted in a precision rate of 89.6% for honest classification, a precision rate of 61.4% for faking classification, and overall classification accuracy of 75.6%. The descriptive statistics and the correlation table for the honest and faking samples that were corrected by the random forest faking detection method are reported in Tables 6 and 7. With the samples that were corrected by the random forest faking detection method, 356 participants remained in the honest sample and 337 participants remained in the faking sample. This resulted in a precision rate of 85.3% for honest classification, a precision rate of 82.1% for faking classification, and an overall classification accuracy rate of 83.7%.

The first research question and the fourth research question concerned how the validity coefficients differed between no correction and the corrected by the identify and removal strategy when using the random forest or covariance index faking detection method. Figure 3 (simulation result based on emotional stability) and figure 4 (simulation result based on conscientiousness) for the simulation on conscientiousness and turnover intention was created for all the simulation conditions with job applicant sample size = 200 and figure 5 and 6 were created for all the simulation conditions with job applicant sample size = 50. With an eyeball examination, there was no obvious pattern difference between the figures for sample size = 200 and sample size = 50. I further created two vectors ($Ʊ_1$ and $Ʊ_2$) that represent all the findings (the mean values) from table 8 and table 9 (simulation results based on emotional stability) for job applicant sample size of 200 and 90 respectively, and another two vectors ($Ʊ_3$ and $Ʊ_4$) that represent all the findings from table 8 and table 9 (simulation results based on conscientiousness). The correlation coefficient between $Ʊ_1$ and $Ʊ_2$ was 0.999 and the correlation coefficient between $Ʊ_3$ and $Ʊ_4$ was 0.998. Given all the variables the simulations shared the same scale. These findings suggested the simulation results from the simulation conditions for job applicant sample size = 200 and job applicant sample size =50 were equivalent. Therefore, I will focus on interpreting figures 3 and 4 to answer these two research questions.

Based on all the simulation samples, as the FHRs increased, the validity coefficient weakened. However, the steepness of the slopes varied. With no correction, the simulation samples had the steepest slope between faker proportion and the validity coefficient, while the simulation samples corrected by the covariance index or the random forest faking detection methods had the flatter slopes. This data pattern suggests that the increase in fakers more strongly weakens the validity coefficient of the no correction simulation than when the sample is corrected by the covariance index or the random forest faking detection methods as the FHRs.

In figure 3, starting at the FHR = 0.15, the validity coefficients of the emotional stability and turnover intention were stronger for the simulation samples corrected by the random forest faking detection method ($r_{(0.15\ FHR)}$ = -0.376 (-0.38, -0.373); $r_{(0.3\ FHR)}$ = -0.369 (-0.373, -0.365); $r_{(0.5\ FHR)}$ = -0.34 (-0.345, -0.335))than the same validity coefficients of the simulation samples with no correction ($r_{(0.15\ FHR)}$ = -0.365 (-0.369, -0.362); $r_{(0.3\ FHR)}$= -0.315 (-0.319, -0.311); $r_{(0.5\ FHR)}$ = -0.252 (-0.256, -0.249)) (Table 8). Based on the same figure, starting at the FHR = 0.3, the validity coefficients of emotional stability and turnover intention were stronger for the simulation samples that corrected by the covariance index faking detection method ($r_{(0.3\ FHR)}$ = -0.336(-0.34, -0.332); $r_{(0.5\ FHR)}$= -0.296 (-0.3, -0.291)) than the same validity coefficients for the simulation samples with no correction (Table 8). Moreover, the validity coefficients of emotional stability and turnover intention for the simulation samples that were corrected by the random forest faking detection method were always stronger than the same validity coefficients for the simulation samples that corrected by the covariance index faking detection method across all four FHRs (again, see Table 8). This data pattern suggested, even in a job applicant sample with low FHR (0.15), removing the potential fakers who are identified by the random forest faking detection method can increase the validity coefficient for emotional stability and turnover intention over the no correction approach. However, the FHR needs to be 0.3 or higher before removing the potential fakers identified by the covariance index faking detection method can increase the validity coefficient for emotional stability and turnover intention over the no correction approach.

However, a different pattern was unexpectedly observed in figure 4. The validity coefficients of conscientiousness and turnover intention for the simulation samples corrected by the random forest faking detection method($r_{(0\ FHR)}$ = -0.196 (-0.199, -0.192); $r_{(0.15\ FHR)}$ = -0.185 (-0.189, -0.181); $r_{(0.3\ FHR)}$ = -0.173 (-0.177, -0.168)) were weaker than the validity coefficients for the simulation samples with no correction ($r_{(0\ FHR)}$ = -0.196 (-0.199, -0.192); $r_{(0.15\ FHR)}$ = -0.185 (-0.189, -0.181); $r_{(0.3\ FHR)}$ = -0.173 (-0.177, -0.168)) for 0 – 0.3 FHRs. The 95% CI of the same correlations for the simulation samples corrected by the random forest faking detection method (r (0.5 FHR) = -0.146 (-0.152, -0.14)) and the simulation samples with no correction ($r_{(0.5\ FHR)}$ = -0.139 (-0.143, -0.135)) overlapped for 0.5 FHR in figure 4 (Table 8). All the validity coefficients of conscientiousness and turnover intention for the simulation samples corrected by the covariance index faking detection method ($r_{(0\ FHR)}$= -0.188 (-0.191, -0.185); $r_{(0.15\ FHR)}$ = -0.167(-0.171, -0.163); $r_{(0.3\ FHR)}$ = -0.14(-0.144,-0.135); $r_{(0.5\ FHR)}$ = -0.101(-0.106,-0.096)) were significantly weaker than the same validity coefficients for the simulation sample with no correction and the same validity coefficients for the simulation sample corrected by the random forest faking detection method (Table 8). This data pattern suggested only when the FHR is 0.5 or higher, removing the potential fakers who are identified by the random forest faking detection method can increase the validity coefficient for conscientiousness and turnover intention over the no correction approach. Moreover, because the slopes are flatter, that means that as more fakers are part of the applicant pool (up to 50%), the random forest faking detection method deals with the increase more effectively. On the other hand, even when the FHR is 0.5, removing the potential fakers who are identified by the covariance index faking detection method cannot increase the validity coefficient for conscientiousness and turnover intention over the no correction approach.

Therefore, to answer research question 1, which concerns whether the identify removal strategy based on these two faking detection methods can improve the validity coefficient over no correction, I only found partial support that applying the two pattern-based faking correction methods to job applicant samples with different FHRs can improve validity coefficients of the predictors and criteria. To answer research question 4, which concerns which faking detection method can be more effectively improve criterion validity, I found the validity coefficients of the predictors and criteria for the simulation samples corrected by the random forest faking detection (the higher classification accuracy method) were always stronger than the same validity coefficients for the simulation samples corrected by the covariance index faking detection method (the lower classification accuracy method). This finding suggested, random forest faking detection method is always more effective to improve the validity coefficients across all FHRs.

The second and the fourth research question concern the faker inclusion ratios difference in the final selected group for the simulation samples with no correction and the simulation samples corrected by the identify and removal strategy based on either the random forest or the covariance index faking detection method across FHRs and SRs. Figures 7(simulation results based on emotional stability) and 8 (simulation results based on conscientiousness) were created for all the simulation conditions with job applicant sample size = 200 and figure 9 (simulation results based on emotional stability) and 10 (simulation results based on conscientiousness) were created for all the simulation conditions with job applicant sample size = 50. I further created two vectors ($Ʋ_5$ and $Ʋ_6$) that represent all the findings from table 10 and table 12 (simulation results based on emotional stability) respectively, and another two vectors ($Ʋ_7$ and $Ʋ_8$) that represent all the findings from table 11 and table 13 (simulation results based on conscientiousness). The correlation coefficient between $Ʋ_5$ and $Ʋ_6$ was 0.999, and the correlation coefficient between $Ʋ_7$ and $Ʋ_8$ was 0.999. Given all the simulation variables shared the same scale, these findings suggested the simulation results from the simulation conditions that job applicant sample size = 200 and job applicant sample size =50 were equivalent. Therefore, I will focus on interpreting figures 7 and 8 to answer these two research questions.

In figure 7, which illustrated selecting participants into selected groups with emotional stability, the proportion of fakers being included in the selected group increased as the SRs decreased for the simulation samples with no correction across all fake-to honest ratios. This data pattern suggested when no correction method is applied to the job applicant sample, the proportion of fakers being included in the final selected group increase as the SR decrease. This finding is aligned with the findings from Rosse et al. (1998) and Zickar et al. (1996) that fakers have the unfair advantage of getting hired over honest respondents. The proportion of fakers being included in the selected group for the simulation samples corrected by both of the correction methods only increased as FHRs increased but did not change significantly when the SRs decreased. More importantly, even the highest faker inclusion ratio in the selected group for the simulation samples corrected by either the random forest faking detection method or the covariance index faking detection method were always significantly lower than the lowest faker inclusion ratio for the simulation samples with no correction across FHRs. For the results of the random forest faking detection method, the highest faker inclusion ratio for the FHR 0.15 across SRs was 0.045 (0.039, 0.044), the highest faker inclusion ratio for the FHR 0.3 across SRs was 0.11 (0.101, 0.118). The highest faker inclusion ratio for the FHR 0.5 across SRs was 0.207 (0.196, 0.128) (Table 11). For the results of the covariance index faking detection method, the highest faker inclusion ratio for the FHR 0.15 across SRs was 0.135 (0.129, 0.141), the highest faker inclusion ratio for the FHR 0.3 across SRs was 0.259 (0.253, 0.265). The highest faker inclusion ratio for the FHR 0.5 across SRs was 0.479 (0.471, 0.488) (Table 11). For the results of no correction, the lowest faker inclusion ratio for the FHR 0.15 across SRs was 0.253 (0.251, 0.255), the lowest faker inclusion ratio for the FHR 0.3 across SRs was 0.513 (0.51, 0.516), and the lowest faker inclusion ratio for the FHR 0.5 across SRs was 0.762 (0.759, 0.766) (Table 11). This data pattern suggested, the application of either one of the faking correction methods can significantly decrease the proportion of fakers being included in the final selected group. The highest faker inclusion ratios in the selected group across SRs for the simulation samples corrected by the random forest faking detection method for each FHR was always lower than the lowest faker inclusion ratios across SRs for the simulation samples corrected by the covariance index faking detection method for each FHRs (the lowest faker inclusion ratio for the FHR 0.15 across SRs is 0.099 (0.098, 0.101); the lowest faker inclusion ratio for the FHR 0.3 across SRs is 0.205 (0.203, 0.207); the lowest faker inclusion ratio for the FHR 0.5 across SRs is 0.361 (0.358, 0.363)) (Table 11). This data pattern suggested the identify and removal strategy based on the random forest faking detection method is more effective in decreasing the faker inclusion ratios in the final selected group across SRs and FHRs.

The same data pattern was repeated in figure 8 which illustrated selecting participants into selected groups with conscientiousness. The highest faker inclusion ratios in the selected group across SRs for the simulation samples corrected by the random forest faking detection method for each FHR(the highest faker inclusion ratio for the FHR 0.15 across SRs is 0.045 (0.043, 0.046); the highest faker inclusion ratio for the FHR 0.3 across SRs is 0.102 (0.1, 0.104); the highest faker inclusion ratio for the FHR 0.5 across SRs is 0.208 (0.204, 0.211)) was consistently lower than the lowest faker inclusion ratios across SRs for the simulation samples corrected by the covariance index faking detection method(the lowest faker inclusion ratio for the FHR 0.15 across SRs is 0.096 (0.094, 0.097); the lowest faker inclusion ratio for the FHR 0.3 across SRs is 0.194 (0.192, 0.196); the lowest faker inclusion ratio for the FHR 0.5 across SRs is 0.34 (0.337, 0.342)). The highest faker inclusion ratios in the selected group across SRs for the simulation samples corrected by the random forest faking detection was always lower than the simulation samples with no correction method (the lowest faker inclusion ratio for the FHR 0.15 across SRs is 0.265 (0.264, 0.266); the lowest faker inclusion ratio for the FHR 0.3 across SRs is 0.5 (0.499, 0.502); the lowest faker inclusion ratio for the FHR 0.5 across SRs is 0.749 (0.747, 0.751)) for each FHR (Table 12).

Therefore, to answer research question 2, which concerns whether the identify and removal strategy based on the two faking detection methods can reduce the proportion of fakers being selected, I found support for that the two pattern-based faking correction methods can reduce the proportion of fakers being included in the selected group in job applicant samples that with different FHRs and. To answer research question 4, which concern the performance comparison of the two faking detection methods, I found the faker inclusion ratio in the selected group for the simulation samples corrected by the random forest faking detection was always lower than the faker inclusion ratio in the selected group for the simulation samples corrected by the covariance index faking detection method. These findings suggested the identify and removal strategy based on the random forest faking detection methods is consistently more effective in terms of reducing the faker inclusion ratio in the final selected group.

However, given the faking detection methods could not perfectly classify fakers and honest respondents. It misclassified some honest respondents as fakers. This led to a concern that applying these faking detection methods to remove suspected fakers might remove a significant proportion of honest respondents in the job applicant sample. To assess how many honest respondents were mistakenly classified as fakers by these faking detection methods, I further calculated the proportion of honest respondents being misclassified as fakers by these two faking detection methods across FHRs for the simulation samples that the sample sizes = 200. Across the four FHRs (0, 0.15, 0.3, 0.5), the proportions of honest respondents being misclassified as fakers by the covariance index method were 0.103, 0.102, 0.103 and, 0.101, and the proportions of honest respondents being misclassified as fakers by the random forest method were 0.146, 0.146, 0.145, 0.146. Compared with the random forest method, the covariance index method has a lower rate of misclassifying the honest respondents.

The third and the fourth research question concern the mean criterion score difference in the selected group of the simulation samples with no correction and the simulation samples corrected by the identify and removal strategy based on either the random forest or covariance index faking detection method across FHRs and SRs. Figures 11 (simulation results based on emotional stability) and 12 (simulation results based on conscientiousness) were created for all the simulation conditions with job applicant sample size = 200, and figures 13 (simulation results based on emotional stability) and 14 (simulation results based on conscientiousness) were created for all the simulation conditions with job applicant sample size = 50 to answer these two research questions. I further created two vectors (υ_9 and υ_{10}) that represent all the findings from table 14 and table 16 (simulation results based on emotional stability) respectively, and another two vectors (Ʊ$_{11}$ and Ʊ$_{12}$) that represent all the findings from table 15 and table 17 (simulation results based on conscientiousness). The correlation coefficient between Ʊ$_{9}$ and Ʊ$_{10}$ was 0.997 and the correlation coefficient between Ʊ$_{11}$ and Ʊ$_{12}$ was 0.994. Given all the variables in the simulations shared the same scale, these findings suggested the simulation results from the simulation conditions that job applicant sample size = 200 and job applicant sample size =50 were equivalent. Therefore, I will focus on interpreting figures 11 and 12 to answer these two research questions.

For both figures 11 and 12, the mean turnover intention scores increased as the SRs and the FHRs increased for all the simulation samples regardless of with or without correction. In figure 11, which illustrated selecting participants into selected groups with emotional stability, the mean turnover intention scores for simulation samples with no correction were always lower than the mean turnover intention scores for the simulation samples corrected by either one of faking detection methods across all FHRs and SRs. This data pattern suggested the application of the identify and removal strategy based on either one of faking detection methods increased the turnover intention of the final select groups across all SRs and FHRs. The mean turnover intention scores for the simulation samples corrected by the covariance index were almost always lower than the mean turnover intention scores for the simulation samples corrected by the random forest faking detection method. The only the exception to this finding was when FHR = 0.3 and SRs = 0.1(covariance index faking detection method = 2.587 (2.576, 2.598); random forest faking detection method = 2.572 (2.563, 2.581)) and when FHR = 0.5 and SR = 0.1(covariance index faking detection method = 2.688 (2.677, 2.7); random forest faking detection method = 2.643 (2.633, 2.653)) (Table 15). This data pattern suggested the identify and removal strategy based on the covariance index faking detection method increase the turnover intention of the final select groups to a lower degree compare with the identify and removal strategy based on the random forest faking detection method across most of the SRs and FHRs. A similar pattern was repeated in figure 12, which illustrated selecting participants into selected groups with conscientiousness. The mean turnover intention scores for simulation samples with no correction were always lower than the mean turnover intention scores for the simulation samples corrected by either one of faking detection methods across all FHRs and SRs. The mean turnover intention for the simulation samples corrected by the covariance index was almost always lower than the mean turnover intention for the simulation samples corrected by the random forest faking detection method with only the exception when FHR = .05 and SRs = 0.2 (covariance index faking detection method = 2.91 (2.9, 2.92); random forest faking detection method = 2.9 (2.89, 2.91)) (Table 16).

Therefore, to answer research question 3, which concerned whether the application of identify and removal strategy based on the two faking detection methods can improve the group mean criterion score in the selected group, I found the application identify and removal strategy based on either one of faking detection methods decreased the group mean criterion score. This means the group mean criterion score in the selected group worsen with the application of the identify and removal strategy based on either of the two faking detection methods. To answer research question 4 which concerned the comparison of the group mean criterion score in the selected group when used the random forest faking detection method and the covariance index faking detection method to remove suspected fakers, I found the group mean criterion score in the selected group for the simulation samples corrected by the random forest faking detection was almost always higher than the group mean turnover intention score for the simulation samples that corrected by the covariance index faking detection method. This finding suggested even though the application of the identify and removal strategy based on both faking detection methods worsen the group mean criterion score in the selected group, the covariance index faking method worsened the group mean criterion score to a lower degree than the random forest faking method.

7. Discussion

In the current study, I investigated the effectiveness of the identify and removal strategy with two response pattern based faking detection methods in terms of the proportion of fakers beingh selected and the group mean criterion score in the selected group. To answer the research question concerning whether the identify and removal strategy based on the response pattern based faking detection method can reduce the proportion of fakers being selected when personality assessment is used, I found that identify and removal strategy (with both random forest and covariance index faking detection method) could significantly reduce the proportion of fakers being included in the final selection group across job applicant sample sizes, SRs, and FHRs. Moreover, the application of identify and removal strategy with the random forest faking detection method achieved better performance in terms of the proportion of faker reduction in the selected group than the same strategy with the covariance index faking detection method. These finding suggested while the application of identify and removal strategy based on the two response pattern based faking detection methods can effectively reduce the proportion of fakers being selected when personality assessment is used, the random forest faking detection method can be more effective in terms of reducing the proportion of fakers being selected.

To answer the research question concerning whether the identify and removal strategy based on the response pattern based faking detection method can improve the group mean criterion score in the selected group, the findings did not support the use of such correction methods. When I investigated the correlation between predictor and the criterion in the job applicant sample, I found mixed results. The application of the identify and removal strategy based on the two response pattern based methods could only improve the validity coefficient between the emotional stability and the criterion but not the conscientiousness and the criterion.

This mixed result was due to two reasons: 1. faking has a different impact on the two correlations as FHR increases; 2. the identify and removal correction strategy has a different impact on the two correlations.

In the current dataset, the predictors and the criterion had significant correlations in the honest condition sample but not in the faking condition sample. Therefore, as the FHR increased in the simulation samples, it is equivalent to increasing the degree of random noise in the simulation samples. As a result, the correlation between the predictors and the criterion decreased as the fake-to-honest increased. However, faking did not impact the correlation between emotional stability and the correlation between conscientiousness equally. The absolute distance of the range of the estimated validity coefficients of conscientiousness and the criterion (0.138) is narrower than the absolute distance of the range of the estimated validity coefficients of emotional stability and the criterion (0.187) in the simulation samples with no correction across FHRs. In other words, faking has a stronger impact on the correlation of emotional stability and the criterion than the correlation between conscientiousness and the criterion. This effect can be observed by the slope of the line representing the validity coefficient of emotional stability and the criterion in the simulation samples with no correction in figure 3 is steeper than the slope of the line representing the validity coefficient of conscientiousness and the criterion in the simulation samples with no correction in figure 4. To the best of my knowledge, the current study is the first study that found empirical evidence suggesting the impact of the faking on the correlation coefficient between the personality predictors and the criterion is influenced by the true correlation coefficient between the personality predictors and the criterion. Morgeson, Campion, Dipboye, Hollenbeck, Murphy, & Schmitt (2007) discussed the reason that the impact of faking on criterion validity is low is that the true correlation coefficients between personality predictors and criteria are generally low as well. This finding partially supports their argument and further suggests the impact of faking on criterion related validity of personality assessment becomes stronger as the correlation coefficients between the personality predictors and the criteria become stronger.

Given faking has a stronger impact on the correlation coefficient between emotional stability and the criterion than the correlation coefficient between conscientiousness and the criterion, the correction strategy should have a stronger impact on the correlation between emotional stability and the criterion than the correlation between conscientiousness and the criterion. This is because when applying the identify and removal strategy, removing the fakers is equal to removing the random noise from the simulation samples. By removing the random noise, the decrease in the validity coefficient of the predictor and the criterion in the simulation samples corrected by either faking detection method was much smaller than the same validity coefficient in the simulation samples with no correction as the FHR increased. The absolute distance of the range of the estimated validity coefficients of conscientiousness and the criterion (0.05) in the simulation samples corrected by the random forest faking detection method across FHRs. The absolute distance of the range of the estimated validity coefficients of conscientiousness and the criterion (0.087) in the simulation samples corrected by the covariance index faking detection method across FHRs. The absolute distance of the range of the estimated validity coefficients of emotional stability and the criterion (0.06) in the simulation samples corrected by the random forest faking detection method across FHRs. The absolute distance of the range of the estimated validity coefficients of emotional stability and the criterion (0.092) in the simulation samples corrected by the covariance index faking detection method across FHRs. In other words, the changes in the validity coefficients of both conscientiousness and emotional stability with the criterion the simulation samples corrected by either faking detection method are smaller than the changes in the same validity coefficient in the simulation samples with no correction as FHRs increase. Therefore, the slopes of the lines representing the validity coefficient of emotional stability and the criterion in the simulation samples corrected by either faking detection method are steeper than the slope of the line representing the same validity coefficient in the simulation samples with no correction in figure 3. The slopes of the lines representing the validity coefficient of conscientiousness and the criterion in the simulation samples corrected by either faking detection method are steeper than the slope of the line representing the same validity coefficient in the simulation samples with no correction in figure 4.

Therefore, when the impact of faking on the validity coefficient is strong, the line representing the validity coefficient in the simulation sample with no correction crosses the lines representing the validity coefficients in the simulation samples corrected by either faking detection methods with low FHR (Figure 3). On the other hand, when the impact of faking on the validity coefficient is weak, the line representing the validity coefficient in the simulation sample with no correction only crosses the line representing the validity coefficient in the simulation samples corrected by the random forest faking detection methods (the higher accuracy detection method) with high FHR (Figure 4). This data pattern suggested, whether the faking detection method can effectively improve the validity coefficient is determined by the impact of faking has on the validity coefficient.

To answer the research question concerning the group mean criterion score in the selected group of the identify and removal strategy from the lens of the group mean criterion score in the selective group, I found the application of the identify and removal strategy had a detrimental effect on the group mean criterion score. Across all conditions, the application of the identify and removal strategy with either faking detection method always resulted in higher turnover intention in the selected groups. This is due to the positive correlations between the classification labels which were created by the two faking detection methods, and the predictors in the honest condition. The correlation between the classification labels created by the covariance index faking detection method and conscientiousness was 0.11, and the correlation between the same classification label and emotional stability was 0.25. The correlation between the classification labels created by the random forest faking detection method and conscientiousness was 0.47 and the correlation between the same classification label and emotional stability was 0.5. These positive correlations in the honest condition sample suggested participants who score high in these predictors were more likely to be misclassified as fakers and then taken out of the sample in the simulation process. The mean score of emotional stability and conscientiousness for the participants who were misclassified as fakers in the honest condition by the covariance index faking detection method were 3.25 and 3.28 and the same mean scores for the participants who were misclassified as fakers in the honest condition by the random forest faking detection method were 3.57 and 3.68. The mean scores of conscientiousness and emotional stability for the participants who were misclassified as fakers were significantly higher than the same mean score in the honest condition sample without any correction (emotional stability = 2.77, and conscientiousness = 3.12). Both predictors had significant negative correlations with the criterion in the honest condition sample but not in the faking condition sample. Even though the identify and removal strategy based on the two faking detection methods did remove a large proportion of fakers in the simulation sample, by misclassifying and removing a significant proportion of honest respondents who were supposed to be included in the selected group, the mean score of the criterion in the selected group would result in a higher mean score of turnover intention in the selected group.

In summary, the current study found the identify and removal strategy based on the response pattern based faking detection methods could successfully reduce a significant proportion of fakers being included in the selected group across all simulation conditions. This is the first time faking research found a correction strategy that could reduce the proportion of fakers being selected in personnel selection when using personality as a selection tool. In a prior study conducted by Ellingson et al. (1999), they used the suppression strategy along with the SDR scale score found the suppression strategy did not have any effect in terms of reducing the proportion of faker being included in the selected group. The main difference between the current study and the study conducted by Ellingson et al. was the correction strategy used. The suppression strategy assumes the faking is a suppressor and can be measured by the SDR score. Therefore, correcting for faking can be achieved by partialing out the variance SDR score in the correlation between the predictor and the criterion. The identify and removal strategy assume faking is a moderator and the validity coefficient of the predictor and the criterion for fakers is lower than honest respondents. Therefore, correcting for faking can be achieved by identifying and removing the fakers from the sample. Given the assumption of faking is different from the suppression correction strategy and the identify and removal strategy, the fact that only the identify and removal strategy is effective in reducing the proportion of fakers being included in the selected group provides evidence to support that we should consider faking as a moderator instead of suppressor for the trait score. On the other hand, I could not find any evidence to support the identify and removal strategy based on the two response pattern based faking detection methods that could effectively improve the group mean criterion score in the selected group. The positive correlation between the predictors and the classification labels created by the faking detection method resulted in removing a significant number of honest respondents with a high score in both predictors. Even though the response pattern based faking detection methods are more effective to classify fakers and honest respondents, they suffer a similar problem as the traditional SDR scale faking detection method that tends to misclassify honest respondents with high trait scores as fakers. Given these misclassified honest respondents should have higher criterion scores compare to the other job applicants when the trait scores are valid predictors of the criterion, practitioners need to be cautious when deploying such faking detection methods to classify fakers.

8. Limitations

There are three major limitations to the current study. First, the current study only investigated the identify and removal strategy with the response pattern based faking detection methods but did not include other faking detection methods such as the overclaiming scale. Even though the overclaiming scale does not classify honest respondent and fakers as effectively as the pattern base faking detection method, research suggests the overclaiming scale score has a non-significant correlation with all big five personality traits (Bensch, Paulhus, Stankov, Ziegler, 2019). Therefore, it might provide some advantage for the identify and removal strategy based on the overclaiming scale to improve the group mean criterion score in the selected group. Future studies should investigate the effectiveness of the identify and removal strategy with the faking detection methods that do not produce the result that correlates with the predictors.

Second, participants who were instructed to fake the personality assessment were only instructed to respond to the personality assessment as if they were applying for a desirable job. Past research suggested participants who were instructed to fake could tailor their response to the specific job description (Raymark & Tafero, 2009). In other words, job applicants who decide to engage in faking will also tailor their response to the personality assessment when they apply for a different job. Therefore, it is questionable whether the response pattern based faking detection methods develop based on the training sample in which the faking condition is simply instructed to fake good as if they were applying for a desirable job can still be effectively detect faking in the real high-stakes selection context. If these response pattern based faking detection methods can only detect faking in the selection context when the subject in the faking condition was instructed to fake as if they are applying for a specific job, it will significantly increase the cost of developing these faking detection methods. This is because a new training sample is needed for every new job being selected for these response pattern based faking detection methods to effectively detect faking. Future study is needed to address this generalizability question.

Third, I used turnover intention as the evaluation criterion to answer research question 1 and research question 2. As discussed in the method section, turnover intention was the only criterion that met the assumption of the identify and removal strategy in the current dataset. This assumption states the correlation between the personality predictor and the criterion for the honest respondents is significantly stronger than the fakers. However, turnover intention is not a common criterion to evaluate the effectiveness of a selection tool in a research setting. Instead, researchers and practitioners often use job performance as the evaluation criterion in both researches and practice (e.g. Boudreau, 1988; Cascio, 1996; Cascio & Silbey, 1979). Using job performance as the criterion to evaluate the effectiveness of the personality assessment in the selection context might not meet this assumption. Empirical research suggests job applicants who have high cognitive ability (e.g., Christiansen, Burns, & Montgomery, 2005; Vasilopoulos, Cucina, Dyomina, Morewitz, & Reilly, 2006) and have more knowledge of the job (e.g., Dwight & Alliger, 1997; Holden & Jackson, 1981) can inflate their personality trait score to a higher degree. Given both cognitive ability and knowledge of the job have positive correlations with job performance (Schmidt & Hunter, 1998; Schmidt, & Shaffer, 2016), even though there is a significant proportion of job applicants fake in the job applicant pool, the criterion related validity for the fakers might not be significantly weaker than the honest respondents. Therefore, further research is needed to determine whether these faking detection methods are effective to improve the effectiveness of the personality assessment when using job performance as the evaluation criterion.

9. Conclusion

In summary, I found the identify and removal strategy based on the two response pattern based faking detection method could effectively correct for faking when reducing the proportion of fakers being selected is the primary gold of the selection process, but could not effectively correct for faking when the primary goal is to improve the group mean criterion score in the selcted group. Therefore, practitioners need to carefully consider the trade-off between the proportion of fakers being selected and improve the group mean criterion score in the selected group when applying this method in a real-world selection context.

References

Ajzen, I. (1991). Theory of planned behavior. *Organizational behavior and human decision processes*, 50, 179-211.

Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of applied psychology*, *81*, 261-278.

Bensch, D., Paulhus, D. L., Stankov, L., & Ziegler, M. (2019). Teasing apart overclaiming, overconfidence, and socially desirable responding. *Assessment*, *26*(3), 351-363.

Bing, M. N., Kluemper, D., Davison, H. K., Taylor, S., & Novicevic, M. (2011). Overclaiming as a measure of faking. *Organizational Behavior and Human Decision Processes*, *116*(1), 148-162.

Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, *14*(4), 317-335.

Boudreau, J. W. (1988). *Utility analysis for decisions in human resource management*. Cornell University, Center for Advanced Human Resource Studies, School of Industrial and Labor Relations.

Bourdage, J. S., Lee, K., Lee, J.-H., & Shin, K.-H. (2012). Motives for organizational citizenship behavior: Personality correlates and coworker ratings of OCB. *Human Performance, 25*(3), 179–200. https://doi-org.ezproxy.lib.vt.edu/10.1080/08959285.2012.683904

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Brunetti, D. G., Schlottmann, R. S., Scott, A. B., & Hollrah, J. L. (1998). Instructed faking and MMPI-2 response latencies: The potential for assessing response validity. *Journal of Clinical Psychology*, *54*(2), 143-153.

Burns, G. & Christiansen, N.D. (2006). Sensitive or senseless: On the use of social desirability measures in selection and assessment. In R. Griffith (Ed.) A Closer Examination of Applicant Faking Behavior. Greenwich, CT: Information Age Publishing.

Burns, G. N., & Christiansen, N. D. (2011). Methods of measuring faking behavior. *Human Performance*, *24*(4), 358-372.

Burns, G. N., & Christiansen, N. D. (2011). Methods of measuring faking behavior. *Human Performance*, *24*(4), 358-372.

Calanna, P., Lauriola, M., Saggino, A., Tommasi, M., & Furlan, S. (2020). Using a supervised machine learning algorithm for detecting faking good in a personality self-report. *International Journal of Selection and Assessment*, *28*(2), 176-185.

Caldwell-Andrews, A., Baer, R. A., & Berry, D. T. (2000). Effects of response sets on NEO-PI-R scores and their relations to external criteria. *Journal of Personality Assessment (3), 74*, 472-488.

Camara, W. J., & Schneider, D. L. (1994). Integrity tests: Facts and unresolved issues. *American Psychologist, 49,* 112–119.

CASCIO, W. F. (1996). The role of utility analysis in the strategic management of organizations. *Journal of Human Resource Costing & Accounting*.

Cascio, W. E, Outtz, J., Zedeck. S., & Goldstein, I. L. (1991-this issue). Statistical implica- tions of six methods of test score use in personnel selection. Human Performance, 4, 233- 264.

Cascio, W. F., & Silbey, V. (1979). Utility of the assessment center as a selection device. *Journal of Applied Psychology*, *64*(2), 107.

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).

Chen, X. P., Hui, C., & Sego, D. J. (1998). The role of organizational citizenship behavior in turnover: Conceptualization and preliminary tests of key hypotheses. *Journal of applied psychology*, *83*(6), 922-931.

Christiansen, N. D., Burns, G. N., & Montgomery, G. E. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Human Performance*, *18*(3), 267-307.

Christiansen, N. D., Robie, C., Burns, G. N., & Speer, A. B. (2017). Using item-level covariance to detect response distortion on personality measures. *Human Performance*, *30*(2-3), 116-134.

Cofer, C. N., Chance, J., & Judson, A. J. (1949). A study of malingering on the Minnesota Multiphasic Personality Inventory. *The Journal of psychology*, *27*, 491-499.

Cunningham, M. R., Wong, D. T., & Barbee, A. P. (1994). Self-presentation dynamics on overt integrity tests: Experimental studies of the Reid Report. *Journal of Applied Psychology*, *79*, 643-668.

Dalal, R. S. (2005). A Meta-Analysis of the Relationship Between Organizational Citizenship Behavior and Counterproductive Work Behavior. *Journal of Applied Psychology*, *90*(6), 1241–1255. https://doi-org.ezproxy.lib.vt.edu/10.1037/0021-9010.90.6.1241

Damarin, F., & Messick, S. (1965). RESPONSE STYLES AS PERSONALITY VARIABLES: A THEORETICAL INTEGRATION OF MULTIVARIATE RESEARCH 1. *ETS Research Bulletin Series*, *1965*, i-116.

Dana, J., & Dawes, R. M. (2004). The superiority of simple alternatives to regression for social science predictions. *Journal of Educational and Behavioral Statistics*, *29*(3), 317-331.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34*(7), 571–582. https://doi-org.ezproxy.lib.vt.edu/10.1037/0003-066X.34.7.571

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin, 81*(2), 95–106. https://doi-org.ezproxy.lib.vt.edu/10.1037/h0037613

Donovan, J. J., Dwight, S. A., & Schneider, D. (2008). Faking in the real world: Evidence from a field study. In *annual conference of the Society of Industrial and Organizational Psychology*. *San Francisco, CA*.

Donovan, J. J., Dwight, S. A., & Schneider, D. (2014). The impact of applicant faking on selection measures, hiring decisions, and employee performance. *Journal of Business and Psychology*, *29*(3), 479-493.

Douglas, E. F., McDaniel, M. A., & Snell, A. F. (1996, August). THE VALIDITY OF NON-COGNITIVE MEASURES DECAYS WHEN APPLICANTS FAKE. In *Academy of Management Proceedings* (Vol. 1996, No. 1, pp. 127-131). Briarcliff Manor, NY 10510: Academy of Management.

Dunlop, P. D., Bourdage, J. S., de Vries, R. E., McNeill, I. M., Jorritsma, K., Orchard, M., ... & Choe, W. K. (2020). Liar! Liar!(when stakes are higher): Understanding how the overclaiming technique can be used to measure faking in personnel selection. *Journal of Applied Psychology*, *105*(8), 784.

Dunn, T. G., Lushene, R. E., & O'Neil, H. F. (1972). Complete automation of the MMPI and a study of its response latencies. *Journal of Consulting and Clinical Psychology*, *39*(3), 381-387.

Dwight, S. A., & Alliger, G. M. (1997). Reactions to overt integrity test items. *Educational and Psychological Measurement*, *57*, 937-948.

Edwards, A. L. (1957). The social desirability variable in personality assessment and research.

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

Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied psychology*, *84*(2), 155.

Feeney, J. R., & Goffin, R. D. (2015). The overclaiming questionnaire: A good way to measure faking? *Personality and Individual Differences*, *82*, 248-252.

Fekken, G. C., & Holden, R. R. (1992). Response latency evidence for viewing personality traits as schema indicators. *Journal of Research in Personality*, *26*, 103-120.

Fine, S., & Pirak, M. (2016). Faking fast and slow: Within-person response time latencies for measuring faking in personnel testing. *Journal of Business and Psychology*, *31*(1), 51-64.

Galić, Z., & Jerneić, Ž. (2013). Measuring faking on Five-Factor personality questionnaires: The usefulness of Communal and Agentic Management Scales. *Journal of Personnel*

Psychology, *12*(3), 115–123. https://doi-org.ezproxy.lib.vt.edu/10.1027/1866-5888/a000087

Ganster, D. C., Hennessey, H. W., & Luthans, F. (1983). Social desirability response effects: Three alternative models. *Academy of Management Journal*, *26(2)*, 321-331.

Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media.

Goffin, R. D., & Christiansen, N. D. (2003). Correcting personality tests for faking: A review of popular personality tests and an initial survey of researchers. *International Journal of Selection and assessment*, *11*(4), 340-344.

Goldberg, D. P. (1972). *The detection of psychiatric illness by questionnaire: A technique for the identification and assessment of non-psychotic psychiatric illness.* Oxford U. Press.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological assessment*, *4*(1), 26-61.

Gopalan, P., Nisan, N., Servedio, R. A., Talwar, K., & Wigderson, A. (2016). Smooth Boolean Functions are Easy: Efficient Algorithms for Low-Sensitivity Functions. *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science - ITCS '16*, 59–70. https://doi.org/10.1145/2840728.2840738

Griffith, R. L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review, 36*, 341-355.

Griffith, R. L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review*.

Hershcovis, M. S., Turner, N., Barling, J., Arnold, K. A., Dupré, K. E., Inness, M., ... & Sivanathan, N. (2007). Predicting workplace aggression: a meta-analysis. *Journal of applied Psychology*, *92*(1), 228.

Hogan, J., Hogan, R., & Murtha, T. (1992). Validation of a personality measure of managerial performance. *Journal of Business and Psychology*, *7*(2), 225-237.

Holden, R. R. (2007). Socially desirable responding does moderate personality scale validity both in experimental and in nonexperimental contexts. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *39*(3), 184.

Holden, R. R. (2008). Underestimating the effects of faking on the validity of self-report personality scales. *Personality and Individual Differences*, *44*(1), 311–321. https://doi-org.ezproxy.lib.vt.edu/10.1016/j.paid.2007.08.012

Holden, R. R., & Jackson, D. N. (1981). Subtlety, information, and faking effects in personality assessment. *Journal of Clinical Psychology, 37*, 379-386.

Holden, R. R., Kroner, D. G., Fekken, G. C., & Popham, S. M. (1992). A model of personality test item response dissimulation. *Journal of Personality and Social Psychology*, *63*, 272-296.

Hough, L. M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance, 11*, 209-244.

Hough, L. M., & Oswald, F. L. (2008). Personality testing and industrial–organizational psychology: Reflections, progress, and prospects. *Industrial and Organizational Psychology*, *1*(3), 272-290.

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those valididities. *Journal of Applied Psychology*, *75*(5), 581–595. https://doi-org.ezproxy.lib.vt.edu/10.1037/0021-9010.75.5.581

Hsu, L. M., Santelli, J., & Hsu, J. R. (1989). Faking detection validity and incremental validity of response latencies to MMPI subtle and obvious items. *Journal of Personality assessment*, *53*(2), 278-295.

Ilies, R., Fulmer, I. S., Spitzmuller, M., & Johnson, M. D. (2009). Personality and citizenship behavior: The mediating role of job satisfaction. *Journal of Applied Psychology*, *94*(4), 945–959. https://doi-org.ezproxy.lib.vt.edu/10.1037/a0013329

Jackson, D. N., & Messick, S. (1958). Content and style in personality assessment. *Psychological bulletin*, *55*(4), 243.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning with applications in R. New York, NY: Springer.

Jansen, A., KöNIG, C. J., Kleinmann, M., & Melchers, K. G. (2012). The Interactive Effect of Impression Motivation and Cognitive Schema on Self-Presentation in a Personality Inventory 1. *Journal of Applied Social Psychology*, *42*, 1932-1957.

Jensen, J. M., & Patel, P. C. (2011). Predicting counterproductive work behavior from the interaction of personality traits. *Personality and Individual Differences*, *51*(4), 466-471.

Japkowicz, N. (2000). The Class Imbalance Problem: Significance and Strategies. *In Proceedings of the 2000 International Conference on Artificial Intelligence (ICAI*, 111–117.

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity*, *47*(4), 2025-2047.

Kuhn, M., & Johnson, K. (2013). Applied predictive modeling. New York, NY: Springer.

Kuncel, N. R., & Borneman, M. J. (2007). Toward a new method of detecting deliberately faked personality tests: The use of idiosyncratic item responses. *International Journal of Selection and Assessment*, *15*(2), 220-231.

Kuncel, N.R. & Tellegen, A. (2001). The Social Desirability of Personality Items and Endorsed Trait Level: A reconceptualization of the measurement of item desirability. Poster Session Presented at the Annual Meeting of the Society for Industrial and Organizational Psychology, April, San Diego, CA.

Kusyszyn, I., & Jackson, D. N. (1968). A multimethod factor analytic appraisal of endorsement and judgment methods in personality assessment. *Educational and Psychological Measurement*, *28*(4), 1047-1061.

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological bulletin*, *107*, 34-56.

Leary, M. R., Robertson, R. B., Barnes, B. D., & Miller, R. S. (1986). Self-presentations of small group leaders: Effects of role requirements and leadership orientation. *Journal of Personality and Social Psychology*, *51*, 742-765.

McDaniel, M. A., & Timm, H. (1990, August). Lying takes time: Predicting deception in biodata using response latency. In *98th Annual Convention of the American Psychological Association, Boston*.

McDaniel, M. A., Douglas, E. F., & Snell, A. F. (1997, April). *A survey of deception among job seekers.* Paper presented at the Twelfth Annual Conference of the Society of Industrial and Organizational Psychology, St. Louis, MO.

McFarland, L. A., & Ryan, A. M. (2006). Toward an Integrated Model of Applicant Faking Behavior 1. *Journal of Applied Social Psychology*, *36*(4), 979-1016.

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel psychology*, *60*(3), 683-729.

Mueller-Hanson, R. A., Heggestad, E. D., & Thornton, G. C. III. (2006). Individual differences in impression management: An exploration of the psychological processes underlying faking. Psychology Science, 48(3), 288–312.

Murphy, A. H. (1996). The Finley affair: A signal event in the history of forecast verification. *Weather and Forecasting*, *11*(1), 3-20.

Napier, J. D. (1979). Effects of knowledge of cognitive-moral development and request to fake on Defining Issues Test P-scores. *The Journal of Psychology: Interdisciplinary and Applied*, *101*(1), 45–52. https://doi-org.ezproxy.lib.vt.edu/10.1080/00223980.1979.9915051

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81(6)*, 660-685.

Organ, D. W., & Lingl, A. (1995). Personality, satisfaction, and organizational citizenship behavior. *The Journal of Social Psychology, 135*(3), 339–350. https://doi-org.ezproxy.lib.vt.edu/10.1080/00224545.1995.9713963

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of personality and social psychology*, *46*(3), 598-612.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of social psychological attitudes, Vol. 1.*

Measures of personality and social psychological attitudes (p. 17–59). Academic Press. https://doi-org.ezproxy.lib.vt.edu/10.1016/B978-0-12-590241-0.50006-X

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of social psychological attitudes, Vol. 1. Measures of personality and social psychological attitudes* (p. 17–59). Academic Press. https://doi-org.ezproxy.lib.vt.edu/10.1016/B978-0-12-590241-0.50006-X

Paulhus, D. L., Harms, P. D., Bruce, M. N., & Lysy, D. C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of personality and social psychology*, *84 (4)*, 890-912.

Pelt, D. H., van der Linden, D., & Born, M. P. (2018). How emotional intelligence might get you the job: The relationship between trait emotional intelligence and faking on personality tests. *Human Performance, 31*, 33-54.

Peterson, M. H., Griffith, R. L., Isaacson, J. A., O'Connell, M. S., & Mangos, P. M. (2011). Applicant faking, social desirability, and the prediction of counterproductive work behaviors. *Human performance*, *24*(3), 270-290.

Peterson, M. H., Griffith, R. L., O'Connell, M. S., & Isaacson, J. A. (2008). Examining faking in real job applicants: A within-subjects investigation of score changes across applicant and research settings. In *RL Griffith & MH Peterson (Chairs), Examining faking using within-subjects designs and applicant data. Symposium conducted at the 23 rd Annual Conference for the Society for Industrial and Organizational Psychology: San Francisco, CA.*

Phillips, D. L., & Clancy, K. J. (1972). Some effects of" social desirability" in survey studies. *American Journal of Sociology*, *77*(5), 921-940.

Podsakoff, P. M., & MacKenzie, S. B. (1997). Impact of organizational citizenship behavior on organizational performance: A review and suggestion for future research. *Human performance*, *10*(2), 133-151.

Podsakoff, P. M., Ahearne, M., & MacKenzie, S. B. (1997). Organizational citizenship behavior and the quantity and quality of work group performance. *Journal of Applied Psychology*, *82*(2), 262–270. https://doi-org.ezproxy.lib.vt.edu/10.1037/0021-9010.82.2.262

Putka, D. J., Beatty, A. S., & Reeder, M. C. (2018). Modern prediction methods: New perspectives on a common problem. *Organizational Research Methods*, *21*(3), 689-732.

Raymark, P. H., & Tafero, T. L. (2009). Individual differences in the ability to fake on personality measures. *Human Performance*, *22*(1), 86-103.

Reeder, M. C., & Ryan, A. M. (2012). Methods for correcting for faking. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (p. 131–150). Oxford University Press.

Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, *83*(4), 634-663.

Rothstein, M. G., & Goffin, R. D. (2006). The use of personality measures in personnel selection: What does current research support? *Human resource management review*, *16*(2), 155-180.

Rotundo, M., & Sackett, P. R. (2002). The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: A policy-capturing approach. *Journal of Applied Psychology*, *87*(1), 66–80. https://doi-org.ezproxy.lib.vt.edu/10.1037/0021-9010.87.1.66

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

Sackeim, H. A., & Gur, R. C. (1978). Lateral asymmetry in intensity of emotional expression. *Neuropsychologia*, *16*(4), 473–481. https://doi-org.ezproxy.lib.vt.edu/10.1016/0028-3932(78)90070-2

Sackett, P. R., & Roth, L. (1991). A Monte Carlo examination of banding and rank order methods of test score use in personnel selection. *Human Performance, 4*(4), 279–295. https://doi-org.ezproxy.lib.vt.edu/10.1207/s15327043hup0404_3

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological bulletin*, *124*(2), 262-283.

Schmit, M. J., & Ryan, A. M. (1993). The Big Five in personnel selection: Factor structure in applicant and nonapplicant populations. *Journal of applied psychology*, *78*(6), 966-982.

Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology*, *80*(5), 607.

Schmitt, N., & Oswald, F. L. (2006). The impact of corrections for faking on the validity of noncognitive measures in selection settings. *Journal of Applied Psychology*, *91*(3), 613-621.

Simmering, L. (2012). Deception Detection: Using Eye-Tracking Technology to Measure Faking in a Simulated Applicant Setting.

Smith, D. B., Hanges, P. J., & Dickson, M. W. (2001). Personnel selection and the five-factor model: Reexamining the effects of applicant's frame of reference. *Journal of Applied Psychology*, *86*(2), 304.

Snell, A. F., Sydell, E. J., & Lueke, S. B. (1999). Towards a theory of applicant faking: Integrating studies of deception. *Human Resource Management Review*, *9*(2), 219-242.

Spector, P. E. (2011). The relationship of personality to counterproductive work behavior (CWB): An integration of perspectives. *Human Resource Management Review*, *21*(4), 342-352.

Spector, P. E., Bauer, J. A., & Fox, S. (2010). Measurement artifacts in the assessment of counterproductive work behavior and organizational citizenship behavior: Do we know what we think we know? *Journal of Applied Psychology*, *95*(4), 781–790. https://doi-org.ezproxy.lib.vt.edu/10.1037/a0019477

Spector, P. E., Fox, S., Penney, L. M., Bruursema, K., Goh, A., & Kessler, S. (2006). The dimensionality of counterproductivity: Are all counterproductive behaviors created equal? *Journal of Vocational Behavior*, *68*(3), 446–460. https://doi-org.ezproxy.lib.vt.edu/10.1016/j.jvb.2005.10.005

Topping, G. D., & O'Gorman, J. G. (1997). Effects of faking set on validity of the NEO-FFI. *Personality and Individual Differences*, *23*(1), 117–124. https://doi-org.ezproxy.lib.vt.edu/10.1016/S0191-8869(97)00006-8

Van Hooft, E. A., & Born, M. P. (2012). Intentional response distortion on personality tests: Using eye-tracking to understand response processes when faking. *Journal of Applied Psychology*, *97*(2), 301-316.

Vardi, Y., & Weitz, E. (2004). Misbehavior in organizations: *Theory, research, and management.* Mahwah, NJ: Erlbaum.

Vasilopoulos, N. L., Cucina, J. M., Dyomina, N. V., Morewitz, C. L., & Reilly, R. R. (2006). Forced-choice personality tests: A measure of personality and cognitive ability?. *Human Performance*, *19*(3), 175-199.

Vasilopoulos, N. L., Reilly, R. R., & Leaman, J. A. (2000). The influence of job familiarity and impression management on self-report measure scale scores and response latencies. *Journal of Applied Psychology*, *85*(1), 50-64.

Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and psychological measurement*, *59*(2), 197-210.

Viswesvaran, C., & Ones, D. S. (2000). Perspectives on models of job performance. *International Journal of Selection and Assessment*, *8*(4), 216–226. https://doi-org.ezproxy.lib.vt.edu/10.1111/1468-2389.00151

Walczyk, J. J., Roper, K. S., Seemann, E., & Humphrey, A. M. (2003). Cognitive mechanisms underlying lying to questions: Response time as a cue to deception. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *17*(7), 755-774.

Walczyk, J. J., Schwartz, J. P., Clifton, R., Adams, B., Wei, M. I. N., & Zha, P. (2005). Lying person-to-person about life events: A cognitive framework for lie detection. *Personnel Psychology*, *58*(1), 141-170.

Walz, S. & Niehoff, B. P. (1996). Organizational citizenship behaviors and their effect on organizational effectiveness in limited-menu restaurants, paper presented at Academy of Management Meetings, Cincinnati, OH.

Wiggins, J. S. (1959). Interrelationships among MMPI measures of dissimulation under standard and social desirability instruction. *Journal of Consulting Psychology*, *23*, 419-432.

Wiggins, J. S. (1959). Interrelationships among MMPI measures of dissimulation under standard and social desirability instruction. *Journal of Consulting Psychology*, *23*, 419-432.

Zickar, M., Rosse, J., Levin, R., & Hulin, C. (1996, April). Modeling the effects of faking on personality scales. In *11th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA*.

Zhang, B., Cao, M., Tay, L., Luo, J., & Drasgow, F. (2020). Examining the item response process to personality measures in high-stakes situations: Issues of measurement validity and predictive validity. *Personnel Psychology*, *73*(2), 305-332.

Ziegler, M. E., MacCann, C. E., & Roberts, R. D. (2012). Faking: Knowns, unknowns, and points of contention.

Ziegler, M., & Buehner, M. (2009). Modeling socially desirable responding and its effects. *Educational and Psychological Measurement*, *69*(4), 548-565.

Ziegler, M., Danay, E., Schölmerich, F., & Bühner, M. (2010). Predicting academic success with the Big 5 rated from different points of view: Self-rated, other rated and faked. *European Journal of Personality, 24*(4), 341–355.

Ziegler, M., MacCann, C., & Roberts, R. D. (Eds.). (2012). *Faking: Knowns, unknowns, and points of contention.* In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (p. 3–16). Oxford University Press.

Ziegler, M., Schmidt-Atzert, L., Buhner, M., & Krumm, S. (2007). Fakability of different measurement methods for achievement motivation: Questionnaire, semi-projective, and objective. *Psychology Science*, *49*(4), 291.

Zimmermann, S., Klusmann, D., & Hampe, W. (2017). Correcting the predictive validity of a selection test for the effect of indirect range restriction. *BMC medical education*, *17*(1), 1-10.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In *Advances in experimental social psychology* (Vol. 14, pp. 1-59). Academic Press.

Table 1

Comparisons of Big Five Trait Score Distribution Between Training Dataset and Testing Dataset

	Extroversion	Agreeableness	Neuroticism	Conscientiousness	Openness
Honest(train)	2.464 (2.411, 2.516)	3.052 (3.01, 3.094)	2.604 (2.553, 2.656)	2.833 (2.792, 2.874)	3.113 (3.073, 3.153)
Honest(test)	2.376 (2.3, 2.453)	3.089 (3.028, 3.15)	2.732 (2.653, 2.811)	3.084 (3.02, 3.148)	3.103 (3.047, 3.16)
Faking(train)	3.151 (3.103, 3.199)	3.465 (3.423, 3.508)	3.529 (3.487, 3.571)	3.678 (3.639, 3.717)	3.462 (3.425, 3.5)
Faking(test)	3.113 (3.055, 3.171)	3.478 (3.423, 3.533)	3.508 (3.449, 3.567)	3.659 (3.606, 3.711)	3.429 (3.377, 3.481)

Note. train = training dataset; test = testing dataset; Numbers not in parentheses represent the mean trait scores. Numbers in parentheses represent the 95% confidence intervals of the mean trait scores.

Table 2

Descriptive Statistics and Correlation for Precorrection Honest Sample in the Testing Dataset

	mean	SD	1	2	3
1. Conscientiousness	3.12	0.49	-	0.47**	-0.27**
2. Emotional Stability	2.78	0.65	-	-	-0.43**
3. Turnover Intention	2.86	0.93	-	-	-

Note. ** $p \leq 0.01$; * $p \leq 0.05$

Table 3

Descriptive Statistics and Correlation for Precorrection Faking Sample in the Testing Dataset

	mean	SD	1	2	3
1. Conscientiousness	3.65	0.38	-	0.77**	-0.02
2. Emotional Stability	3.5	0.43	-	-	-0.1
3. Turnover Intention	2.86	0.95	-	-	-

Note. ** $p \leq 0.01$; * $p \leq 0.05$

Table 4

Descriptive Statistics and Correlation for Honest Sample Corrected by Covariance Index Faking Detection Method

	mean	SD	1	2	3
1. Conscientiousness	3.1	0.44	-	0.35**	-0.18*
2. Emotional Stability	2.72	0.59	-	-	-0.38**
3. Turnover Intention	2.91	0.92	-	-	-

Note. ** $p \leq 0.01$; * $p \leq 0.05$

Table 5

Descriptive Statistics and Correlation for Faking Sample Corrected by Covariance Index Faking Detection Method

	mean	SD	1	2	3
1. Conscientiousness	3.85	0.18	-	0.39**	0.01
2. Emotional Stability	3.74	0.2	-	-	-0.07
3. Turnover Intention	2.79	1.04	-	-	-

Note. ** $p \leq 0.01$; * $p \leq 0.05$

Table 6

Descriptive Statistics and Correlation for Honest Sample Corrected by Random Forest Faking Detection Method

	mean	SD	1	2	3
1. Conscientiousness	3.03	0.45	-	0.31**	-0.19*
2. Emotional Stability	2.64	0.58	-	-	-0.39**
3. Turnover Intention	2.96	0.9	-	-	-

Note. ** $p \leq 0.01$; * $p \leq 0.05$

Table 7

Descriptive Statistics and Correlation for Faking Sample Corrected by Random Forest Faking Detection Method

	mean	SD	1	2	3
1. Conscientiousness	3.78	0.23	-	0.56**	0.03
2. Emotional Stability	3.64	0.28	-	-	-0.09
3. Turnover Intention	2.81	1	-	-	-

Note. ** $p \leq 0.01$; * $p \leq 0.05$

Table 8

Correlations Between Predictors and Criteria and 95% Confidence Intervals (N = 200)

FHRs	Con (No Correction)	Con (Covariance Index)	Con (Random Forest)	ES (No Correction)	ES (Covariance Index)	ES (Random Forest)
0	-0.277(-0.28,-0.274)	-0.188(-0.191,-0.185)	-0.196(-0.199,-0.192)	-0.439(-0.442,-0.437)	-0.388(-0.391,-0.385)	-0.392(-0.395,-0.389)
0.15	-0.226(-0.229,-0.222)	-0.167(-0.171,-0.163)	-0.185(-0.189,-0.181)	-0.365(-0.369,-0.362)	-0.36(-0.364,-0.357)	-0.376(-0.38,-0.373)
0.3	-0.187(-0.191,-0.183)	-0.14(-0.144,-0.135)	-0.173(-0.177,-0.168)	-0.315(-0.319,-0.311)	-0.336(-0.34,-0.332)	-0.369(-0.373,-0.365)
0.5	-0.139(-0.143,-0.135)	-0.101(-0.106,-0.096)	-0.146(-0.152,-0.14)	-0.252(-0.256,-0.249)	-0.296(-0.3,-0.291)	-0.34(-0.345,-0.335)

Note. 95% confidence interval are inside parenthesis; FHRs = fake-to-honest ratios; Con = conscientiousness; ES = Emotional stability; *No Correction = simulation samples with no correction; Covariance Index = simulation samples corrected by covariance index faking detection method; Random Forest = simulation samples corrected by random forest faking detection method.*

Table 9

Correlations Between Predictors and Criteria and 95% Confidence Intervals (N = 50)

FHRs	Con (No Correction)	Con (Covariance Index)	Con (Random Forest)	ES (No Correction)	ES (Covariance Index)	ES (Random Forest)
0	-0.27(-0.278,-0.262)	-0.182(-0.191,-0.173)	-0.187(-0.197,-0.177)	-0.434(-0.442,-0.427)	-0.383(-0.391,-0.375)	-0.386(-0.394,-0.378)
0.15	-0.233(-0.241,-0.224)	-0.174(-0.184,-0.164)	-0.187(-0.198,-0.177)	-0.369(-0.377,-0.361)	-0.36(-0.368,-0.351)	-0.375(-0.383,-0.366)
0.3	-0.194(-0.202,-0.185)	-0.151(-0.161,-0.141)	-0.179(-0.189,-0.168)	-0.319(-0.328,-0.31)	-0.337(-0.347,-0.327)	-0.361(-0.37,-0.351)
0.5	-0.146(-0.154,-0.138)	-0.112(-0.123,-0.101)	-0.156(-0.168,-0.143)	-0.26(-0.269,-0.251)	-0.307(-0.318,-0.297)	-0.347(-0.358,-0.335)

Note. 95% confidence interval are inside parenthesis; F-T-R = fake-to-honest ratios; Con = conscientiousness; ES = Emotional stability; *No Correction = simulation samples with no correction; Covariance Index = simulation samples corrected by covariance index faking detection method; Random Forest = simulation samples corrected by random forest faking detection method.*

Table 10

The proportion of Fakers Being Included in the Selected Groups (Selected Based on Emotional Stability) and 95% Confidence Intervals (N = 200)

SR	No Correction (15%)	Covariance Index (15%)	Random Forest (15%)	No Correction (30%)	Covariance Index (30%)	Random Forest (30%)	No Correction (50%)	Covariance Index (50%)	Random Forest (50%)
0.1	0.433 (0.427,0.438)	0.127 (0.123,0.131)	0.044 (0.041,0.046)	0.665 (0.659,0.67)	0.272 (0.267,0.278)	0.105 (0.101,0.109)	0.82 (0.816,0.825)	0.47 (0.464,0.476)	0.208 (0.204,0.213)
0.2	0.413 (0.41,0.417)	0.139 (0.136,0.142)	0.045 (0.043,0.047)	0.635 (0.632,0.639)	0.28 (0.276,0.284)	0.102 (0.1,0.105)	0.814 (0.811,0.818)	0.476 (0.472,0.48)	0.22 (0.216,0.223)
0.3	0.371 (0.369,0.373)	0.13 (0.128,0.132)	0.048 (0.046,0.049)	0.622 (0.619,0.625)	0.264 (0.261,0.267)	0.106 (0.104,0.109)	0.806 (0.803,0.809)	0.45 (0.447,0.453)	0.211 (0.208,0.213)
0.4	0.313 (0.312,0.315)	0.115 (0.113,0.116)	0.044 (0.043,0.045)	0.575 (0.573,0.577)	0.232 (0.23,0.234)	0.102 (0.101,0.104)	0.796 (0.794,0.798)	0.401 (0.399,0.404)	0.2 (0.198,0.202)
0.5	0.27 (0.269,0.271)	0.099 (0.098,0.101)	0.046 (0.044,0.047)	0.512 (0.511,0.514)	0.205 (0.203,0.207)	0.097 (0.096,0.099)	0.762 (0.76,0.763)	0.361 (0.358,0.363)	0.178 (0.176,0.18)

Note. fake-to-honest ratios are inside the parenthesis next to the column name; 95% confidence interval is inside the parenthesis next to the data; SR = selection ratios; *No Correction = simulation samples with no correction; Covariance Index = simulation samples corrected by covariance index faking detection method; Random Forest = simulation samples corrected by random forest faking detection method.*

Table 11

The proportion of Fakers Being Included in the Selected Groups (Selected Based on Conscientiousness) and 95% Confidence Intervals (N = 200)

SR	No Correction (15%)	Covariance Index (15%)	Random Forest (15%)	No Correction (30%)	Covariance Index (30%)	Random Forest (30%)	No Correction (50%)	Covariance Index (50%)	Random Forest (50%)
0.1	0.528 (0.522,0.533)	0.107 (0.103,0.111)	0.025 (0.023,0.027)	0.762 (0.757,0.767)	0.231 (0.226,0.236)	0.075 (0.072,0.078)	0.883 (0.879,0.888)	0.422 (0.416,0.428)	0.177 (0.173,0.182)
0.2	0.415 (0.411,0.418)	0.127 (0.124,0.13)	0.041 (0.039,0.042)	0.687 (0.683,0.691)	0.26 (0.256,0.263)	0.097 (0.095,0.1)	0.87 (0.868,0.873)	0.451 (0.447,0.455)	0.208 (0.204,0.211)
0.3	0.353 (0.351,0.356)	0.12 (0.117,0.122)	0.045 (0.043,0.046)	0.614 (0.612,0.617)	0.248 (0.246,0.251)	0.102 (0.1,0.104)	0.827 (0.824,0.829)	0.432 (0.429,0.435)	0.192 (0.19,0.195)
0.4	0.305 (0.304,0.307)	0.111 (0.109,0.113)	0.041 (0.04,0.042)	0.56 (0.558,0.563)	0.224 (0.222,0.226)	0.088 (0.086,0.089)	0.788 (0.786,0.79)	0.389 (0.387,0.392)	0.174 (0.172,0.176)
0.5	0.265 (0.264,0.266)	0.096 (0.094,0.097)	0.039 (0.038,0.04)	0.5 (0.499,0.502)	0.194 (0.192,0.196)	0.083 (0.081,0.085)	0.749 (0.747,0.751)	0.34 (0.337,0.342)	0.174 (0.172,0.176)

Note. fake-to-honest ratios are inside the parenthesis next to the column name; 95% confidence interval is inside the parenthesis next to the data; SR = selection ratios; *No Correction = simulation samples with no correction; Covariance Index = simulation samples corrected by covariance index faking detection method; Random Forest = simulation samples corrected by random forest faking detection method.*

Table 12

The proportion of Fakers Being Included in the Selected Groups (Selected Based on Emotional Stability) and 95% Confidence Intervals (N = 50)

SR	No Correction (15%)	Covariance Index (15%)	Random Forest (15%)	No Correction (30%)	Covariance Index (30%)	Random Forest (30%)	No Correction (50%)	Covariance Index (50%)	Random Forest (50%)
0.1	0.431 (0.419,0.443)	0.12 (0.111,0.129)	0.037 (0.032,0.042)	0.664 (0.652,0.676)	0.283 (0.272,0.295)	0.11 (0.101,0.118)	0.823 (0.813,0.833)	0.463 (0.45,0.477)	0.207 (0.196,0.218)
0.2	0.4 (0.393,0.407)	0.135 (0.129,0.141)	0.042 (0.038,0.045)	0.642 (0.634,0.651)	0.284 (0.276,0.292)	0.105 (0.099,0.111)	0.824 (0.817,0.832)	0.479 (0.471,0.488)	0.219 (0.212,0.227)
0.3	0.351 (0.347,0.355)	0.121 (0.117,0.126)	0.042 (0.039,0.045)	0.63 (0.624,0.636)	0.259 (0.253,0.265)	0.099 (0.095,0.104)	0.811 (0.805,0.816)	0.442 (0.435,0.449)	0.205 (0.199,0.211)
0.4	0.296 (0.293,0.299)	0.107 (0.103,0.11)	0.041 (0.038,0.043)	0.572 (0.568,0.576)	0.23 (0.225,0.235)	0.102 (0.098,0.106)	0.793 (0.789,0.798)	0.4 (0.394,0.406)	0.202 (0.197,0.207)
0.5	0.253 (0.251,0.255)	0.096 (0.093,0.099)	0.042 (0.039,0.044)	0.513 (0.51,0.516)	0.205 (0.201,0.209)	0.095 (0.091,0.098)	0.762 (0.759,0.766)	0.356 (0.351,0.361)	0.174 (0.17,0.178)

Note. fake-to-honest ratios are inside the parenthesis next to the column name; 95% confidence interval are inside the parenthesis next to the data; SR = selection ratios; Con = conscientiousness; ES = Emotional stability; *No Correction = simulation samples with no correction; Covariance Index = simulation samples corrected by covariance index faking detection method; Random Forest = simulation samples corrected by random forest faking detection method.*

Table 13

The proportion of Fakers Being Included in the Selected Groups (Selected Based on Conscientiousness) and 95% Confidence Intervals (N = 50)

SR	No Correction (15%)	Covariance Index (15%)	Random Forest (15%)	No Correction (30%)	Covariance Index (30%)	Random Forest (30%)	No Correction (50%)	Covariance Index (50%)	Random Forest (50%)
0.1	0.524 (0.512,0.537)	0.113 (0.104,0.121)	0.028 (0.023,0.033)	0.758 (0.747,0.769)	0.256 (0.244,0.268)	0.081 (0.073,0.088)	0.882 (0.874,0.89)	0.438 (0.425,0.451)	0.179 (0.169,0.19)
0.2	0.403 (0.396,0.41)	0.121 (0.115,0.127)	0.034 (0.031,0.038)	0.69 (0.682,0.698)	0.263 (0.255,0.271)	0.098 (0.092,0.103)	0.867 (0.861,0.873)	0.443 (0.434,0.452)	0.208 (0.201,0.215)
0.3	0.335 (0.33,0.34)	0.113 (0.108,0.117)	0.039 (0.036,0.042)	0.624 (0.618,0.629)	0.246 (0.24,0.252)	0.095 (0.09,0.099)	0.828 (0.823,0.834)	0.426 (0.419,0.433)	0.187 (0.181,0.193)
0.4	0.289 (0.286,0.292)	0.102 (0.098,0.105)	0.037 (0.034,0.039)	0.559 (0.554,0.563)	0.222 (0.217,0.227)	0.089 (0.085,0.093)	0.785 (0.781,0.79)	0.385 (0.379,0.391)	0.176 (0.171,0.181)
0.5	0.247 (0.245,0.249)	0.091 (0.088,0.094)	0.035 (0.033,0.037)	0.502 (0.498,0.505)	0.195 (0.191,0.199)	0.083 (0.08,0.086)	0.749 (0.745,0.753)	0.336 (0.331,0.341)	0.168 (0.164,0.172)

Note. fake-to-honest ratios are inside the parenthesis next to the column name; 95% confidence interval is inside the parenthesis next to the data; SR = selection ratios; Con = conscientiousness; ES = Emotional stability; *No Correction = simulation samples with no correction; Covariance Index = simulation samples corrected by covariance index faking detection method; Random Forest = simulation samples corrected by random forest faking detection method.*

Table 14

Mean Turnover Intention in the Selected Groups (Selected Based on Emotional Stability) and 95% Confidence Intervals (n = 200)

SR	No Correction (0%)	Covariance Index (0%)	Random Forest (0%)	No Correction (15%)	Covariance Index (15%)	Random Forest (15%)
0.1	2.252(2.243,2.262)	2.458(2.448,2.467)	2.497(2.489,2.505)	2.401(2.389,2.414)	2.522(2.512,2.533)	2.532(2.524,2.541)
0.2	2.33(2.324,2.336)	2.522(2.515,2.528)	2.548(2.542,2.553)	2.499(2.491,2.507)	2.567(2.56,2.574)	2.573(2.566,2.579)
0.3	2.429(2.424,2.434)	2.542(2.537,2.547)	2.584(2.579,2.588)	2.523(2.517,2.529)	2.591(2.585,2.596)	2.624(2.619,2.629)
0.4	2.471(2.467,2.475)	2.578(2.574,2.583)	2.641(2.637,2.645)	2.561(2.556,2.567)	2.63(2.626,2.634)	2.697(2.693,2.701)
0.5	2.519(2.515,2.522)	2.635(2.632,2.639)	2.707(2.703,2.71)	2.58(2.575,2.584)	2.685(2.681,2.689)	2.76(2.756,2.763)

Note. fake-to-honest ratios are inside the parenthesis next to the column name; 95% confidence interval is inside the parenthesis next to the data; SR = selection ratios; *No Correction = simulation samples with no correction; Covariance Index = simulation samples corrected by covariance index faking detection method; Random Forest = simulation samples corrected by random forest faking detection method.*

Table 14 (Continue)

Mean Turnover Intention in the Selected Groups (Selected Based on Emotional Stability) and 95% Confidence Intervals (n = 200)

SR	No Correction (30%)	Covariance Index (30%)	Random Forest (30%)	No Correction (50%)	Covariance Index (50%)	Random Forest (50%)
0.1	2.505(2.492,2.518)	2.587(2.576,2.598)	2.572(2.563,2.581)	2.563(2.548,2.578)	2.688(2.677,2.7)	2.643(2.633,2.653)
0.2	2.573(2.564,2.582)	2.611(2.603,2.618)	2.607(2.6,2.613)	2.632(2.622,2.641)	2.685(2.677,2.692)	2.701(2.695,2.708)
0.3	2.612(2.605,2.619)	2.645(2.639,2.651)	2.679(2.673,2.684)	2.659(2.652,2.666)	2.721(2.715,2.727)	2.772(2.767,2.777)
0.4	2.63(2.624,2.636)	2.682(2.678,2.687)	2.755(2.75,2.76)	2.708(2.701,2.714)	2.758(2.753,2.763)	2.844(2.838,2.849)
0.5	2.639(2.634,2.644)	2.735(2.731,2.74)	2.822(2.818,2.827)	2.707(2.702,2.712)	2.804(2.799,2.808)	2.949(2.944,2.954)

Note. fake-to-honest ratios are inside the parenthesis next to the column name; 95% confidence interval is inside the parenthesis next to the data; SR = selection ratios; *No Correction = simulation samples with no correction; Covariance Index = simulation samples corrected by covariance index faking detection method; Random Forest = simulation samples corrected by random forest faking detection method.*

Table 15

Mean Turnover Intention in the Selected Groups (Selected Based on Conscientiousness) and 95% Confidence Intervals (n = 200)

SR	No Correction (0%)	Covariance Index (0%)	Random Forest (0%)	No Correction (15%)	Covariance Index (15%)	Random Forest (15%)
0.1	2.369(2.359,2.379)	2.632(2.621,2.642)	2.807(2.796,2.818)	2.515(2.502,2.528)	2.709(2.698,2.72)	2.834(2.823,2.845)
0.2	2.511(2.503,2.519)	2.782(2.774,2.789)	2.866(2.859,2.873)	2.601(2.593,2.609)	2.818(2.81,2.825)	2.872(2.864,2.879)
0.3	2.645(2.639,2.65)	2.785(2.779,2.791)	2.836(2.83,2.841)	2.673(2.666,2.68)	2.805(2.799,2.811)	2.845(2.839,2.85)
0.4	2.67(2.665,2.675)	2.785(2.781,2.79)	2.832(2.827,2.836)	2.722(2.716,2.728)	2.807(2.802,2.812)	2.845(2.84,2.85)
0.5	2.683(2.679,2.687)	2.787(2.783,2.791)	2.845(2.841,2.85)	2.719(2.714,2.724)	2.805(2.8,2.809)	2.865(2.861,2.87)

Note. fake-to-honest ratios are inside the parenthesis next to the column name; 95% confidence interval is inside the parenthesis next to the data; SR = selection ratios; *No Correction = simulation samples with no correction; Covariance Index = simulation samples corrected by covariance index faking detection method; Random Forest = simulation samples corrected by random forest faking detection method.*

Table 15 (Continue)

Mean Turnover Intention in the Selected Groups (Selected Based on Conscientiousness) and 95% Confidence Intervals (n = 200)

SR	No Correction (30%)	Covariance Index (30%)	Random Forest (30%)	No Correction (50%)	Covariance Index (50%)	Random Forest (50%)
0.1	2.595(2.582,2.609)	2.781(2.77,2.793)	2.876(2.865,2.888)	2.679(2.665,2.693)	2.893(2.881,2.905)	2.947(2.936,2.959)
0.2	2.678(2.669,2.687)	2.862(2.854,2.87)	2.874(2.866,2.881)	2.722(2.712,2.731)	2.916(2.908,2.924)	2.902(2.894,2.91)
0.3	2.705(2.698,2.712)	2.83(2.824,2.836)	2.858(2.852,2.864)	2.757(2.75,2.764)	2.864(2.858,2.871)	2.879(2.873,2.885)
0.4	2.756(2.75,2.762)	2.827(2.822,2.833)	2.87(2.864,2.875)	2.789(2.782,2.795)	2.855(2.85,2.861)	2.908(2.903,2.914)
0.5	2.753(2.747,2.758)	2.827(2.823,2.832)	2.891(2.887,2.896)	2.803(2.797,2.808)	2.867(2.862,2.872)	2.959(2.954,2.964)

Note. fake-to-honest ratios are inside the parenthesis next to the column name; 95% confidence interval is inside the parenthesis next to the data; SR = selection ratios; *No Correction = simulation samples with no correction; Covariance Index = simulation samples corrected by covariance index faking detection method; Random Forest = simulation samples corrected by random forest faking detection method.*

Table 16

Mean Turnover Intention in the Selected Groups (Selected Based on Emotional Stability) and 95% Confidence Intervals (n =50)

	No Correction (0%)	**Covariance Index (0%)**	**Random Forest (0%)**	**No Correction (15%)**	**Covariance Index (15%)**	**Random Forest (15%)**
0.1	2.258(2.234,2.281)	2.459(2.435,2.483)	2.525(2.505,2.545)	2.418(2.39,2.446)	2.522(2.498,2.546)	2.536(2.515,2.557)
0.2	2.354(2.338,2.37)	2.521(2.505,2.536)	2.556(2.541,2.571)	2.493(2.474,2.512)	2.566(2.549,2.582)	2.584(2.57,2.599)
0.3	2.421(2.409,2.434)	2.545(2.533,2.558)	2.598(2.586,2.609)	2.53(2.515,2.544)	2.597(2.584,2.61)	2.636(2.624,2.647)
0.4	2.476(2.465,2.488)	2.586(2.575,2.597)	2.65(2.64,2.661)	2.549(2.537,2.56)	2.637(2.626,2.648)	2.703(2.692,2.713)
0.5	2.536(2.526,2.546)	2.646(2.636,2.656)	2.717(2.707,2.727)	2.585(2.575,2.596)	2.692(2.682,2.701)	2.763(2.754,2.773)

Note. fake-to-honest ratios are inside the parenthesis next to the column name; 95% confidence interval is inside the parenthesis next to the data; SR = selection ratios; *No Correction = simulation samples with no correction; Covariance Index = simulation samples corrected by covariance index faking detection method; Random Forest = simulation samples corrected by random forest faking detection method.*

Table 16(Continue)

Mean Turnover Intention in the Selected Groups (Selected Based on Emotional Stability) and 95% Confidence Intervals (n =50)

	No Correction (30%)	**Covariance Index (30%)**	**Random Forest (30%)**	**No Correction (50%)**	**Covariance Index (50%)**	**Random Forest (50%)**
0.1	2.508(2.479,2.538)	2.599(2.574,2.624)	2.593(2.571,2.614)	2.548(2.516,2.579)	2.672(2.647,2.696)	2.668(2.646,2.69)
0.2	2.566(2.546,2.586)	2.619(2.603,2.635)	2.618(2.603,2.633)	2.627(2.606,2.648)	2.694(2.677,2.711)	2.705(2.69,2.72)
0.3	2.602(2.586,2.618)	2.653(2.64,2.666)	2.689(2.676,2.701)	2.674(2.658,2.691)	2.723(2.71,2.736)	2.781(2.77,2.793)
0.4	2.637(2.624,2.65)	2.695(2.684,2.707)	2.761(2.75,2.772)	2.701(2.688,2.715)	2.766(2.755,2.777)	2.862(2.851,2.874)
0.5	2.634(2.623,2.644)	2.731(2.72,2.741)	2.827(2.817,2.838)	2.709(2.697,2.721)	2.817(2.807,2.828)	2.951(2.941,2.962)

Note. fake-to-honest ratios are inside the parenthesis next to the column name; 95% confidence interval is inside the parenthesis next to the data; SR = selection ratios; *No Correction = simulation samples with no correction; Covariance Index = simulation samples corrected by covariance index faking detection method; Random Forest = simulation samples corrected by random forest faking detection method.*

Table 17

Mean Turnover Intention in the Selected Groups (Selected Based on Conscientiousness) and 95% Confidence Intervals (n =50)

	No Correction (0%)	Covariance Index (0%)	Random Forest (0%)	No Correction (15%)	Covariance Index (15%)	Random Forest (15%)
0.1	2.365(2.34,2.39)	2.638(2.612,2.663)	2.793(2.767,2.819)	2.54(2.511,2.568)	2.71(2.683,2.737)	2.811(2.784,2.838)
0.2	2.538(2.519,2.557)	2.774(2.756,2.792)	2.866(2.848,2.884)	2.593(2.574,2.612)	2.801(2.782,2.819)	2.868(2.851,2.886)
0.3	2.63(2.615,2.645)	2.775(2.76,2.79)	2.842(2.827,2.857)	2.673(2.657,2.689)	2.807(2.792,2.821)	2.853(2.839,2.867)
0.4	2.665(2.652,2.677)	2.779(2.767,2.791)	2.833(2.821,2.845)	2.709(2.696,2.722)	2.807(2.795,2.82)	2.855(2.843,2.868)
0.5	2.691(2.68,2.702)	2.794(2.783,2.805)	2.855(2.845,2.866)	2.714(2.703,2.725)	2.811(2.8,2.821)	2.864(2.853,2.874)

Note. fake-to-honest ratios are inside the parenthesis next to the column name; 95% confidence interval is inside the parenthesis next to the data; SR = selection ratios; *No Correction = simulation samples with no correction; Covariance Index = simulation samples corrected by covariance index faking detection method; Random Forest = simulation samples corrected by random forest faking detection method.*

Table 17(Continue)

Mean Turnover Intention in the Selected Groups (Selected Based on Conscientiousness) and 95% Confidence Intervals (n =50)

	No Correction (30%)	**Covariance Index (30%)**	**Random Forest (30%)**	**No Correction (50%)**	**Covariance Index (50%)**	**Random Forest (50%)**
0.1	2.612(2.583,2.64)	2.815(2.789,2.841)	2.885(2.86,2.91)	2.662(2.632,2.692)	2.894(2.869,2.92)	2.934(2.909,2.959)
0.2	2.664(2.644,2.684)	2.827(2.809,2.845)	2.864(2.847,2.882)	2.7(2.68,2.72)	2.907(2.889,2.925)	2.892(2.875,2.909)
0.3	2.703(2.687,2.719)	2.835(2.821,2.85)	2.859(2.845,2.873)	2.765(2.749,2.781)	2.863(2.849,2.877)	2.88(2.866,2.894)
0.4	2.759(2.745,2.772)	2.833(2.82,2.845)	2.877(2.865,2.889)	2.789(2.776,2.803)	2.868(2.856,2.88)	2.917(2.905,2.929)
0.5	2.753(2.742,2.765)	2.828(2.817,2.839)	2.894(2.883,2.905)	2.796(2.784,2.808)	2.877(2.866,2.888)	2.965(2.955,2.975)

Note. fake-to-honest ratios are inside the parenthesis next to the column name; 95% confidence interval is inside the parenthesis next to the data; SR = selection ratios; *No Correction = simulation samples with no correction; Covariance Index = simulation samples corrected by covariance index faking detection method; Random Forest = simulation samples corrected by random forest faking detection method.*

Figure 1

Simulation Process Example 1

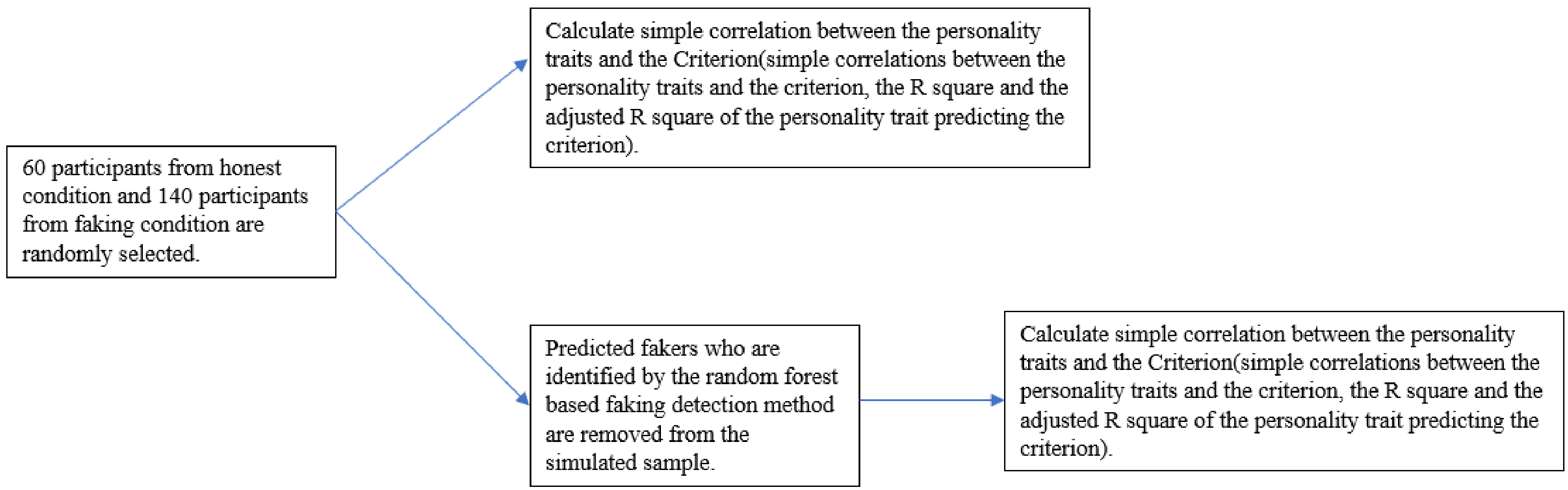


Figure 2

Simulation Process Example 2

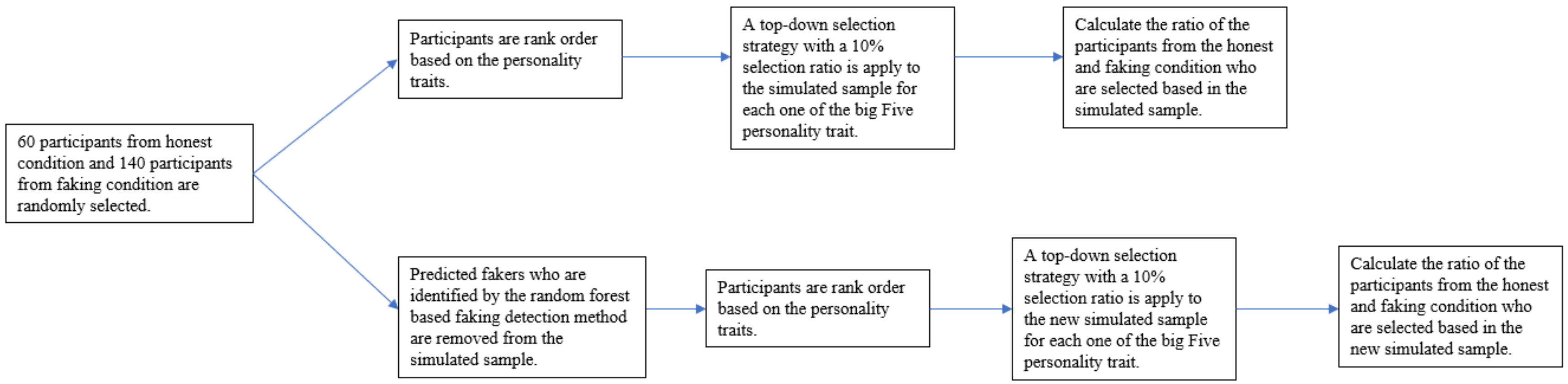
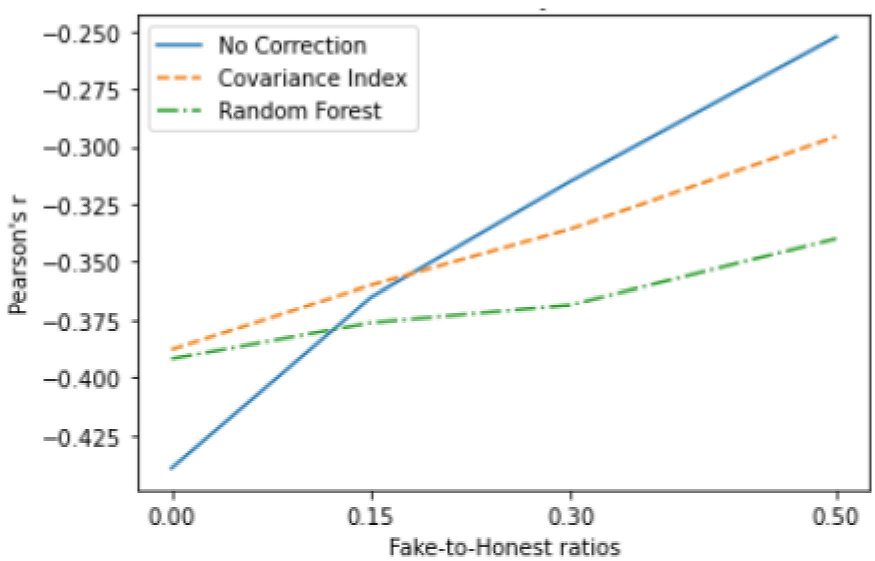


Figure 3

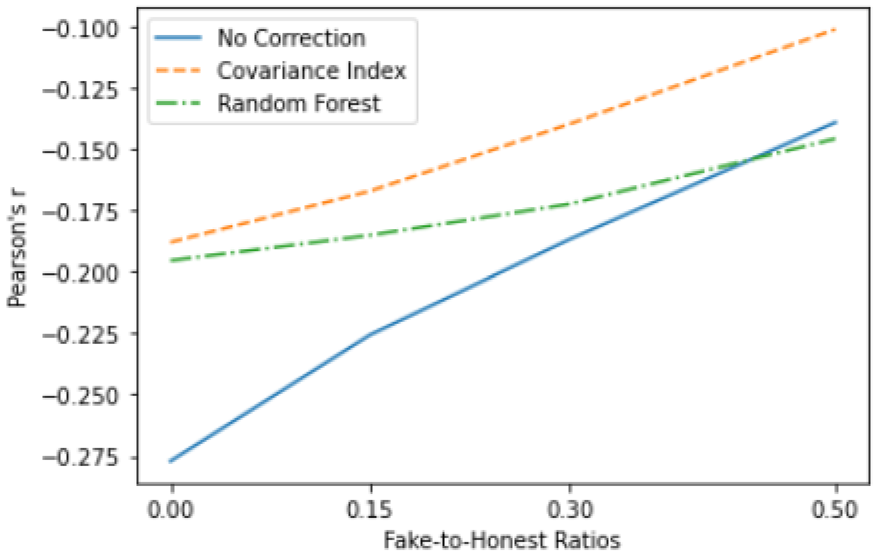
Correlation between Emotional Stability and Turnover Intention (n = 200)



Note. *No Correction = simulation samples with no correction; Covariance Index = simulation samples corrected by covariance index faking detection method; Random Forest = simulation samples corrected by random forest faking detection method.*

Figure 4

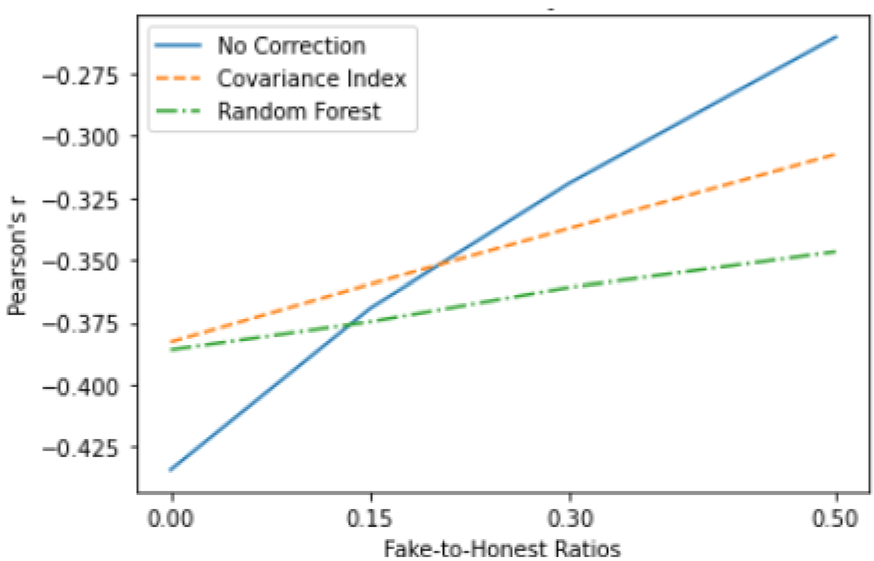
Correlation between Conscientiousness and Turnover Intention (n = 200)



Note. *No Correction = simulation samples with no correction; Covariance Index = simulation samples corrected by covariance index faking detection method; Random Forest = simulation samples corrected by random forest faking detection method.*

Figure 5

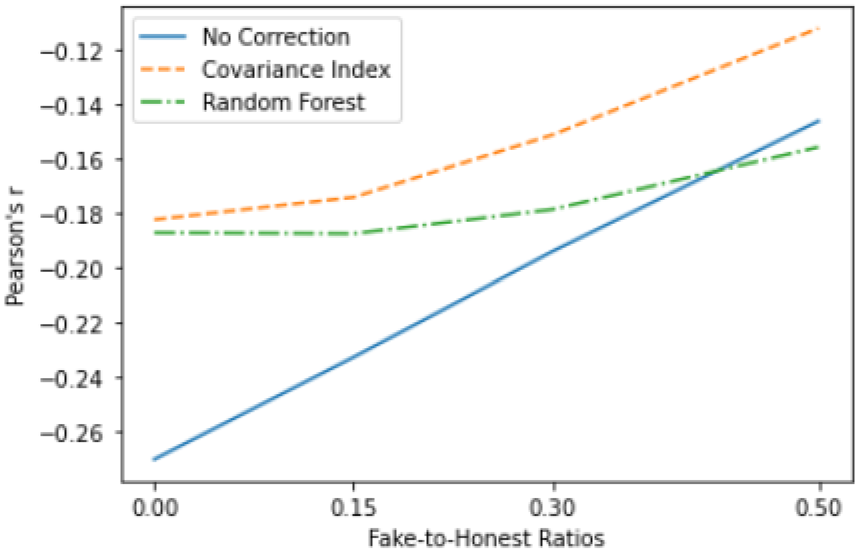
Correlation between Emotional Stability and Turnover Intention (n = 50)



Note. *No Correction = simulation samples with no correction; Covariance Index = simulation samples corrected by covariance index faking detection method; Random Forest = simulation samples corrected by random forest faking detection method.*

Figure 6

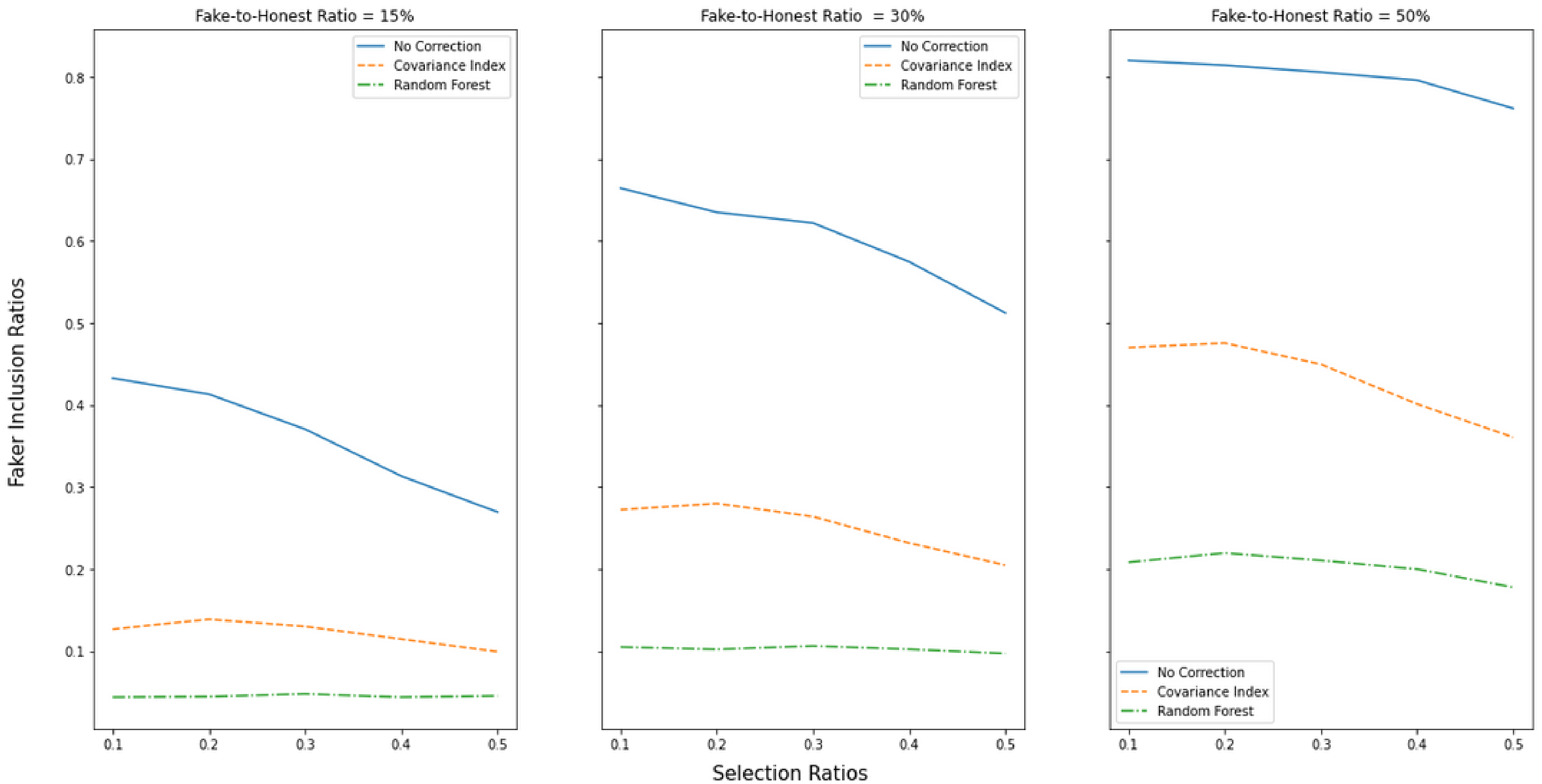
Correlation between Conscientiousness and Turnover Intention (n = 50)



Note. *No Correction = simulation samples with no correction; Covariance Index = simulation samples corrected by covariance index faking detection method; Random Forest = simulation samples corrected by random forest faking detection method.*

Figure 7

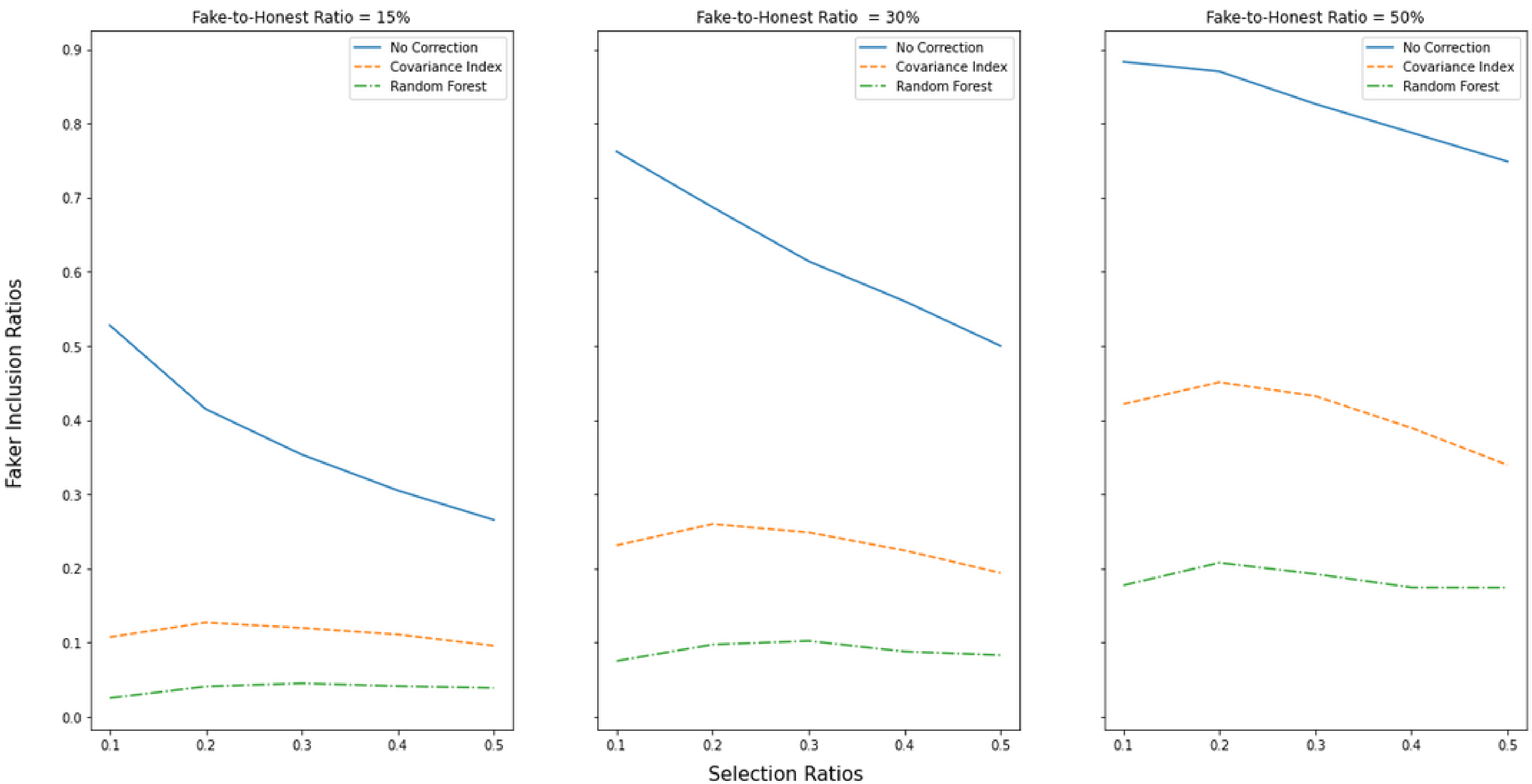
Proportion of Fakers being Included in the Selected Group Based on Emotional Stability Trait Score (n = 200)



Note. *No Correction = simulation samples with no correction; Covariance Index = simulation samples corrected by covariance index faking detection method; Random Forest = simulation samples corrected by random forest faking detection method.*

Figure 8

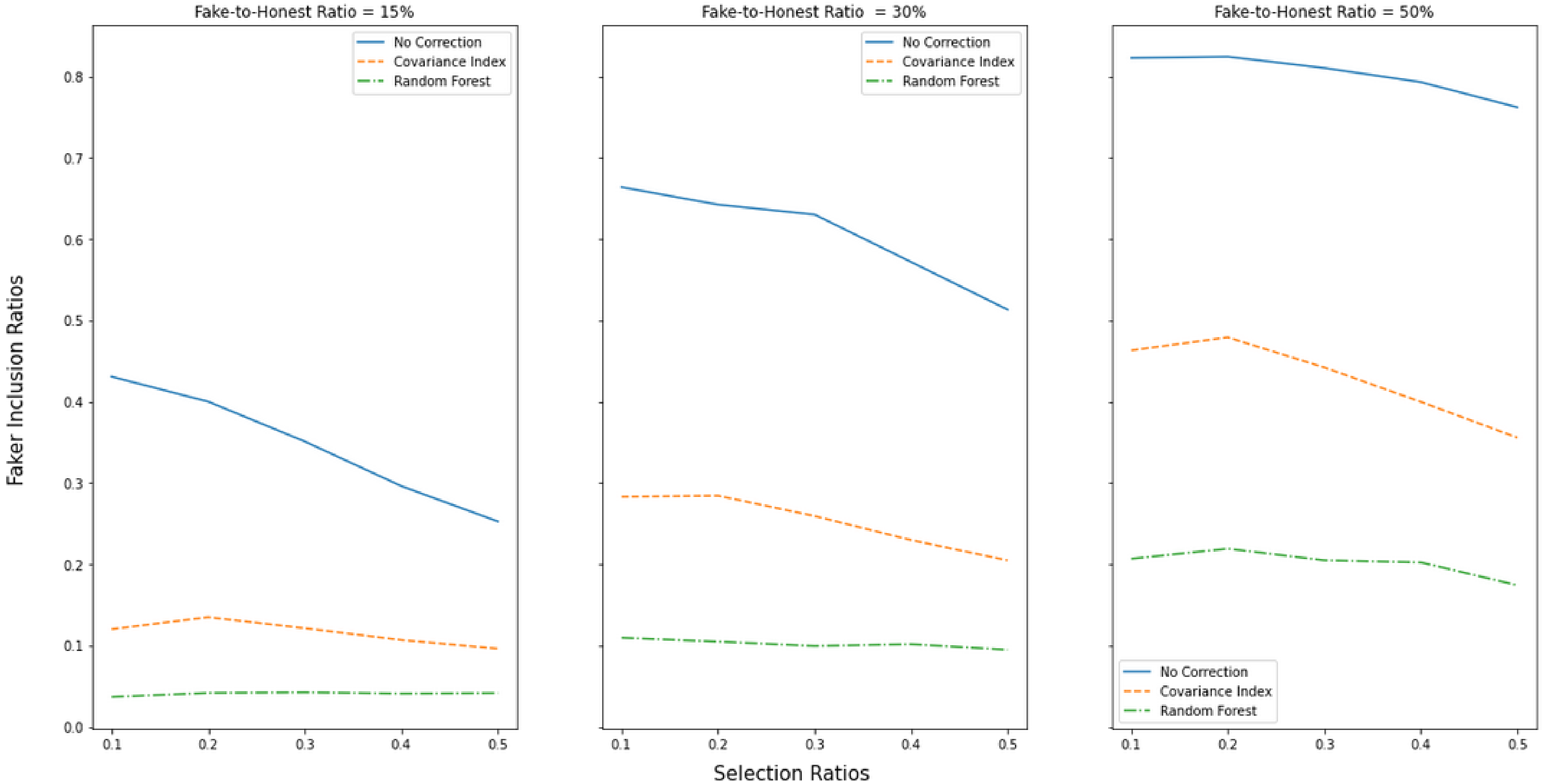
Proportion of Fakers being Included in the Selected Group Based on Conscientiousness Trait Score (n = 200)



Note. *No Correction = simulation samples with no correction; Covariance Index = simulation samples corrected by covariance index faking detection method; Random Forest = simulation samples corrected by random forest faking detection method.*

Figure 9

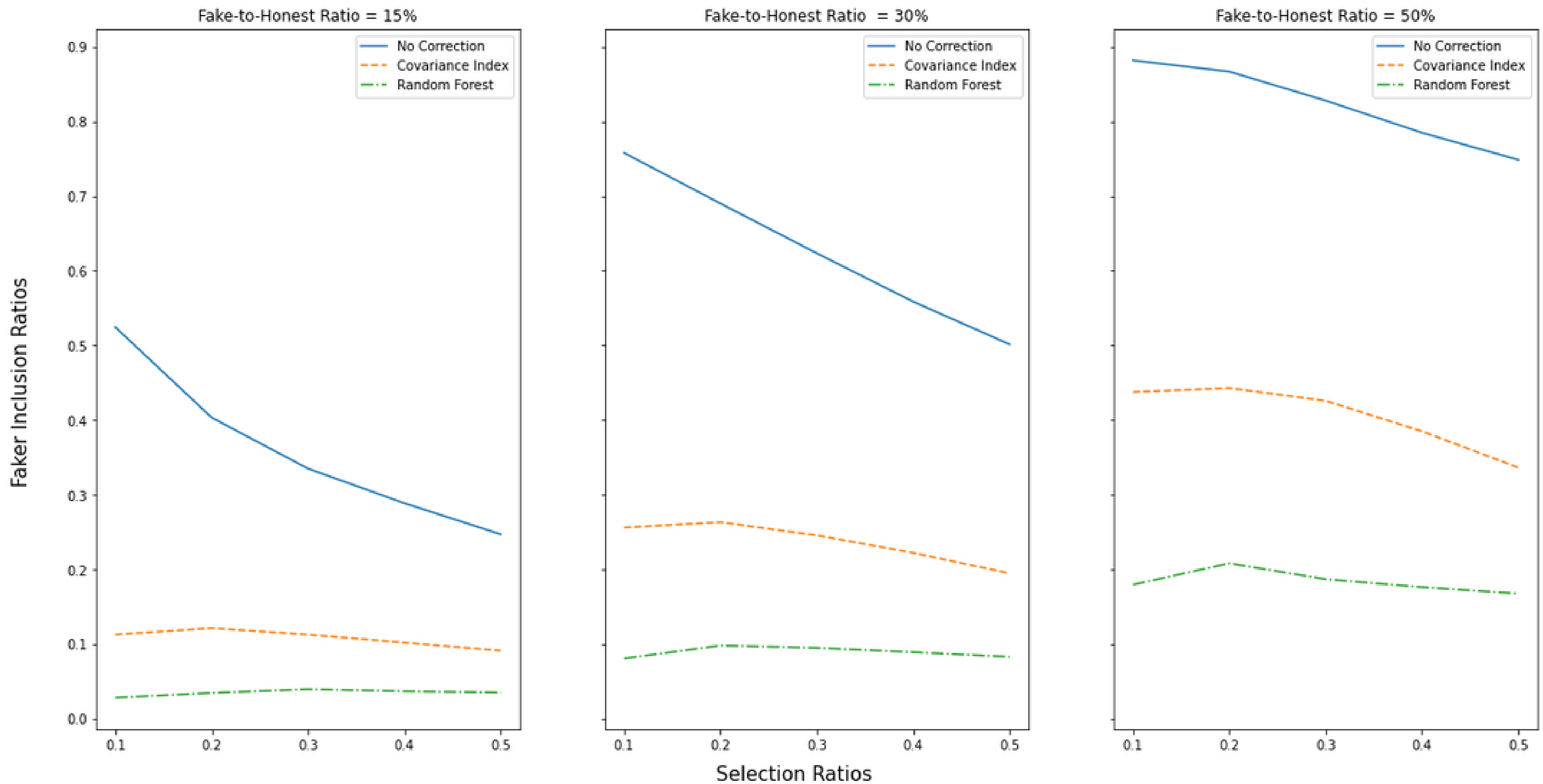
Proportion of Fakers being Included in the Selected Group Based on Emotional Stability Trait Score (n = 50)



Note. *No Correction = simulation samples with no correction; Covariance Index = simulation samples corrected by covariance index faking detection method; Random Forest = simulation samples corrected by random forest faking detection method.*

Figure 10

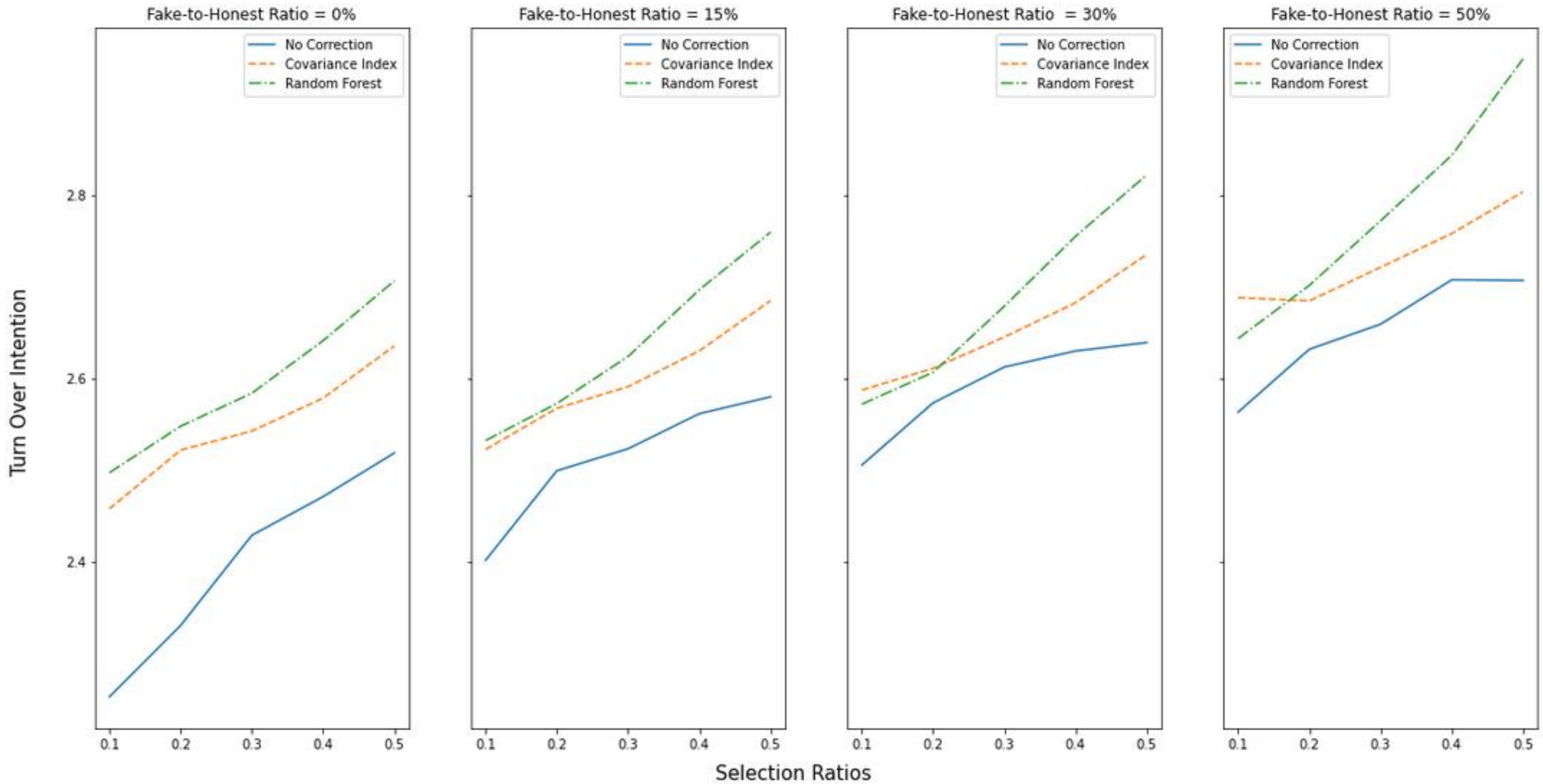
Proportion of Fakers being Included in the Selected Group Based on Conscientiousness Trait Score (n = 50)



Note. *No Correction = simulation samples with no correction; Covariance Index = simulation samples corrected by covariance index faking detection method; Random Forest = simulation samples corrected by random forest faking detection method.*

Figure 11

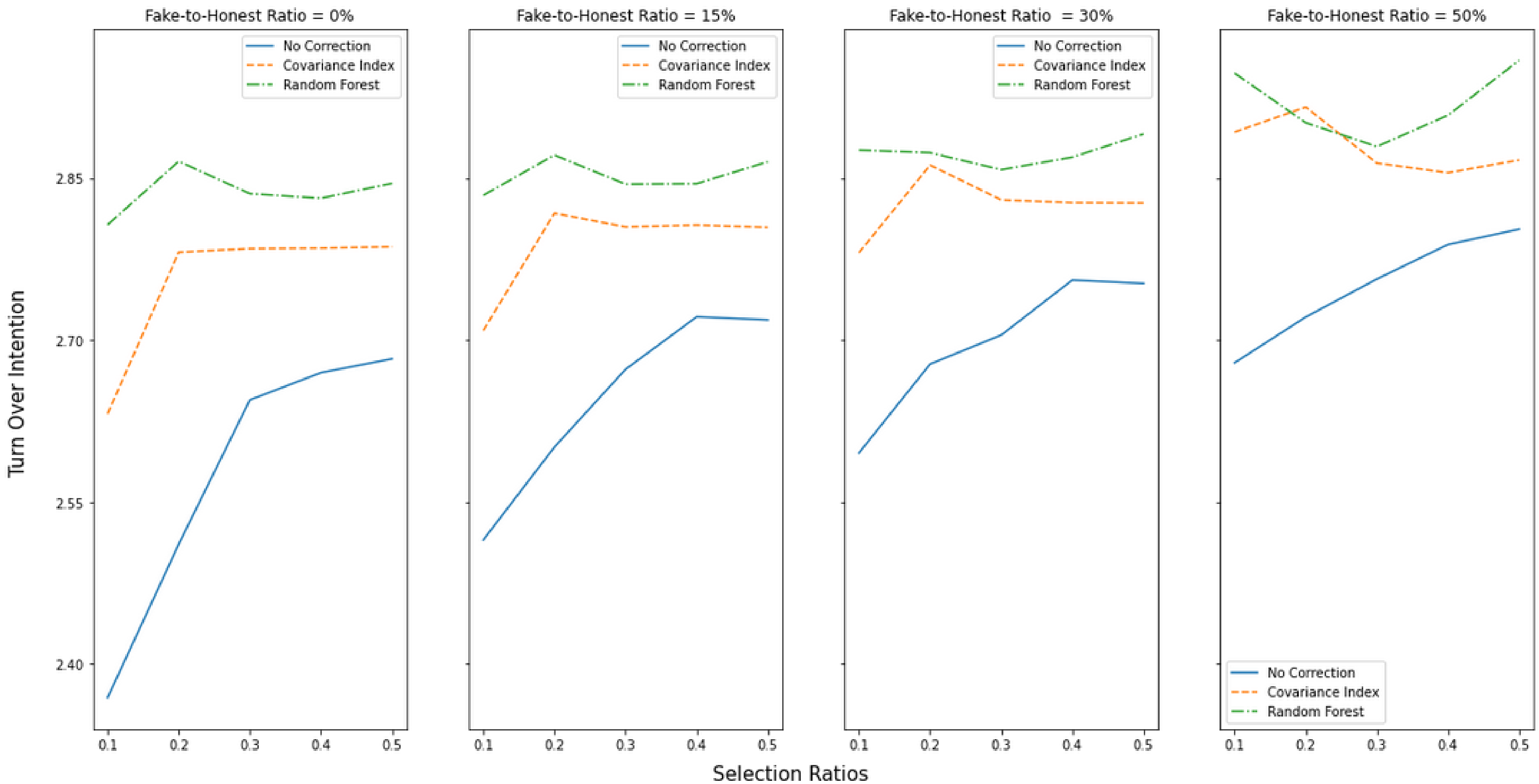
Mean Turnover Intention Score in the Selected Group Based on Emotional Stability Trait Score (n = 200)



Note. *No Correction = simulation samples with no correction; Covariance Index = simulation samples corrected by covariance index faking detection method; Random Forest = simulation samples corrected by random forest faking detection method.*

Figure 12

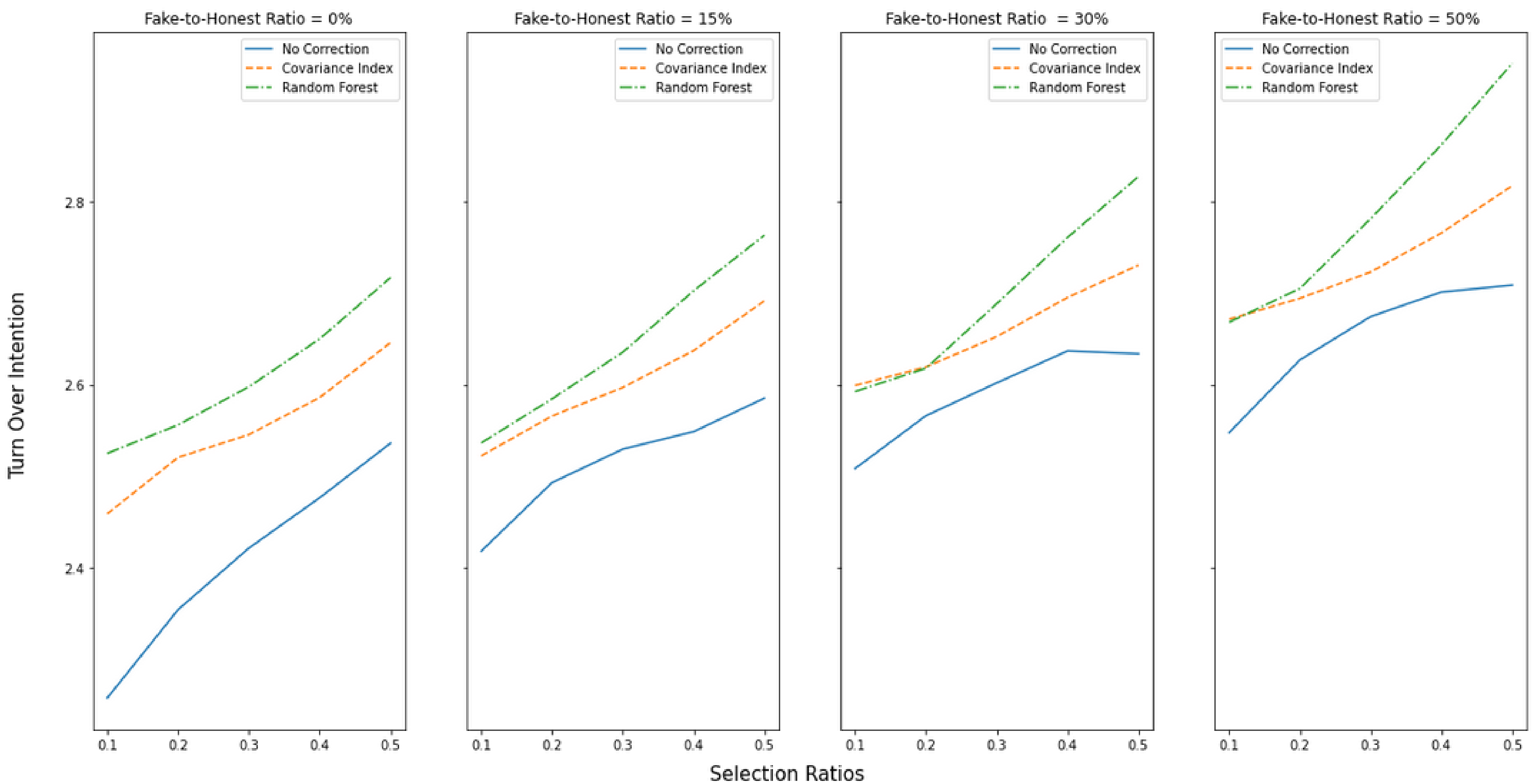
Mean Turnover Intention Score in the Selected Group Based on Conscientiousness Trait Score (n = 200)



Note. *No Correction = simulation samples with no correction; Covariance Index = simulation samples corrected by covariance index faking detection method; Random Forest = simulation samples corrected by random forest faking detection method.*

Figure 13

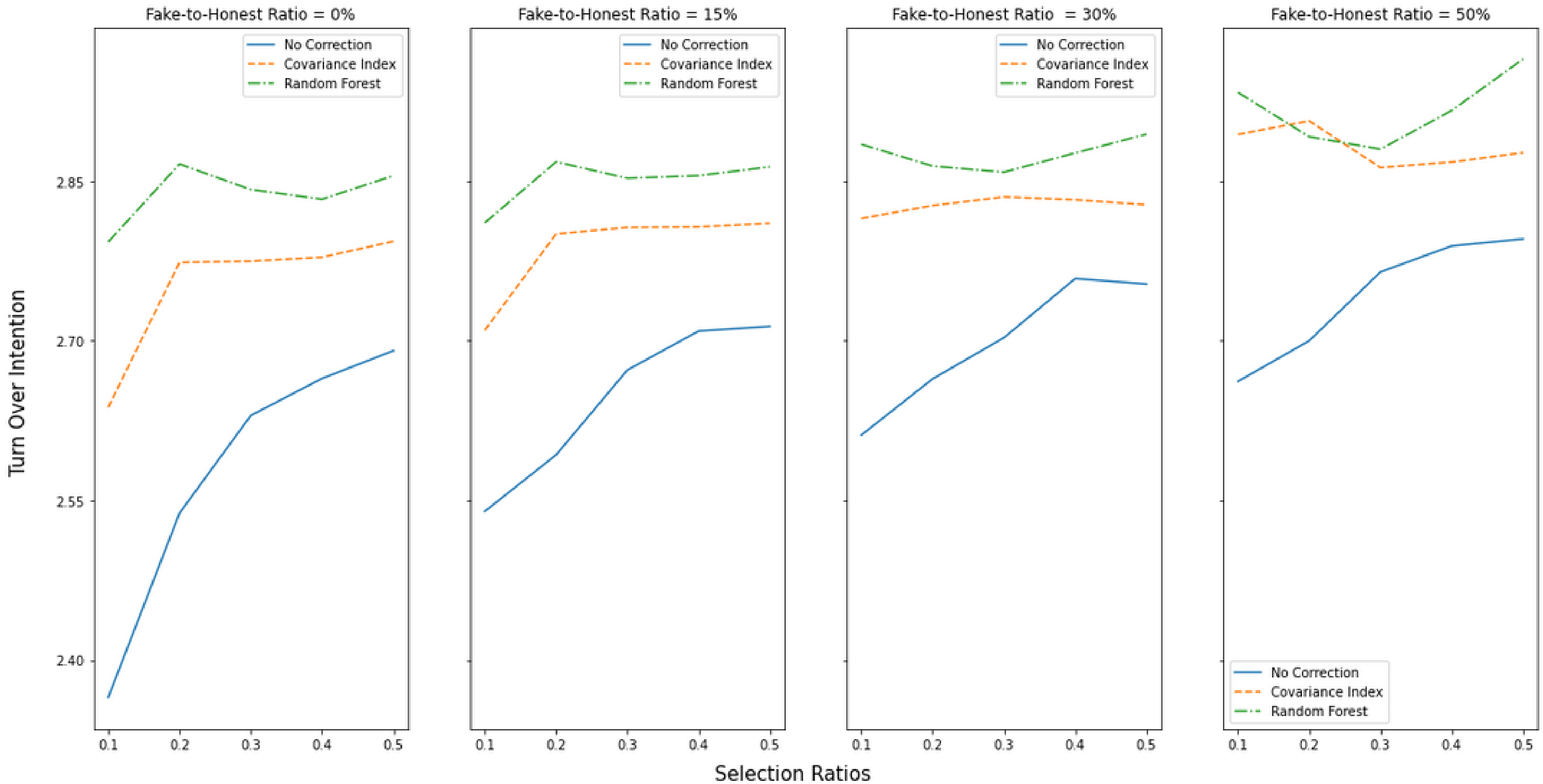
Mean Turnover Intention Score in the Selected Group Based on Emotional Stability Trait Score (n = 50)



Note. *No Correction = simulation samples with no correction; Covariance Index = simulation samples corrected by covariance index faking detection method; Random Forest = simulation samples corrected by random forest faking detection method.*

Figure 14

Mean Turnover Intention Score in the Selected Group that Selected Based on Conscientiousness Trait Score (n = 50)



Note. *No Correction = simulation samples with no correction; Covariance Index = simulation samples corrected by covariance index faking detection method; Random Forest = simulation samples corrected by random forest faking detection method.*