Received: 2 June 2023 | Revised: 11 May 2024 | Accepted: 13 May 2024

DOI: 10.1111/ijsa.12482

RESEARCH ARTICLE

International Journal of Selection and Assessment | WILEY

Effects of a constructed response retest strategy on faking, test perceptions, and criterion-related validity of situational judgment tests

Liyan Xi[1,2] | Qingxiong Weng[1] | Jan Corstjens[3] | Xiujuan Wang[1] | Lixin Chen[1]

[1]School of Management, University of Science and Technology of China, Hefei, China

[2]Department of Work, Organization, and Society, Ghent University, Ghent, Belgium

[3]Lee Kong Chian School of Business, Singapore Management University, Singapore, Singapore

Correspondence
Qingxiong Weng, School of Management, University of Science and Technology of China, Hefei, China.
Email: wqx886@ustc.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Number: 7190107003; China Scholarship Council

Abstract

This research proposes a faking-mitigation strategy for situational judgment tests (SJTs), referred to as the constructed response retest (CR-retest). The CR-retest strategy involves presenting SJT items in a constructed response format first, followed by equivalent closed-ended items with the same situation description. Two field experiments (N_1 = 733, N_2 = 273) were conducted to investigate the effects of this strategy and contrast it with a commonly used pretest warning message. Study 1 revealed that the CR-retest strategy was more effective than the warning message in reducing score inflation and improving criterion-related validity. Study 2 delved deeper by investigating the effects of the CR-retest strategy on applicant reactions in a 2 (with or without CR-retest strategy) × 2 (warning or control message) between-subjects design. The results showed that applicants reported positive fairness perceptions on SJT items with the CR-retest strategy. The CR-retest strategy was effective in reducing faking by evoking threat perceptions, whereas the warning message heightened threat and fear. Combining two strategies further decreased faking without undermining fairness perceptions. Overall, our results indicate that the CR-retest strategy could be a valuable method to mitigate faking in real-life selection settings.

KEYWORDS
constructed response retest, fairness perceptions, faking, personnel selection, situational judgment tests, warning message

Practitioner points

- The constructed response retest (CR-retest) strategy is a potentially effective faking-mitigation method for situational judgment tests (SJTs) in selection settings.
- The CR-retest strategy is more effective than a warning message in mitigating faking, while also promoting more favorable test perceptions.

We thank two anonymous reviewers for providing valuable feedback on this work.

Portions of this paper were presented at the 83rd Annual Meeting of the Academy of Management (AOM).

 © 2024 John Wiley & Sons Ltd.

- Combining the CR-retest strategy and a warning message yields additive effects in reducing faking without undermining test fairness perceptions.
- Organizations could incorporate the CR-retest strategy in their selection system, as the advantages may outweigh potential disadvantages, such as extended testing time.

Situational judgment tests (SJTs) usually present applicants with real-world challenges and subsequently ask them to select the best response (Motowidlo et al., 1990). SJTs' versatility in tapping into various job-related constructs, combined with their good criterion-related validity, relatively low subgroup differences, and positive applicant reactions, have made them popular selection tools (Christian et al., 2010; Kanning & Kuhne, 2006; Lievens & Motowidlo, 2016; Sackett et al., 2022). However, applicants could distort their responses, defined as faking (Griffith et al., 2006), which can lead to SJT score inflation (e.g., de Leng et al., 2019; Lievens, Peeters & Schollaert, 2008; Zhang et al., 2021). This issue is of significant concern because faking could bias selection outcomes (Lopez et al., 2019; Melchers et al., 2020) and lead to the inclusion of dishonest employees who may engage in counterproductive and deviant behaviors (O'Neill et al., 2013; Schneider & Goffin, 2012).

An effective selection tool should (1) address faking to enhance selection quality, (2) encourage applicants to self-disclose comfortably, and (3) ensure that applicants perceive the process as fair (de Leng et al., 2018; Lievens & Peeters, 2008; Lievens et al., 2021). Accordingly, prior research has proposed several faking-mitigation strategies (e.g., Oostrom et al., 2017). Unfortunately, these strategies struggle to strike a balance between robust performance (see Kuncel & Borneman, 2007) and applicants' fairness perceptions (Bill & Melchers, 2022; Burns et al., 2015; Fan et al., 2012; Li et al., 2022). Although SJTs are commonly used in practice, the mainstream faking-mitigation research has centered on personality tests. Kasten et al. (2020) highlighted differences in the explanatory models for faking in SJTs versus personality tests. This raises questions about the efficacy of applying faking-mitigation strategies developed for personality tests in SJTs. Consequently, there is a pressing need for refined strategies to address applicant faking in SJTs.

The constructed response (CR) strategy, which requires applicants to respond in an open-ended manner instead of selecting from the given response options (Lievens et al., 2019), served as the inspiration to conduct this study. SJTs with a CR format have comparable criterion-related validity with closed-ended SJTs (Edwards & Arthur, 2007; Lievens et al., 2019) and blend this validity with more chances for applicants to present themselves realistically and comprehensively. As a result, applicants have favorable reactions towards the CR format (De Soete et al., 2013; Lievens et al., 2019). However, the open-ended nature of this format entails high processing and scoring costs, limiting its practicality (Arthur et al., 2002; Edwards & Arthur, 2007). Our aim is to integrate the CR and closed-ended formats to develop a novel and mild faking-mitigation strategy for SJTs, termed the CR-retest strategy.

This study contributes to selection research in three ways. First, we comparatively evaluate the CR-retest strategy against a control group and a warning message (Rothstein & Goffin, 2006) in addressing faking and predicting selection outcomes. Notably, our focus is primarily on would-do SJTs with a rating format, given this format is particularly susceptible to faking (Arthur et al., 2014; Nguyen et al., 2005). Then, we investigate applicants' perceived fairness towards the CR-retest strategy and compare it with reactions toward the warning message. Lastly, drawing from cognitive appraisal theory, we explore the psychological mechanisms underlying the CR-retest strategy to investigate the potential cognitive and affective processes that drive faking-mitigation strategies.

1 | OVERVIEW OF PRIOR FAKING-MITIGATION STRATEGIES

McFarland and Ryan (2006) posited that faking hinges on applicants' intentions to fake and tests' susceptibility to faking. Two main types of faking-mitigation strategies have emerged. The first type, intention-reducing strategies such as warning messages, aims to curb applicants' desire to fake by clearly outlining the adverse outcomes of faking (e.g., Burns et al., 2015; Dwight & Donovan, 2003; Fan et al., 2012; Landers et al., 2011; Lopez et al., 2019). However, studies have indicated that such messages can compromise perceptions of test fairness (Burns et al., 2015; Converse et al., 2008; Fan et al., 2012; Li et al., 2022). The second type involves fakability-reducing strategies, aiming to limit fake opportunities by refining item properties. Tactics include using a forced-choice response format (Cao & Drasgow, 2019), imposing time constraints (Holden et al., 2001), asking what others would do (Oostrom et al., 2017), and requiring applicants to elaborate on their selected answers (Levashina et al., 2012; Lievens & Peeters, 2008). However, the effectiveness of forced-choice and time constraint strategies is not well-established in real-world selection (Heggestad et al., 2006; Komar et al., 2010; Kuncel & Borneman, 2007). Dalal et al. (2021) contended that forced-choice could threaten applicants' self-concepts and compromise their fairness perceptions.

Elaboration offers a markedly nuanced faking-mitigation strategy, prompting applicants to justify their answer choices in a constructed response format after answering closed-ended SJTs (Lievens & Peeters, 2008). The core logic is that retrospection promotes honest responses (Levashina et al., 2012). This method has effectively curtailed score inflation (Ramsay et al., 2006;

Schmitt & Kunce, 2002) and gained positive feedback from participants (Lievens & Peeters, 2008). However, a potential drawback is that exposure to closed-ended items may enable applicants to compare and evaluate response options and adopt new behaviors (Leeds, 2012, 2018). Consequently, their subsequent elaborations could be influenced by the response options they previously encountered. Hence, post-answer elaboration might not sufficiently deter applicants from faking in selection settings.

2 | CR-RETEST STRATEGY AS A NOVEL FAKING-MITIGATION STRATEGY

We draw inspiration from Lievens and Peeters (2008) and propose a new approach, prompting applicants to answer SJT items in a constructed response format *prior* to focusing on equivalent closed-ended SJT items. Prior studies have shown that applicants may engage in various cognitive processes when facing closed-ended SJTs, such as comparing options (Krumm et al., 2015), using implicit knowledge (Motowidlo & Beier, 2010), learning new behaviors (Leeds, 2012, 2018), and picking socially desirable answers (Kaminski et al., 2019; Schmitt et al., 2019), which could potentially lead to "faking good." This strategy contrasts with the traditional method of having applicants justify their choices post-selection from multiple-choice items. For the proposed model, applicants respond to SJTs in a constructed response format with no preset options, encouraging them to recall or introspectively revisit past behaviors (Edwards & Arthur, 2007). This approach limits applicants' inclination to tailor responses to provided choices, therefore curtailing the opportunity for faking. We termed this strategy the CR-retest strategy, leveraging the constructed response format to mitigate faking.

The CR-retest strategy is a retest because applicants respond to constructed response items first and thereafter to closed-ended SJT items with the same situation descriptions. The repeated items may make applicants self-aware of their previous responses (Desimone et al., 2015), implicitly signaling the risk of identifying and unfavorably evaluating inconsistent responses. Thus, we posit that this procedure decreases applicants' intention to fake. The CR-retest strategy incorporates the effectiveness of constructed response formats in measuring applicants' actual behavioral tendencies (Lievens et al., 2015; Lievens et al., 2019). At the same time, the strategy avoids the disadvantages of the constructed response format's high complexity and administrative costs (Arthur et al., 2002; Edwards & Arthur, 2007) by scoring only the responses of the closed-ended format.

3 | OVERVIEW OF STUDIES

The next sections present two experimental field studies. In Study 1, we tested the CR-retest strategy's effect on applicant faking and the SJT's criterion-related validity by comparing it with a prevalent faking mitigation strategy (i.e., a pre-test warning message) and with a control group. In Study 2, we examined the CR-retest strategy's effect on applicants' fairness perceptions, while also delving into the psychological mechanisms of this strategy.

4 | STUDY 1

In this study, we developed a two-phase CR-retest strategy. In the first phase, applicants were required to complete a portion of the SJT in a constructed response format. Constructed responses activate applicants' retrospective thought patterns (De Soete et al., 2013; Lievens et al., 2019), encouraging realistic and comprehensive self-presentation (Edwards & Arthur, 2007). This could potentially prevent applicants from distorting their responses when subsequently being presented with given options.

In the second phase, applicants were required to answer closed-ended SJT items with identical situation descriptions. The repeated items may hold applicants accountable for their responses, as highlighted by Desimone et al. (2015). Thus, we hypothesize that applicants would not change or optimize their answers based on the provided options in the closed-ended SJT items but would respond directly in line with the constructed response SJT items. We posit that adding the CR-retest strategy to SJT items would reduce applicant faking.

H1: Applicants will be faking less on closed-ended SJT items with the CR-retest strategy compared with closed-ended SJT items without the CR-retest strategy.

Criterion-related validity is decisive in evaluating selection tools' utility. Most studies support that faking undermines the criterion-related validity of selection tools (Buehl et al., 2019; de Leng et al., 2019; Lee et al., 2017; Peeters & Lievens, 2005; Peterson et al., 2011; van Hooft & Born, 2012; Wood et al., 2022). A longitudinal study by Donovan et al. (2014) found that fakers performed worse than non-fakers after entry in the organization, providing strong evidence for the potential damage of faking on validity. We expect that the CR-retest strategy could help applicants respond honestly by activating their retrospective thought patterns, possibly enhancing their responses' realism. This realism would enhance SJTs' criterion-related validity because it would closely resemble how they would perform on the job. Thus, the CR-retest strategy would benefit SJTs' criterion-related validity.

H2: Closed-ended SJT items with the CR-retest strategy will have a higher level of criterion-related validity than closed-ended SJT items without the CR-retest strategy.

We argue that the CR-retest strategy could reduce faking in SJTs and improve criterion-related validity, resulting in considerably realistic responses when presented with repeated situation description items. However, compared with warning messages that directly and explicitly tell applicants not to fake (Burns et al., 2015;

Dwight & Donovan, 2003), repeated items do not literally convey "do not fake" to applicants. Therefore, we are concerned that the CR-retest strategy may not be as influential as a warning message in deterring faking.

Conversely, although warning messages convey the negative consequences of faking, applicants may perceive warning messages as merely scaring them and posing no real threat (Kuncel & Borneman, 2007). Their credibility may thus be questioned because there is no identified "camera" during the test. By contrast, the CR-retest strategy, in which the repeated items may serve as deterrent against faking, could potentially be more effective in discouraging faking than a warning message.

To extend our understanding of faking-mitigation strategies applied in SJTs, we include a warning message, which is commonly used in personality tests, in our research design and propose the following research question:

RQ1: *Will the CR-retest strategy perform better than a warning message in reducing faking and increasing criterion-related validity in closed-ended SJTs?*

4.1 | Method

Study 1 aims to investigate the effectiveness of the CR-retest strategy in mitigating faking and optimizing criterion-related validity. To assess this strategy's incremental value, we compared it to a traditional faking-mitigation strategy (i.e., a warning message). Overall, we developed a threethree-factor-factor (i.e., CR-retest strategy, warning message, and control message groups) between-subjects field experimental design.

4.1.1 | Sample and procedure

The participants were 733 applicants competing for jobs at a large public hospital in Eastern China in 2022. These positions included nurses, technicians, and clinical staff, all requiring direct contact with patients. A total of 61% were female, average age was 24.67 years (SD = 3.32), and 50% held master's degrees.

4.1.2 | Field experiment

The recruitment included an online psychological assessment (i.e., a SJT and a personality test), an offline structured interview, and a medical knowledge test (see Figure 1). The psychological test contained an initial block and a main block. In the initial block, all applicants were required to complete part of the psychological test (i.e., five randomly selected closed-ended SJT items taken from the full SJT), which was used to estimate applicants' baseline SJT score in the high-stakes context.

In the main block, the applicants were randomly assigned to one of three groups (i.e., CR-retest strategy, warning message, and control group).[1] Those assigned to the *CR-retest strategy group* (N = 245) needed to complete seven constructed response SJT items, 20 closed-ended SJT items, and a personality test. The situation descriptions of the seven constructed response SJT items were identical to seven of the 20 closed-ended SJT items. All items were presented in random sequence, but the applicants were always shown the closed-ended SJT items with the same situational description after completing the corresponding constructed response SJT items. Applicants assigned to the *warning message group* (N = 246) received a warning message (Figure 2) before 20 closed-ended SJT items and the personality test. Lastly, applicants assigned to the *control group* (N = 242) received a control message (Figure 2) before 20 SJT items and the personality test. One day after the psychological test, the applicants attended an offline interview and completed medical knowledge tests. A total of 162 applicants were hired (22% of the total sample) after the selection procedure. We received their supervisor-rated job performance data from the HR department 6 months after they started their jobs.

The post hoc power analyses revealed that the applicant sample gave us a power of 99% to detect medium effects of mean differences between the CR-retest strategy and control groups (d = 0.50, Cohen, 1992), and a power of 91% to detect medium

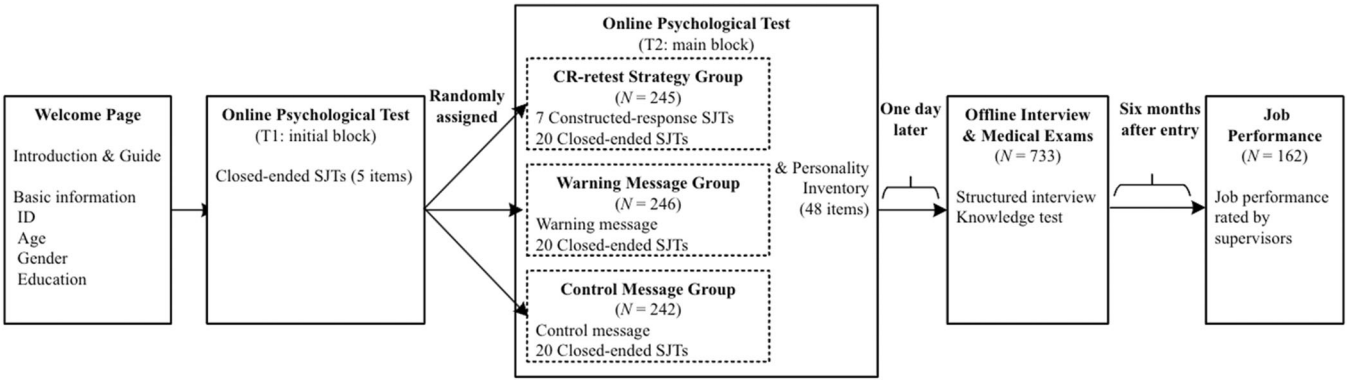

FIGURE 1 *The research design flow chart of Study 1.*

Warning Message

*感谢您参与**医院的招聘并参与此部分线上测试。该测试用于测量心理特质，您的回答并不涉及对错，所有的回答均为合理，我们期望您能在测试中呈现真实的自我。*

基于过去的测评数据，我们注意到一些申请人在回答时会作伪，即回答前后矛盾、存在明显的夸大甚至虚构经历，这极大影响了心理测量的真实性与准确性，严重破坏了测试的公平性。在此提醒，请您在回答时真实呈现自我，作伪的申请人会被识别出来并取消成绩。

*Thank you for participating in the recruitment process at *** Hospital and for taking this part of the online test. The test is designed to measure psychological traits and your answers are not right or wrong, all responses are reasonable, and we expect you to present yourself authentically in the test.*

Based on previous assessment data, we noticed that some applicants faked their answers, i.e., their answers had inconsistent, obvious exaggerations, or even fictitious experiences, which greatly impaired the accuracy and validity of the psychometric test and seriously undermined the fairness of the test. We would like to remind you to be true to yourself in your answers, as fakers will be identified, and their scores will be annulled.

Control Message

*感谢您参与**医院的招聘并参与此部分线上测试。该测试用于测量心理特质，您的回答并不涉及对错，所有的回答均为合理，我们期望您能在测试中呈现真实的自我。*

*Thank you for participating in the recruitment process at *** Hospital and for taking this part of the online test. The test is designed to measure psychological traits and your answers are not right or wrong, all responses are reasonable, and we expect you to present yourself authentically in the test.*

FIGURE 2 *Warning and control messages used in Study 1.*

effects of the correlation difference between two groups ($q = 0.30$, Cohen, 1992). The hired sample only gave a power of 33% to detect medium effects of correlation differences between two groups ($q = 0.30$, Cohen, 1992).

4.1.3 | Measures

SJT

The 25-item SJT was developed by a professional talent assessment company, which provides selection services for many public hospitals in China. The SJT items present applicants with various situations that can be encountered in the hospital workplace, such as care of patients, conflict management, interpersonal communication, teamwork, and integrity (see detailed information in the online Supporting Information S1). Each SJT item consisted of four response options with would-do response instructions, and applicants were asked to rate each response option from 1 (*least likely*) to 5 (*most likely*). McDonald's ω reliability of the initial block SJT items was 0.66, 0.59, and 0.60 for the three groups, respectively; and McDonald's ω of the main block SJT items was 0.89, 0.81, and 0.68, respectively.

Criterion measure

We followed Lievens et al. (2015) and used an internal validation measure (selection decision) and an external (overall job performance) validation measure provided by the hospital's HR department as criteria. Hiring decision (i.e., being hired vs. not being hired) was based on the participants' scores on the structured interview and knowledge test. Their SJT scores had no bearing on the selection decision.[2] Overall job performance was rated on a single-item scale (i.e., average, excellent, or outstanding) by the department head based on the first 6-month performance of the hired applicants.

Faking behavior

We utilized the faking literature (Burns & Christiansen, 2011; Feeney et al., 2023; Levashina et al., 2014) as basis for including the following two measurements of applicant faking.[3]

Regression-adjusted difference scores

Regression-adjusted difference scores (RADS) (Burns & Christiansen, 2011; Feeney et al., 2023) were derived by regressing the SJT scores from the main block onto scores from the initial block and computing the residuals. A high magnitude of differential scores implies an elevated level of faking.[4]

Blatant Extreme Responding

Blatant extreme responding (BER) (Feeney et al., 2023; Levashina et al., 2014) is calculated as the proportion of the number of times an applicant endorses the most favorable response (i.e., strongly agreed) compared with the least favorable response (i.e., strongly disagreed) in the total SJT rating because fakers typically endorse extreme responses (Levashina et al., 2014; Sun et al., 2022).

4.2 | Results

Table 1 shows the means, standard deviations, and correlations of this study. H1 stated that applicants will be faking less on SJT items with the CR-retest strategy compared with SJT items without the CR-retest strategy. With RADS as proxy of faking, applicants in the CR-retest strategy group ($M = -0.19$, SD = 0.88) exhibited lower faking than those in the control group ($M = 0.22$, SD = 0.51), $t(390) = -6.35$, $p < 0.001$, $d = -0.57$. With BER as proxy of faking, applicants in the CR-retest strategy group ($M = 0.18$, SD = 0.13) also showed lower faking than those in the control group ($M = 0.28$, SD = 0.15), $t(473) = -7.81$, $p < 0.001$, $d = -0.71$. These results supported that the CR-retest strategy was effective mitigating faking in SJT items. Hence, H1 was supported.

H2 proposed that SJT items with the CR-retest strategy will have higher criterion-related validity than SJT items without the CR-retest strategy. With selection decision as the criterion, the criterion-related validity of the CR-retest strategy group was 0.30, which was significantly higher compared with the control group ($r = 0.13$, *Fisher's* $z = 1.96$, $p < 0.05$, $q = 0.18$). For predicting job performance, the criterion-related validity of the CR-retest strategy group ($r = 0.27$) was three times as high as that of the control group ($r = 0.09$, *Fisher's* $z = 0.94$, $p = 0.35$, $q = 0.20$). Although the difference was not statistically significant due to the low statistical power of 33%, it had a modest effect size (Cohen, 1992). H2 was thus supported.

We next examined RQ1, which concerns whether the CR-retest strategy is more effective than a warning message in mitigating faking and enhancing the criterion-related validity of SJTs. In terms of faking, with RADS as proxy, although the warning message ($M = -0.02$, SD = 0.67) also reduced faking compared to the control message ($M = 0.22$, SD = 0.51), $t(454) = -4.40$, $p < 0.001$, $d = -0.40$, the CR-retest strategy ($M = -0.19$, SD = 0.88) was more effective than the warning message, $t(457) = -2.47$, $p < 0.05$, $d = -0.22$. With BER as proxy, the warning message ($M = 0.19$, SD = 0.17) reduced faking compared to the control message ($M = 0.28$, SD = 0.15), $t(480) = -6.58$, $p < 0.001$, $d = -0.59$. The difference between the CR-retest strategy group ($M = 0.18$, SD = 0.13) and the warning group was not significant, $t(456) = -0.25$, $p = 0.80$, $d = -0.02$.

In terms of criterion-related validity, the warning message group ($r = 0.24$) had higher validity than the control group ($r = 0.13$) with selection decision as the criterion, however the difference was not significant (*Fisher's* $z = 1.25$, $p = 0.21$, $q = 0.11$). Similarly, the criterion-related validity of the CR-retest strategy group ($r = 0.30$) was higher than that of the warning message group ($r = 0.24$), albeit not statistically significant (*Fisher's* $z = 0.71$, $p = 0.48$, $q = 0.06$). With job performance as the criterion, no statistically significant differences were found among the CR-retest strategy group ($r = 0.27$), the warning message group ($r = 0.08$), and the control group ($r = 0.09$) under the current statistical power.

Overall, Study 1 suggested that the CR-retest strategy could reduce faking and increase the criterion-related validity of SJTs in high-stakes selection settings compared to a warning message. Although the warning message also reduced faking, it did not significantly increase the criterion-related validity of SJTs. Therefore, the CR-retest strategy was a more effective strategy than the warning message in mitigating faking on SJTs.[5]

4.3 | Discussion

Although Study 1 verified the efficacy of the CR-retest strategy in reducing SJT faking, there are still unanswered questions. First, although we argued that the CR-retest strategy is a relatively mild strategy to mitigate faking in SJTs, Study 1 did not investigate whether the CR-retest strategy leads to more positive applicants' fairness perceptions. Second, Study 1 only explored the outcomes of the CR-retest strategy; we still do not know *how* the CR-retest strategy reduces faking. By understanding how and why the CR-retest strategy mitigates faking in SJTs, scholars can open the black box of faking, and practitioners can explore ways to optimize faking-mitigation strategies. Therefore, we conducted Study 2 to test applicants' fairness perceptions toward the CR-retest strategy and uncover cognitive and emotional mechanisms of the CR-retest strategy.

5 | STUDY 2

5.1 | Applicants' fairness perceptions of the CR-retest strategy

Applicants' fairness perceptions of selection procedures depend on the perceived fairness of outcome and the perceived fairness of procedure (Bauer & Truxillo, 2006; Gilliland, 1993). The CR-retest strategy may promote favorable fairness perceptions for two main

TABLE 1 Means, standard deviations, correlations, and reliabilities of variables in Study 1.

	M	SD	1	2	3	4	5	6	7	8	9
Group 1: CR-retest strategy (N^a = 245, N^b = 53)											
1. Gender	0.61	0.49									
2. Age	24.58	3.72	−0.23***								
3. Degree	1.48	0.50	−0.21***	0.85***							
4. SJT score_t1	1.39	1.20	0.05	0.24***	0.28***	(0.66)					
5. SJT score_t2	0.95	0.94	0.11	0.26***	0.27***	0.34***	(0.89)				
6. BER_t1	0.27	0.20	0.03	0.18**	0.22***	0.58***	0.24***				
7. BER_t2	0.18	0.13	0.16	0.24***	0.27***	0.37***	0.50***	0.34***			
8. RADS	−0.19	0.88	0.10	0.20**	0.20**	0.04	0.95***	0.07	0.41***		
9. Selection decision	0.22	0.41	−0.05	0.33***	0.39***	**0.14***	**0.30*****	0.10	0.27***	0.28***	
10. Job performance	1.23	0.51	0.24	−0.06	0.19	**0.09**	**0.27**	0.23	0.10	0.26	–
Group 2: Warning message (N^a = 246, N^b = 55)											
1. Gender	0.60	0.49									
2. Age	24.83	3.19	−0.18**								
3. Degree	1.45	.50	−0.11	0.75***							
4. SJT score_t1	1.40	1.11	−0.01	0.05	0.12	(0.59)					
5. SJT score_t2	1.12	0.69	−0.12	0.16*	0.19**	0.24***	(0.81)				
6. BER_t1	0.29	0.19	0.05	0.06	0.12	0.50***	0.04				
7. BER_t2	0.19	0.17	−0.08	0.11	0.13*	0.06	0.14*	0.49***			
8. RADS	−0.02	0.67	−0.12	0.14*	0.15*	−0.14*	0.93***	−0.15*	0.12		
9. Selection decision	0.22	0.42	−0.16*	0.28***	0.38***	**0.13***	**0.24*****	0.04	0.08	0.19**	
10. Job performance	1.15	0.36	0.14	−0.02	0.21	**0.07**	**0.08**	0.09	0.07	0.06	–
Group 3: Control message (N^a = 242, N^b = 54)											
1. Gender	0.63	0.48									
2. Age	24.60	3.01	−0.18**								
3. Degree	1.49	0.50	−0.16*	0.57***							
4. SJT score_t1	1.41	1.03	0.04	−0.04	0.12	(0.60)					
5. SJT score_t2	1.36	0.58	−0.00	0.04	0.08	0.51***	(0.68)				
6. BER_t1	0.28	0.17	0.13*	0.00	0.06	0.42***	0.09				
7. BER_t2	0.28	0.15	0.10	0.06	−0.02	0.17**	0.17**	0.51***			
8. RADS	0.22	0.51	−0.02	0.07	0.03	0.11	0.91***	−0.10	0.11		
9. Selection decision	0.22	0.42	−0.11	0.12	0.35***	**0.13***	**0.13***	−0.04	0.05	0.09	
10. Job performance	1.17	0.38	0.02	0.29*	0.21	**0.08**	**0.09**	0.15	−0.01	0.08	–

Note. N^a = Selection sample size, N^b = Job performance sample size. Gender: Male = 0, Female = 1. Degree: Bachelor's degree = 1, master's degree = 2. T1 = time 1 (initial block), T2 = time 2 (main block). Selection decision: No admission = 0, Admission = 1. Job performance: Average = 1, Excellent = 2, Outstanding = 3. BER = Blatant extreme responding. RADS = Regression-adjusted difference scores. McDonald's ω as a measure of reliability is on the diagonal.

*p < 0.05; **p < 0.01; ***p < 0.001.

reasons. First, incorporating constructed response items provides an open platform for applicants to reflect on their past workplace behaviors (Edwards & Arthur, 2007), and to showcase their knowledge, abilities, and skills (Lievens et al., 2019). The CR-retest strategy, by requiring the participants to first generate open-ended responses and thereafter closed-ended responses, may foster retrospective thinking and promote awareness that they can comprehensively highlight their job-related competencies. Therefore,

applicants are likely to anticipate that the test outcomes will accurately reflect their competencies. Second, the CR-retest strategy is embedded in the testing process rather than added or appearing out of nowhere. Consequently, applicants who respond honestly are unlikely to perceive differential treatment or suspect accusations of dishonesty, promoting uniformity in their perceptions of the selection procedure. Hence, we posit that applicants will report favorable perceptions of fairness in SJTs administered with the CR-retest strategy.

H3: Applicants will report a higher level of perceived fairness on SJT items with the CR-retest strategy than SJT items without the CR-retest strategy.

5.2 | Mechanisms underlying faking-mitigation strategies

Cognitive appraisal theory proposes that cognitive and emotional pathways are observable factors in explaining individual behavior. Ample research has explored the factors driving applicants' faking behavior in personnel selection. Some studies have focused on cognitive pathways and supported social norms, attitudes, and expected consequences drive applicants' faking behavior (Goffin & Boyd, 2009; McFarland & Ryan, 2006; Snell et al., 1999). Thus, by conveying the unethical nature and negative consequences of faking, warning messages can persuade applicants to stop faking (Burns et al., 2015; Feeney et al., 2023). Recently, scholars have also started to delve into the emotional pathway. Although some emotions (e.g., anxiety and stress) can cause faking (Ho et al., 2021; Powell et al., 2021), other emotions induced by faking-mitigation strategies (e.g., guilt and fear) can reduce faking (Ellingson et al., 2012; Li et al., 2022). We posit that the psychological mechanisms of the CR-retest strategy include cognitive factors (i.e., perceived threat) and emotional factors (i.e., fear). The following section presents the detailed explanation of our reasoning.

Individuals perceive threat when they receive relevant information inconsistent with their goals and that has not yet caused damage (Lazarus & Folkman, 1984). In high-stakes selection, applicants' most essential goals are to obtain high test scores and to get hired. However, the CR-retest strategy may signal to applicants that their exaggerations will be verified and identified (Desimone et al., 2015). This signal may elicit applicants' threat appraisal, especially in those intending to fake. In perceiving this threat signal, applicants may change their behavior to avoid threat (So, 2013) and seek protection (Kim & Hawkins, 2020), thereby disincentivizing their potential faking behavior.

H4: The CR-retest strategy reduces applicants'faking behavior by increasing their perceived threat.

Emotions, particularly fear, play a critical role in coping and decision-making processes. Fear appeals are persuasive and effective in driving attitude and behavior change (Nabi & Myrick, 2019; Putwain et al., 2017). Fear typically arises in situations perceived as threatening or beyond one's control (LoBue & Rakison, 2013; Smith & Ellsworth, 1985). Fear, once triggered, often motivates avoidance behaviors and risk-reduction strategies (Glotzbach et al., 2012; Lazarus, 1991; Wake et al., 2020). The CR-retest strategy inherently signals that inconsistent responses may lead to negative consequences, setting the stage for fear among applicants intending to fake their answers. In a selection context using the CR-retest strategy, we hypothesize that potential fakers are likely to experience fear and consequently become motivated to minimize faking to alleviate this negative emotion.

H5: The CR-retest strategy reduces applicants'faking behavior by triggering their fear emotions.

Cognitive appraisal theory suggests that emotions stem from cognitive processes (Cai et al., 2018; Lazarus, 1991; Moors et al., 2013). Therefore, we assume that a chain of mediating roles may be underlying the CR-retest strategy–faking behavior relationship. The CR-retest strategy may initially convey threat signals to applicants who intend to fake that their responses are likely to be verified, and this threat signal will thereafter evoke their fear (So, 2013), thereby ultimately disincentivizing their potential faking behavior.

H6: Perceived threat and fear are chain mediators between the CR-retest strategy and faking behavior; the CR-retest strategy reduces applicants'faking behavior by first increasing perceived threat and thereafter triggering fear emotions.

5.3 | Combination of two faking-mitigation strategies

In addition to our comparative evaluation of the CR-retest strategy against an established approach, we are intrigued by whether issuing explicit warning messages to applicants before the CR-retest strategy could further deter faking. Some applicants may dismiss the CR-retest strategy's threat signal because it does not explicitly outline the consequences of faking, others may even see warning messages as mere bluffs (Kuncel & Borneman, 2007). As an additional exploratory step, we combined the CR-retest strategy and a warning message to investigate if this integrated approach outperforms individual strategies. Given the evolving nature of our understanding, we frame the following research question:

RQ2: *Will the CR-retest strategy preceded by a warning message mitigate applicants'faking more effectively than the CR-retest strategy alone by triggering stronger threat perceptions?*

5.4 | Method

Similar to Study 1, we added a warning message group to compare differences between the warning message and the CR-retest

strategy. We likewise designed a group with the CR-retest strategy preceded by a warning message to answer RQ2 and investigated whether the interaction group would perform better than the CR-retest strategy or warning message groups alone. This undertaking led to a 2 (with CR-retest strategy or without) × 2 (warning or control message) between-subjects design.

5.4.1 | Sample and procedure

Participants were 273 applicants from a large public hospital in Eastern China in 2022. They were applying for the position of physician. In this sample, 48% were female, average age was 29.28 years (SD = 2.53), and 97% held PhDs in medicine. The recruitment procedure was the same as in Study 1.

5.4.2 | Field experiment

Figure 3 shows the experimental design flowchart of Study 2. Participants were randomly assigned to either the CR-retest strategy (N = 68), warning message (N = 68), CR-retest strategy + warning message (N = 69), or control (N = 68) group.[6] Designs of the CR-retest strategy and the warning message were the same as those in Study 1. The post hoc Monte Carlo power analysis simulation indicated that this sample provided a power of 74% to detect the chain mediation effect of threat and fear (Schoemann et al., 2017).

In the online psychological test, participants were asked to complete 25 hospital workplace SJT items, a personality test, and an impression management scale. Thereafter, they were asked to complete several surveys requiring them to recall their cognitions and emotions when they saw either the warning message, the repeated items, or the control message. At the final step, they completed a perceived test fairness survey. One day after the online psychological tests, all applicants attended offline interviews.

5.4.3 | Measures

SJT

We used the same SJT items as those in Study 1. McDonald's ω was 0.77.

Threat appraisal

Threat appraisal was measured with two dimensions adopted from Knapova et al. (2021). *Perceived susceptibility* was measured using three items. An example is "It is likely that I will be caught if I fake." *Perceived severity* was measured using four items. A sample item is "The negative impact of being detected as a fake is severe." Items were rated on a 5-point Likert scale from 1 (*strongly disagree*) to 5 (*strongly agree*). McDonald's ω of the total scale was 0.85.

Fear emotion

Fear was measured using a 6-item scale adapted from Li et al. (2022). An example item is "When I saw [warning message/repeated items] I felt afraid/frightened/jittery/scared/nervous/shaky." Items were rated on a 7-point Likert scale from 1 (*not at all*) to 7 (*extremely*). McDonald's ω was 0.91.

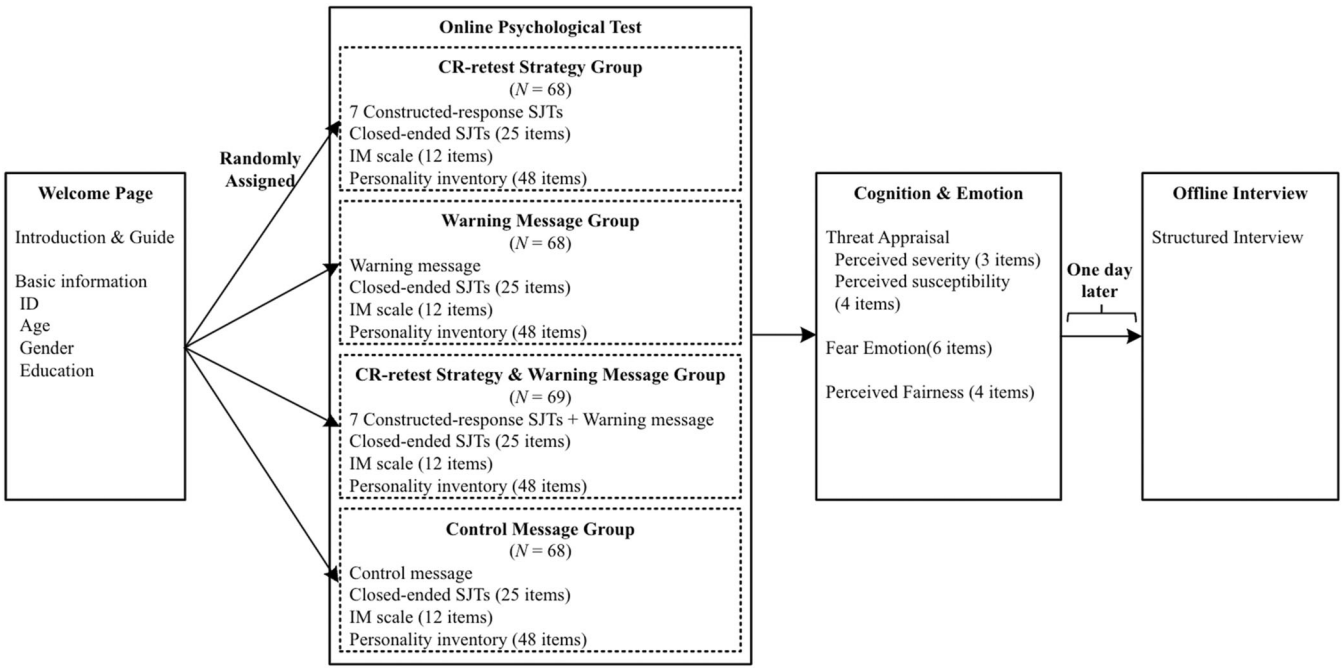

FIGURE 3 *The research design flow chart of Study 2.* IM, Impression management.

Faking behavior

Given that Study 2 focused on the psychological mechanisms of the CR-retest strategy, we used two individual-level faking proxies: BER and impression management (IM).[7] Prior investigations have established a strong relationship between IM and faking because fakers typically engage in more IM behavior in the form of response distortion compared with honest responders (Burns et al., 2015; Li et al., 2022).

BER

Similar to Study 1, we used BER to measure faking behavior. BER was measured identically as in Study 1.

IM

A 12-item scale from Paulhus (1999) was used to measure IM. A sample item is "I never cover up my mistakes." Items were rated on a 7-point Likert scale from 1 (*strongly disagree*) to 7 (*strongly agree*). Following recommendations in the literature, when applicants endorse either a 6 or 7 on the Likert scale, they would be scored a point as indicative of engaging in IM (Paulhus, 1999). McDonald's ω of the scale was 0.83.

Fairness perception

A four-item scale adopted from Smither et al. (1993) was used to measure applicants' fairness perceptions. A sample item is "Overall, I believe that the examination was fair." Items were rated on 5-point Likert scale from 1 (*strongly disagree*) to 5 (*strongly agree*). McDonald's ω was 0.84.

5.4.4 | Analytic strategy

In the analysis of our data, we used ANOVA to investigate the effects of the CR-retest strategy and the warning message on fairness perceptions (H3). Subsequently, we conducted path modeling in M*plus* 8.3 and compared model fit in R 4.3.1 to test and compare the effects of two strategies (H4−H6,). In addressing RQ2, we performed multigroup path analysis to test the interaction effect of the two faking-mitigation strategies combined.[8]

Owing to the fear variable's nonnormal distribution (floor effect), we treated it as a censored variable and used Tobit regression to estimate path coefficients (Greene, 1990). The maximum likelihood with robust standard errors (MLR) method with likelihood-ratio chi-square testing of nested neighboring models were used to test whether a specific restriction upholds (Muthén & Muthén, 1998-2017). The Satorra-Bentler scaled chi-square test was used to compare model fit (Satorra & Bentler, 2010).

5.5 | Results

Table 2 presents the descriptive statistics, correlations, and reliabilities of variables of this study. Our initial investigation aimed to replicate the effects of the CR-retest strategy on mitigating faking observed in Study 1, with our results confirming the prior findings (see online Supporting Information S1 for details).

5.5.1 | Applicants' fairness perceptions

In our investigation of fairness perceptions (H3), the results demonstrated that applicants who received the CR-retest strategy reported a higher level of perceived fairness compared with the applicants who did not receive the CR-retest strategy, $F(1, 255) = 6.03$, $p < 0.05$, $\eta_p^2 = 0.02$. Hence, H3 was supported. Consistent with previous research (e.g., Burns et al., 2015; Fan et al., 2012), applicants receiving the warning message reported lower levels of perceived fairness compared with other applicants, $F(1, 255) = 4.63$, $p < 0.05$, $\eta_p^2 = 0.02$. The fairness perception level of applicants in the combined CR-retest strategy and warning message group ($M = 3.71$, $SD = 0.53$) was similar with the CR-retest alone group ($M = 3.57$, $SD = 0.50$), $t(133) = 0.17$, $p = 0.43$.

5.5.2 | Cognitive and emotional mechanisms of the two faking-mitigation strategies

We examined underlying mechanisms of the CR-retest strategy to test H4-H6 (see Figure 4).[9] H4 stated that the CR-retest strategy would reduce faking by triggering applicants' perceived threat. With BER as proxy of faking, the CR-retest strategy triggered applicants' perceived threat ($b = 0.32$, $p < 0.001$) and reduced faking (indirect effect = −0.02, 95% CI = [−0.06, −0.00]). Hence, H4 was supported. H5 stated that the CR-retest strategy would reduce faking by triggering applicants' fear emotions. H5 was not supported because the link between the CR-retest strategy and fear was insignificant ($b = 0.12$, $p = 0.15$). H6, proposing a chain mediation from the CR-retest strategy to faking via threat appraisal followed by fear, was not supported because of a nonsignificant indirect effect (indirect effect = −0.00, 95% CI = [−0.01, 0.00]).

The warning message also triggered perceived threat ($b = 0.25$, $p < 0.01$) but the mediation effect was not significant (indirect effect = −0.01, 95% CI = [−0.04, 0.00]). Furthermore, the warning message induced applicants' fear ($b = 0.18$, $p < 0.01$), thereby reducing faking (indirect effect = −0.02, 95% CI = [−0.04, −0.00]). Lastly, the warning message triggered applicants' threat perceptions and, consequently, their fear emotions, thereby reducing faking via chain mediation (indirect effect = −0.01, 95% CI = [−0.02, −0.00]). Similar results were obtained when using IM as a measure of faking behavior (see Figure 4).

5.5.3 | Effects of the integrated faking-mitigation strategy

Multigroup path analysis was performed to test RQ2, whether a combination of the CR-retest strategy and a warning message could

TABLE 2 Means, standard deviations, correlations, and reliabilities of variables in Study 2.

		M	SD	1	2	3	4	5	6	7	8	9	10	*M*	SD	
Group 1: CR-retest strategy (*N* = 68)	1. Gender	.49	.50		−0.10	−0.13	−0.13	−0.05	−0.07	−0.02	.10	−0.10	−0.09	.46	.50	Group 4: Control message (*N* = 68)
	2. Age	29.40	2.69	−0.19		.11	−0.20	.02	.13	−0.14	−0.01	−0.16	−0.14	29.24	2.41	
	3. Degree	1.94	.24	−0.13	.34**		−0.14	−0.08	.11	−0.01	−0.03	−0.03	−0.07	1.99	.12	
	4. Selection decision	.46	.50	.23	−0.11	.10		−0.03	.01	.12	.20	−0.08	**.14**	.43	.50	
	5. Threat appraisal	2.98	.48	.02	−0.10	−0.03	.05		.29*	−0.04	.04	.00	.25*	2.61	.60	
	6. Fear	1.78	.71	.05	.18	.03	.07	.20		−0.07	−0.20	−0.21	−0.09	1.52	.56	
	7. IM	7.50	2.93	.04	−.09	.15	−0.09	−0.38**	−0.28*		.46***	−0.01	.20	8.90	2.32	
	8. BER	.26	.17	.13	.14	.12	−0.07	−0.56***	−0.05	.36**		−0.19	.50***	.38	.19	
	9. Fairness	3.57	.50	−0.32**	.04	−0.03	−0.12	−0.07	−0.07	−0.04	.36**		−0.01	3.53	.78	
	10. SJT score	2.20	.35	.28*	−.13	−0.03	**.27***	−0.31*	−0.06	.26*	.42**	.01		2.38	.50	
Group 2: Warning message (*N* = 68)	1. Gender	.50	.50		−.10	−0.01	−0.11	−0.05	−0.05	−0.03	−0.13	.02	−0.13	.46	.50	Group 3: CR-retest strategy & Warning message (*N* = 69)
	2. Age	29.32	2.28	−0.17		.26*	−0.03	−0.07	.05	.17	.07	−0.02	−0.01	29.16	2.75	
	3. Degree	1.99	.12	.12	.02		−0.02	−0.12	.04	.17	.10	−0.17	.04	1.97	.17	
	4. Selection decision	.43	.50	.21	−0.22	.11		.21	.24*	−0.05	−0.12	.09	**.29***	.43	.50	
	5. Threat appraisal	2.91	.58	.17	.01	−0.14	−0.11		.14	−0.74***	−0.56***	.15	−0.34**	3.71	.53	
	6. Fear	1.94	.86	−.04	.17	−0.03	.00	.47***		−0.10	−0.07	.07	−0.21	1.82	.65	
	7. IM	7.78	2.75	−0.04	−.10	.17	.09	−0.50***	−0.54***		.45***	−0.04	.39**	4.81	3.46	
	8. BER	.31	.18	−.12	.13	.14	.11	−0.52***	−0.36**	.43***		.23	.20	.11	.06	
	9. Fairness	3.18	.87	.13	−0.13	−0.01	−0.04	−0.41***	−0.60***	.32**	.20		−0.02	3.56	.58	
	10. SJT score	2.33	.40	.16	.05	.05	**.18**	−0.24*	−0.21	.31**	.42***	.36**		1.95	.47	

Note. N = 273. Gender: Male = 0, Female = 1. Degree: Master's degree = 1, PhD degree = 2. Selection decision: No admission = 0, Admission = 1. IM = Impression management scale. BER = Blatant extreme responding. Reliability of variables is measured by McDonald's ω: Threat appraisal = 0.85, Fear = 0.91, Impression Management = 0.83, Fairness perception = 0.84, SJT score = 0.77.

*$p < .05$; **$p < .01$; ***$p < .001$.

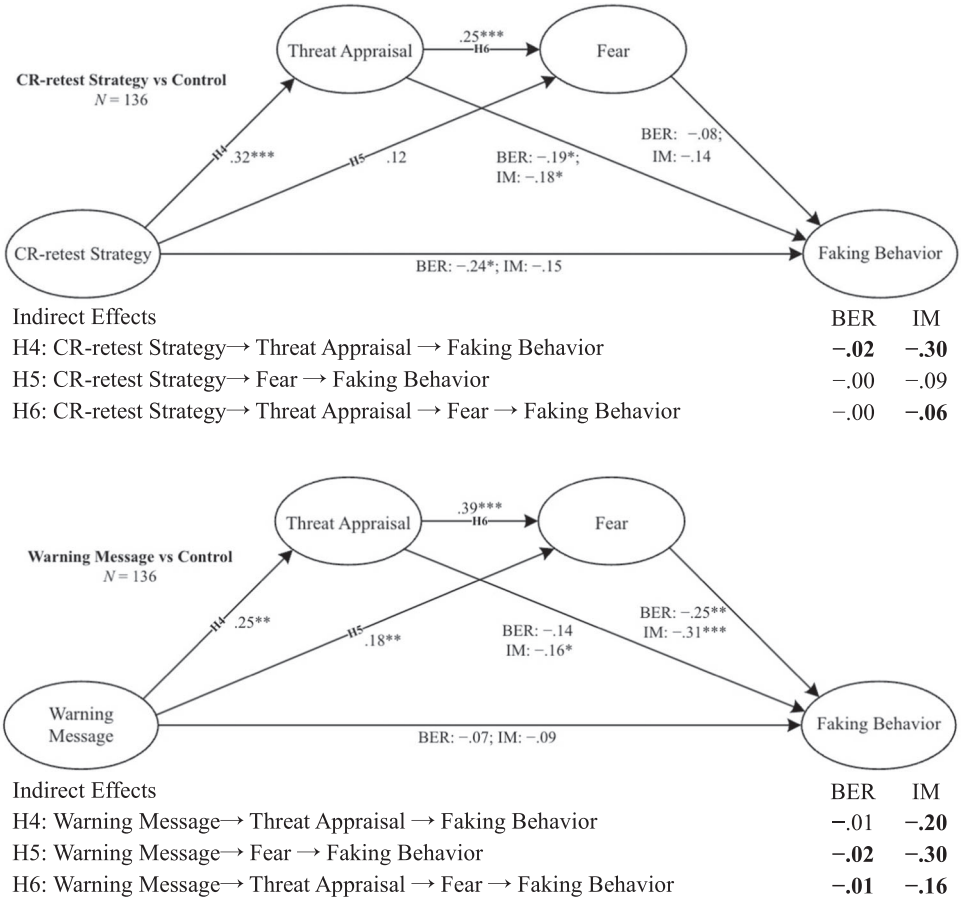

FIGURE 4 *Path analysis results of Study 2*. BER = Blatant extreme responding. IM = Impression management scale. For clarity of presentation, we did not include control variables (age, gender, and degree) in the figure. The complete standardized Tobit regression results are presented in the Appendix A1 (Model 1 and Model 2). Bolded indirect effects mean their confidence intervals do not contain zero and thus are significant. $^{*}p < 0.05$. $^{**}p < 0.01$. $^{***}p < 0.001$.

mitigate faking more effectively than the CR-retest strategy alone by evoking stronger threat perceptions. The Satorra-Bentler scaled chi-square test indicated significant fit improvements between the constrained and original models (BER: $\Delta\chi^2 = 25.11$, $\Delta df = 6$, $p < 0.001$; IM: $\Delta\chi^2 = 38.80$, $\Delta df = 6$, $p < 0.001$), indicating that the inclusion of a warning message changed path coefficients between CR-retest strategy, threat, fear, or faking.[10]

With BER as proxy of faking behavior, the path *CR-retest strategy–threat appraisal–faking behavior* was more robust for groups with a warning message (indirect effect = −0.07, 95% CI = [−0.10, −0.04]) than for those receiving only the CR-retest strategy (indirect effect = −0.03, 95% CI = [−0.06, −0.01]), Wald $\chi^2(df = 1) = 7.49$, $p < 0.01$. No significant difference was observed in the indirect effect of the *CR-retest strategy–fear–faking behavior* link for applicants receiving a warning message and the CR-retest strategy (indirect effect = 0.01, 95% CI = [−0.00, 0.03]) and applicants receiving only the CR-retest strategy (indirect effect = −0.00, 95% CI = [−0.01, 0.00]), Wald $\chi^2(df = 1) = 3.03$, $p = 0.08$. These results indicated that applicants who received the CR-retest strategy with (or without) warning message did not reduce their faking owing to the experience of fear but by perceived threat alone. Similar results were found with IM as proxy of faking.

5.6 | Discussion

Overall, study 2 suggested that the CR-retest strategy was more applicant-friendly than a warning message because applicants in the CR-retest strategy group reported more positive fairness perceptions than applicants in the warning message group. The CR-retest strategy mitigated faking by triggering threat directly, whereas the warning message directly triggered fear. The combination of the two strategies performed best because applicants in the combination group had the lowest faking levels of all four groups, while reporting similar levels of fairness perceptions as the CR-retest strategy group alone.

6 | GENERAL DISCUSSION

To guarantee a selection procedure's validity and fairness, numerous researchers have addressed faking in personnel selection and proposed several faking-mitigation strategies. These strategies, although generally effective in reducing faking, yield some unintended adverse effects, particularly on applicants' fairness

perceptions (Burns et al., 2015; Feeney et al., 2023; Li et al., 2022). Moreover, only a few studies have focused on faking-mitigation strategies applicable to SJTs (e.g., Lievens & Peeters, 2008; Oostrom et al., 2017). Our research proposed the CR-retest strategy, which could effectively reduce faking in SJTs without undermining applicants' fairness perceptions.

Our main findings are threefold. First, the CR-retest strategy effectively mitigated faking and improved the criterion-related validity of SJTs in our study's selection context, surpassing the efficacy of a warning message. Second, the CR-retest strategy increased applicants' fairness perceptions, whereas a warning message decreased fairness perceptions. Third, the effects of the CR-retest strategy and warning message were different, with the former eliciting threat and the latter eliciting fear. The combination of the two strategies resulted in increased perceived threat and decreased faking, without undermining applicants' fairness perceptions.

6.1 | Theoretical contributions

First, this study extends a small body of research on SJT faking-mitigation strategies (e.g., Lievens & Peeters, 2008; Oostrom et al., 2017) by proposing a new strategy: the CR-retest strategy. We demonstrated that the CR-retest strategy could effectively reduce faking across three operationalizations identified in prior studies as indicative of faking (Burns et al., 2015; Feeney et al., 2023; Li et al., 2022).

Second, this study enhances our theoretical understanding of the mechanisms of faking-mitigation strategies. On the basis of cognitive appraisal theory (Lazarus & Folkman, 1984; Lazarus, 1991, 1999), we uncovered the mediating role of perceived threat and fear, extending the research on cognitive (McFarland & Ryan, 2006; Snell et al., 1999) and emotional pathways (Ellingson et al., 2012; Li et al., 2022) underlying faking behavior. Our results extend these studies by showing that the CR-retest strategy evokes threat but does not heighten fear or undermine fairness perceptions. As such, the CR-retest strategy can be regarded as more friendly compared with traditional fear-evoking strategies.

Third, our study contributes to the ongoing debate on whether faking undermines the criterion-related validity of selection instruments. Ingold et al. (2015) found a positive link between faking and supervisor-rated performance. Donovan et al. (2014) observed unfavorable sales performance in fakers. Our results add to the body of studies (Donovan et al., 2014; Komar et al., 2008; Lanyon et al., 2014; Peterson et al., 2011) by demonstrating that faking inflates scores and impairs the criterion-related validity. Faking-mitigation strategies can thus optimize validity by suppressing faking.

Fourth, we also enrich research on warning messages (e.g., Feeney et al., 2023) applied to SJTs in selection settings. Warning messages are considered among the most effective tools to combat faking on noncognitive tests (Fan et al., 2012). However, our results suggested that warning messages may not be as effective in SJTs as in personality tests because applicants did not consider warning messages as a real threat in SJTs.

6.2 | Practical implications

Our study extends practitioners' knowledge and available repertoire relating to effective faking-mitigation strategies in SJTs by proposing an effective and applicant-friendly strategy. The CR-retest strategy combines the practicality of closed-ended response formats with the benefit of having candidates elaborate on their answers before selecting from given options. Moreover, we established that the CR-retest strategy evoked positive applicant reactions. Thus, there is no potentially negative impact on organizations' image.

We likewise provide new perspectives for the selection field to organically combine different strategies in practice. By combining the CR-retest strategy and a warning message, we discovered additive effects on reducing faking by inducing additional threat and reducing fear. These findings suggested that fear invoked by a warning message may be offset by enabling applicants to present themselves fully in an open-ended response format.

6.3 | Limitations and future directions

First, the design of the CR-retest strategy raises concerns on its efficacy and practicality. This research used seven constructed response SJT items based on psychometric considerations and practicality. However, this method begs the question whether seven is the optimal number. Adding more items may improve the CR-retest strategy's impact on faking, while fewer items could offer a more efficient solution with shorter testing times. Moreover, repetition of the same items poses a risk to test security, and the extended testing time may cause applicant fatigue. Future research should address these concerns.

Second, our research design and measurement of key variables may limit the conclusions we could draw. Although collecting data from real-world hiring processes enhances external validity, it limits internal validity in controlling external influences. The cross-sectional nature of our data may introduce common method variance in some key variables (Bullock & Green, 2021). Future research could opt for the repeated measures design in a lab setting (Burns & Christiansen, 2011) to investigate the CR-retest strategy's causal effects. Furthermore, employing a one-item measure of job performance and relying on difference scores as indicators of faking (Edwards, 2001, 2002; Shanock et al., 2010) might limit the robustness of our findings.

Third, the characteristics of our sample may limit the generalizability of this study's results across different organizations and cultures. Chinese culture is high in collectivism and power distance (Hofstede, 2001), and applicants tend to adhere to recruiters' instructions and reduce faking for compliance or the public interest in increasing test fairness. This culture may therefore inflate CR-retest strategy's effects. Moreover, our studies were conducted in a highly specific work context (i.e., a hospital). Future research should focus on the generalizability of the CR-retest strategy across cultures and professions.

Fourth, applicants' responses to the constructed response SJT items call for an in-depth analysis. Although some studies have suggested that applicants can also distort their replies to constructed response items (Forsyth & Anglim, 2020), there is no systematic research on their susceptibility to faking. The content of constructed response SJTs may provide extensive information not captured by closed-ended SJTs, thereby providing incremental predictive validity. We expect future research to apply natural language processing (NLP) to delve into the CR-retest strategy.

7 | CONCLUSION

This study introduced the CR-retest strategy to enhance the selection quality of SJTs. We examined the effectiveness of this strategy in mitigating faking and investigated its mechanisms based on cognitive appraisal theory. The findings indicated that the CR-retest strategy can mitigate SJT score inflation by triggering applicants' threat perceptions, and can potentially enhance the validity in predicting selection outcomes and job performance in practice. Compared to a traditional warning method, the CR-retest strategy was more effective in reducing SJT score inflation while maintaining positive perceptions of test fairness among applicants. Based on these results, we recommend that organizations consider integrating the CR-retest strategy into their assessment systems to address the issue of applicant faking in SJTs.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The selection data are not publicly available due to their proprietary nature. The analysis code, results, research materials, and part of the study 2 data are available at the open science framework (https://osf.io/tk9s5). The study design was not preregistered because the data were collected for an applied selection procedure.

ORCID

Liyan Xi http://orcid.org/0000-0002-1507-8439
Qingxiong Weng http://orcid.org/0000-0002-5426-6029
Jan Corstjens https://orcid.org/0000-0001-8192-339X
Xiujuan Wang http://orcid.org/0009-0006-9241-2038
Lixin Chen http://orcid.org/0000-0001-6664-1634

ENDNOTES

[1] Please see the randomization check results of Studies 1 and 2 in the online Supporting Information S1. No differences were observed in applicants across the different groups.

[2] Please see the detailed information and results of the interview in the online Supporting Information S1.

[3] In the online Supporting Information S1, we explained our faking indicators measurement and the rationale for their inclusion in Studies 1 and 2.

[4] In the online Supporting Information S1, we also used raw difference score to measure faking because this score provides a detailed accounting of score change and direction (Burns & Christiansen, 2011).

[5] A potential concern is that different testing experiences and times may affect the results. Thus, we conducted an additional study to control for this confounder effect in the online Supporting Information S1.

[6] Please see the randomization check results of Studies 1 and 2 in the online Supporting Information S1. No differences were observed in the applicants across the different groups.

[7] In the Supporting Information S1, we offered a comprehensive explanation of our choice of faking measures and the rationale for their inclusion in Studies 1 and 2.

[8] Please see the common method biases check results of the Study 2 variables in the online Supporting Information S1. The results suggest that our measure captured distinct constructs.

[9] The detailed standardized Tobit regression results are shown in Appendix.

[10] The detailed multigroup path analysis results of each path are shown in the online Supporting Information S1.

REFERENCES

Arthur, Jr., W., Edwards, B. D., & Barrett, G. V. (2002). Multiple-choice and constructed response tests of ability: Race-based subgroup performance differences on alternative paper-and-pencil test formats. *Personnel Psychology*, *55*(4), 985–1008. https://doi.org/10.1111/j.1744-6570.2002.tb00138.x

Arthur, W., Glaze, R. M., Jarrett, S. M., White, C. D., Schurig, I., & Taylor, J. E. (2014). Comparative evaluation of three situational judgment test response formats in terms of construct-related validity, subgroup differences, and susceptibility to response distortion. *Journal of Applied Psychology*, *99*(3), 535–545. https://doi.org/10.1037/a0035788

Bauer, T. N., & Truxillo, D. M. (2006). Applicant reactions to situational judgment tests: Research and related practical issues. In J. A. Weekley, & R. E. Ployhart (Eds.), *Situational Judgment Tests: Theory, Measurement, and Application* (pp. 233–249). Lawrence Erlbaum Associates Publishers. https://doi.org/10.4324/9780203774878

Bill, B., & Melchers, K. G. (2022). Thou Shalt not Lie! Exploring and testing countermeasures against faking intentions and faking in selection interviews. *International Journal of Selection and Assessment*, *31*(1), 22–44. https://doi.org/10.1111/ijsa.12402

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

Bullock, J. G., & Green, D. P. (2021). The failings of conventional mediation analysis and a design-based alternative. *Advances in Methods and Practices in Psychological Science*, *4*(4), 25152459211047227. https://doi.org/10.1177/25152459211047227

Burns, G. N., & Christiansen, N. D. (2011). Methods of measuring faking behavior. *Human Performance*, *24*(4), 358–372. https://doi.org/10.1080/08959285.2011.597473

Burns, G. N., Filipkowski, J. N., Morris, M. B., & Shoda, E. A. (2015). Impact of electronic warnings on online personality scores and test-taker reactions in an applicant simulation. *Computers in Human Behavior*, *48*, 163–172. https://doi.org/10.1016/j.chb.2015.01.051

Cai, R., Lu, L., & Gursoy, D. (2018). Effect of disruptive customer behaviors on others' overall service experience: An appraisal theory perspective. *Tourism Management*, *69*, 330–344. https://doi.org/10.1016/j.tourman.2018.06.013

Cao, M., & Drasgow, F. (2019). Does forcing reduce faking? A meta-analytic review of forced-choice personality measures in high-stakes

situations. *Journal of Applied Psychology*, *104*(11), 1347–1368. https://doi.org/10.1037/apl0000414

Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology*, *63*(1), 83–117. https://doi.org/10.1111/j.1744-6570.2009.01163.x

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science*, *1*(3), 98–101. https://doi.org/10.1111/1467-8721.ep10768783

Converse, P. D., Oswald, F. L., Imus, A., Hedricks, C., Roy, R., & Butera, H. (2008). Comparing personality test formats and warnings: Effects on criterion-related validity and test-taker reactions. *International Journal of Selection and Assessment*, *16*(2), 155–169. https://doi.org/10.1111/j.1468-2389.2008.00420.x

Dalal, D. K., Zhu, X., Rangel, B., Boyce, A. S., & Lobene, E. (2021). Improving applicant reactions to forced-choice personality measurement: Interventions to reduce threats to test takers' self-concepts. *Journal of Business and Psychology*, *36*(1), 55–70. https://doi.org/10.1007/s10869-019-09655-6

Desimone, J. A., Harms, P. D., & Desimone, A. J. (2015). Best practice recommendations for data screening. *Journal of Organizational Behavior*, *36*(2), 171–181. https://doi.org/10.1002/job.1962

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. https://doi.org/10.1007/s10869-013-9318-5

Dwight, S. A., & Donovan, J. J. (2003). Do warnings not to fake reduce faking? *Human Performance*, *16*(1), 1–23. https://doi.org/10.1207/S15327043HUP1601_1

Edwards, B. D., & Arthur, W. (2007). An examination of factors contributing to a reduction in subgroup differences on a constructed response paper-and-pencil test of scholastic achievement. *Journal of Applied Psychology*, *92*(3), 794–801. https://doi.org/10.1037/0021-9010.92.3.794

Edwards, J. R. (2001). Ten difference score myths. *Organizational Research Methods*, *4*(3), 265–287. https://doi.org/10.1177/109442810143005

Edwards, J. R. (2002). Alternatives to difference scores: Polynomial regression analysis and response surface methodology. In F. Drasgow & N. Schmitt, (Eds.), *Measuring and analyzing behavior in organizations: Advances in measurement and data analysis* (pp. 350–400). Jossey-Bass/Wiley. https://doi.org/10.1016/j.leaqua.2021.101592

Ellingson, J. E., Heggestad, E. D., & Makarius, E. E. (2012). Personality retesting for managing intentional distortion. *Journal of Personality and Social Psychology*, *102*(5), 1063–1076. https://doi.org/10.1037/a0027327

Fan, J., Gao, D., Carroll, S. A., Lopez, F. J., Tian, T. S., & Meng, H. (2012). Testing the efficacy of a new procedure for reducing faking on personality tests within selection contexts. *Journal of Applied Psychology*, *97*(4), 866–880. https://doi.org/10.1037/a0026655

Feeney, J. R., Goffin, R. D., & Beshai, S. (2023). Applicant faking warnings: Are they really effective? *Personality and Individual Differences*, *200*, 111899. https://doi.org/10.1016/j.paid.2022.111899

Forsyth, L., & Anglim, J. (2020). Using text analysis software to detect deception in written short-answer questions in employee selection. *International Journal of Selection and Assessment*, *28*(3), 236–246. https://doi.org/10.1111/ijsa.12284

Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *The Academy of Management Review*, *18*(4), 694–734. https://doi.org/10.2307/258595

Glotzbach, E., Ewald, H., Andreatta, M., Pauli, P., & Mühlberger, A. (2012). Contextual fear conditioning predicts subsequent avoidance behaviour in a virtual reality environment. *Cognition & Emotion*, *26*(7), 1256–1272. https://doi.org/10.1080/02699931.2012.656581

Goffin, R. D., & Boyd, A. C. (2009). Faking and personality assessment in personnel selection: Advancing models of faking. *Canadian Psychology/Psychologie Canadienne*, *50*(3), 151–160. https://doi.org/10.1037/a0015946

Greene, W. H. (1990). *Econometric Analysis*. Macmillan Publishing Company.

Griffith, R. L., Malm, T., English, A., Yoshita, Y., & Gujar, A. (2006). Applicant faking behavior: Teasing apart the influence of situational variance, cognitive biases, and individual differences. In R. L. Griffith & M. H. Peterson, (Eds.), *A closer examination of applicant faking behavior* (pp. 151–177). Information Age.

Heggestad, E. D., Morrison, M., Reeve, C. L., & McCloy, R. A. (2006). Forced-choice assessments of personality for selection: Evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology*, *91*(1), 9–24. https://doi.org/10.1037/0021-9010.91.1.9

Ho, J. L., Perossa, A., Fancett, R. S., & Powell, D. M. (2021). Examining the situational antecedents of interview faking behavior: A qualitative study. *International Journal of Selection and Assessment*, *29*(3–4), 427–447. https://doi.org/10.1111/ijsa.12354

Hofstede, G. (2001). *Culture's consequences: Comparing values, behavior, institutions, and organizations across nations* (2nd ed.). Sage.

Holden, R. R., Wood, L. L., & Tomashewski, L. (2001). Do response time limitations counteract the effect of faking on personality inventory validity? *Journal of Personality and Social Psychology*, *81*(1), 160–169. https://doi.org/10.1037/0022-3514.81.1.160

van Hooft, E. A. J., & 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. https://doi.org/10.1037/a0025711

Ingold, P. V., Kleinmann, M., König, C. J., & Melchers, K. G. (2015). Shall we continue or stop disapproving of self-presentation? Evidence on impression management and faking in a selection context and their relation to job performance. *European Journal of Work and Organizational Psychology*, *24*(3), 420–432. https://doi.org/10.1080/1359432X.2014.915215

Kaminski, K., Felfe, J., Schäpers, P., & Krumm, S. (2019). A closer look at response options: Is judgment in situational judgment tests a function of the desirability of response options? *International Journal of Selection and Assessment*, *27*(1), 72–82. https://doi.org/10.1111/ijsa.12233

Kanning, U. P., & Kuhne, S. (2006). Social desirability in a multimodal personnel selection test battery. *European Journal of Work and Organizational Psychology*, *15*(3), 241–261. https://doi.org/10.1080/13594320600625872

Kasten, N., Freund, P. A., & Staufenbiel, T. (2020). "Sweet little lies": An in-depth analysis of faking behavior on situational judgment tests compared to personality questionnaires. *European Journal of Psychological Assessment*, *36*(1), 136–148. https://doi.org/10.1027/1015-5759/a000479

Kim, S. C., & Hawkins, K. H. (2020). The psychology of social media communication in influencing prevention intentions during the 2019 U.S. measles outbreak. *Computers in Human Behavior*, *111*, 106428. https://doi.org/10.1016/j.chb.2020.106428

Knapova, L., Kruzikova, A., Dedkova, L., & Smahel, D. (2021). Who is smart with their smartphones? Determinants of smartphone security behavior. *Cyberpsychology, Behavior and Social Networking*, *24*(9), 584–592. https://doi.org/10.1089/cyber.2020.0599

Komar, S., Brown, D. J., Komar, J. A., & Robie, C. (2008). Faking and the validity of conscientiousness: A Monte Carlo investigation. *Journal of Applied Psychology*, *93*(1), 140–154. https://doi.org/10.1037/0021-9010.93.1.140

Komar, S., Komar, J. A., Robie, C., & Taggar, S. (2010). Speeding personality measures to reduce faking: A self-regulatory model. *Journal of Personnel Psychology*, *9*(3), 126–137. https://doi.org/10.1027/1866-5888/a000016

Krumm, S., Lievens, F., Hüffmeier, J., Lipnevich, A. A., Bendels, H., & Hertel, G. (2015). How "situational" is judgment in situational judgment tests? *Journal of Applied Psychology*, *100*(2), 399–416. https://doi.org/10.1037/a0037674

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. https://doi.org/10.1111/j.1468-2389.2007.00383.x

Landers, R. N., Sackett, P. R., & Tuzinski, K. A. (2011). Retesting after initial failure, coaching rumors, and warnings against faking in online personality measures for selection. *Journal of Applied Psychology*, *96*(1), 202–210. https://doi.org/10.1037/a0020375

Lanyon, R. I., Goodstein, L. D., & Wershba, R. (2014). 'Good impression' as a moderator in employment-related assessment. *International Journal of Selection and Assessment*, *22*(1), 52–61. https://doi.org/10.1111/ijsa.12056

Lazarus, R. S. (1991). *Emotion and adaption*. Oxford University Press.

Lazarus, R. S. (1999). *Stress and emotion: A new synthesis*. Springer.

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. Springer.

Lee, H., Smith, W. Z., & Geisinger, K. F. (2017). Faking under a nonlinear relationship between personality assessment scores and job performance. *International Journal of Selection and Assessment*, *25*(3), 284–298. https://doi.org/10.1111/ijsa.12180

Leeds, J. P. (2012). The theory of cognitive acuity: Extending psychophysics to the measurement of situational judgment. *Journal of Neuroscience, Psychology, and Economics*, *5*(3), 166–181. https://doi.org/10.1037/a0027294

Leeds, J. P. (2018). Applying cognitive acuity theory to the development and scoring of situational judgment tests. *Behavior Research Methods*, *50*(6), 2215–2225. https://doi.org/10.3758/s13428-017-0988-1

de Leng, W. E., Stegers-Jager, K. M., Born, M. P., & Themmen, A. P. N. (2018). Influence of response instructions and response format on applicant perceptions of a situational judgement test for medical school selection. *BMC Medical Education*, *18*(1), 282. https://doi.org/10.1186/s12909-018-1390-0

de Leng, W. E., Stegers-Jager, K. M., Born, M. P., & Themmen, A. P. N. (2019). Faking on a situational judgment test in a medical school selection setting: Effect of different scoring methods? *International Journal of Selection and Assessment*, *27*(3), 235–248. https://doi.org/10.1111/ijsa.12251

Levashina, J., Morgeson, F. P., & Campion, M. A. (2012). Tell me some more: Exploring how verbal ability and item verifiability influence responses to biodata questions in a high-stakes selection context. *Personnel Psychology*, *65*(2), 359–383. https://doi.org/10.1111/j.1744-6570.2012.01245.x

Levashina, J., Weekley, J. A., Roulin, N., & Hauck, E. (2014). Using blatant extreme responding for detecting faking in high-stakes selection: Construct validity, relationship with general mental ability, and subgroup differences. *International Journal of Selection and Assessment*, *22*(4), 371–383. https://doi.org/10.1111/ijsa.12084

Li, H., Fan, J., Zhao, G., Wang, M., Zheng, L., Meng, H., Weng, Q., Liu, Y., & Lievens, F. (2022). The role of emotions as mechanisms of mid-test warning messages during personality testing: A field experiment. *Journal of Applied Psychology*, *107*(1), 40–59. https://doi.org/10.1037/apl0000885

Lievens, F., De Corte, W., & Westerveld, L. (2015). Understanding the building blocks of selection procedures: Effects of response fidelity on performance and validity. *Journal of Management*, *41*(6), 1604–1627. https://doi.org/10.1177/0149206312463941

Lievens, F., & Motowidlo, S. J. (2016). Situational judgment tests: From measures of situational judgment to measures of general domain knowledge. *Industrial and Organizational Psychology*, *9*(1), 3–22. https://doi.org/10.1017/iop.2015.71

Lievens, F., & Peeters, H. (2008). Impact of elaboration on responding to situational judgment test items. *International Journal of Selection and Assessment*, *16*(4), 345–355. https://doi.org/10.1111/j.1468-2389.2008.00440.x

Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: A review of recent research. *Personnel Review*, *37*(4), 426–441. https://doi.org/10.1108/00483480810

Lievens, F., Sackett, P. R., Dahlke, J. A., Oostrom, J. K., & De Soete, B. (2019). Constructed response formats and their effects on minority-majority differences and validity. *Journal of Applied Psychology*, *104*(5), 715–726. https://doi.org/10.1037/apl0000367

Lievens, F., Sackett, P. R., & Zhang, C. (2021). Personnel selection: A longstanding story of impact at the individual, firm, and societal level. *European Journal of Work and Organizational Psychology*, *30*(3), 444–455. https://doi.org/10.1080/1359432X.2020.1849386

LoBue, V., & Rakison, D. H. (2013). What we fear most: A developmental advantage for threat-relevant stimuli. *Developmental Review*, *33*(4), 285–303. https://doi.org/10.1016/j.dr.2013.07.005

Lopez, F. J., Hou, N., & Fan, J. (2019). Reducing faking on personality tests: Testing a new faking-mitigation procedure in a U.S. job applicant sample. *International Journal of Selection and Assessment*, *27*(4), 371–380. https://doi.org/10.1111/ijsa.12265

McFarland, L. A., & Ryan, A. M. (2006). Toward an integrated model of applicant faking behavior. *Journal of Applied Social Psychology*, *36*(4), 979–1016. https://doi.org/10.1111/j.0021-9029.2006.00052.x

Melchers, K. G., Roulin, N., & Buehl, A. K. (2020). A review of applicant faking in selection interviews. *International Journal of Selection and Assessment*, *28*(2), 123–142. https://doi.org/10.1111/ijsa.12280

Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal theories of emotion: State of the art and future development. *Emotion Review*, *5*(2), 119–124. https://doi.org/10.1177/1754073912468165

Motowidlo, S. J., & Beier, M. E. (2010). Differentiating specific job knowledge from implicit trait policies in procedural knowledge measured by a situational judgment test. *Journal of Applied Psychology*, *95*(2), 321–333. https://doi.org/10.1037/a0017975

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology*, *75*(6), 640–647. https://doi.org/10.1037/0021-9010.75.6.640

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus User's Guide (8th ed.)*. Muthén & Muthén.

Nabi, R. L., & Myrick, J. G. (2019). Uplifting fear appeals: Considering the role of hope in fear-based persuasive messages. *Health Communication*, *34*(4), 463–474. https://doi.org/10.1080/10410236.2017.1422847

Nguyen, N. T., Biderman, M. D., & McDaniel, M. A. (2005). Effects of response instructions on faking a situational judgment test. *International Journal of Selection and Assessment*, *13*(4), 250–260. https://doi.org/10.1111/j.1468-2389.2005.00322.x

O'Neill, T. A., Lee, N. M., Radan, J., Law, S. J., Lewis, R. J., & Carswell, J. J. (2013). The impact of "non-targeted traits" on personality test faking, hiring, and workplace deviance. *Personality and Individual Differences*, *55*(2), 162–168. https://doi.org/10.1016/j.paid.2013.02.027

Oostrom, J. K., Köbis, N. C., Ronay, R., & Cremers, M. (2017). False consensus in situational judgment tests: What would others do? *Journal of Research in Personality*, *71*, 33–45. https://doi.org/10.1016/j.jrp.2017.09.001

Paulhus, D. L. (1999). *Paulhus Deception Scales (PDS): The balanced inventory of desirable responding – 7*. Multi-Health Systems.

Peeters, H., & Lievens, F. (2005). Situational judgment tests and their predictiveness of college students' success: The influence of faking. *Educational and Psychological Measurement*, *65*(1), 70–89. https://doi.org/10.1177/0013164404268672

Peterson, M. H., Griffith, R. L., Isaacson, J. A., O'Connell, M. S., & Mangos, P. M. (2011). Applicant faking, social desirability, and the prediction of counterproductive work behaviors. *Human Performance*, *24*(3), 270–290. https://doi.org/10.1080/08959285.2011.580808

Powell, D. M., Bourdage, J. S., & Bonaccio, S. (2021). Shake and fake: The role of interview anxiety in deceptive impression management. *Journal of Business and Psychology*, *36*(5), 829–840. https://doi.org/10.1007/s10869-020-09708-1

Putwain, D. W., Symes, W., & Wilkinson, H. M. (2017). Fear appeals, engagement, and examination performance: The role of challenge and threat appraisals. *British Journal of Educational Psychology*, *87*(1), 16–31. https://doi.org/10.1111/bjep.12132

Ramsay, L. J., Schmitt, N., Oswald, F. L., Kim, B. H., & Gillespie, M. A. (2006). The impact of situational context variables on responses to biodata and situational judgment inventory items. *Psychology Science*, *48*(3), 268–287.

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. https://doi.org/10.1016/j.hrmr.2006.03.004

Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2022). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *Journal of Applied Psychology*, *107*(11), 2040–2068. https://doi.org/10.1037/apl0000994

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, *75*(2), 243–248. https://doi.org/10.1007/s11336-009-9135-y

Schmitt, N., & Kunce, C. (2002). The effects of required elaboration of answers to biodata questions. *Personnel Psychology*, *55*(3), 569–587. https://doi.org/10.1111/j.1744-6570.2002.tb00121.x

Schmitt, N., Prasad, J. J., Ryan, A. M., Bradburn, J. C., & Nye, C. D. (2019). Culture as a determinant of option choice in a situational judgement test: A new look. *Journal of Occupational and Organizational Psychology*, *92*(2), 330–351. https://doi.org/10.1111/joop.12251

Schneider, T. J., & Goffin, R. D. (2012). Perceived ability to deceive and incremental prediction in pre-employment personality testing. *Personality and Individual Differences*, *52*(7), 806–811. https://doi.org/10.1016/j.paid.2012.01.015

Schoemann, A. M., Boulton, A. J., & Short, S. D. (2017). Determining power and sample size for simple and complex mediation models. *Social Psychological and Personality Science*, *8*(4), 379–386. https://doi.org/10.1177/1948550617715068

Shanock, L. R., Baran, B. E., Gentry, W. A., Pattison, S. C., & Heggestad, E. D. (2010). Polynomial regression with response surface analysis: A powerful approach for examining moderation and overcoming limitations of difference scores. *Journal of Business and Psychology*, *25*(4), 543–554. https://doi.org/10.1007/s10869-010-9183-4

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, *48*(4), 813–838. https://doi.org/10.1037/0022-3514.48.4.813

Smither, J. W., Reilly, R. R., Millsap, R. E., AT&T, K. P., & Stoffey, R. W. (1993). Applicant reactions to selection procedures. *Personnel Psychology*, *46*(1), 49–76. https://doi.org/10.1111/j.1744-6570.1993.tb00867.x

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. https://doi.org/10.1016/S1053-4822(99)00019-4

So, J. (2013). A further extension of the extended parallel process model (E-EPPM): Implications of cognitive appraisal theory of emotion and dispositional coping style. *Health Communication*, *28*(1), 72–83. https://doi.org/10.1080/10410236.2012.708633

De Soete, B., Lievens, F., Oostrom, J., & Westerveld, L. (2013). Alternative predictors for dealing with the diversity-validity dilemma in personnel selection: The constructed response multimedia test. *International Journal of Selection and Assessment*, *21*(3), 239v250. https://doi.org/10.1111/ijsa.12034

Sun, T., Zhang, B., Cao, M., & Drasgow, F. (2022). Faking detection improved: Adopting a Likert item response process tree model. *Organizational Research Methods*, *25*(3), 490–512. https://doi.org/10.1177/10944281211002904

Wake, S., Wormwood, J., & Satpute, A. B. (2020). The influence of fear on risk taking: A meta-analysis. *Cognition and Emotion*, *34*(6), 1143–1159. https://doi.org/10.1080/02699931.2020.1731428

Wood, J. K., Anglim, J., & Horwood, S. (2022). Effect of job applicant faking and cognitive ability on self-other agreement and criterion validity of personality assessments. *International Journal of Selection and Assessment*, *30*(3), 378–391. https://doi.org/10.1111/ijsa.12382

Zhang, C., Cullen, M. J., & Sackett, P. R. (2021). Effects of response instructions on situational judgment test performance in operational selection and developmental contexts. *International Journal of Selection and Assessment*, *29*(2), 170–184. https://doi.org/10.1111/ijsa.12324

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Xi, L., Weng, Q., Corstjens, J., Wang, X., & Chen, L. (2024). Effects of a constructed response retest strategy on faking, test perceptions, and criterion-related validity of situational judgment tests. *International Journal of Selection and Assessment*, *32*, 561–578. https://doi.org/10.1111/ijsa.12482

APPENDIX

See Table A1.

TABLE A1 Standardized Tobit regression results in study 2.

	Model 1: Indirect effect of CR-retest strategy				Model 2: Indirect effect of warning message				Model 3: Interaction effect			
	Faking behavior				Faking behavior				Faking behavior			
	BER	IM	Fear	Threat	BER	IM	Fear	Threat	BER	IM	Fear	Threat
Gender	0.00	0.00	0.03	−0.03	0.00	−0.05	−0.05	0.06	−0.03	−0.01	−0.02	0.01
Age	0.06	−0.12	0.16*	−0.03	0.09	−0.08	0.13	0.03	0.06	−0.04	0.12*	−0.01
Degree	0.04	0.13	0.02	−0.04	0.03	0.08	0.06	−0.11*	0.03	0.10	0.03	−0.06
CR-retest strategy	−0.24**	−0.15	0.12	0.32***					−0.22*	−0.06	0.07	0.26***
Warning message					−0.07	−0.09	0.18**	0.25**	−0.08	−0.03	0.24**	0.22**
CR-retest strategy×warning message									−0.12	−0.13**	−0.30**	0.28**
Threat appraisal	−0.19*	−0.18*	0.25***		−0.14	−0.16*	0.39***		−0.30**	−0.42***	0.36***	
Fear	−0.08	−0.14			−0.25**	−0.31***			−0.11*	−0.14*		
R-square	0.15**	0.15**	0.12*	0.11*	0.14**	0.22**	0.24***	.08	0.37***	0.39***	0.15***	0.36***
N	*N* = 136				*N* = 136				*N* = 273			

Note. Gender: Male = 0, Female = 1. Degree: master's degree = 1, PhD degree = 2. Abbreviations: BER, Blatant extreme responding; IM, Impression management scale.

*$p < .05$; **$p < .01$; ***$p < .001$.

Copyright of International Journal of Selection & Assessment is the property of Wiley-Blackwell and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.