Freudenstein, Jan-Philipp; Schäpers, Philipp; Reznik, Nomi; Stolte, Talea; Krumm, Stefan

The influence of situational strength on the relation of personality and situational judgment test performance

International Journal of Selection and Assessment

Provided in Cooperation with:
John Wiley & Sons

Suggested Citation: Freudenstein, Jan-Philipp; Schäpers, Philipp; Reznik, Nomi; Stolte, Talea; Krumm, Stefan (2023) : The influence of situational strength on the relation of personality and situational judgment test performance, International Journal of Selection and Assessment, ISSN 1468-2389, Wiley, Hoboken, NJ, Vol. 32, Iss. 1, pp. 1-11, https://doi.org/10.1111/ijsa.12444

Received: 9 May 2022 | Revised: 6 April 2023 | Accepted: 22 June 2023

DOI: 10.1111/ijsa.12444

RESEARCH ARTICLE

International Journal of Selection and Assessment | WILEY

The influence of situational strength on the relation of personality and situational judgment test performance

Jan-Philipp Freudenstein[1,2] | Philipp Schäpers[3] | Nomi Reznik[1] | Talea Stolte[1] | Stefan Krumm[1]

[1]Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany

[2]Group R&D, Hogrefe Verlagsgruppe GmbH, Göttingen, Germany

[3]Institute of Psychology, University of Münster, Münster, Germany

Correspondence
Jan-Philipp Freudenstein, Group R&D, Hogrefe Verlagsgruppe GmbH, Merkelstraße 3, 37085 Göttingen, Germany.
Email: jpfreudenstein@gmail.com

Stefan Krumm.
Email: stefan.krumm@fu-berlin.de

Funding information
Deutsche Forschungsgemeinschaft

Abstract

Situational strength theory has been used as a theoretical underpinning of person–situation processes that are linked to job performance. Accordingly, the link between personality traits and job performance increases in weak situations. Building on this research, similar mechanisms have been proposed for simulation-based selection tools, such as Situational Judgment Tests (SJTs), to explain how these measures work as predictors of job performance. However, underlying processes of SJT performance are subject to debate with some scholars arguing in favor of context-independent processes while others maintain that situations play an essential role. This study ($N = 707$) examined whether the strength of situations in SJT items moderated the relation between personality and SJT performance. Results did not support the notion that personality is more strongly related to SJT performance when situations are weak. In fact, for some traits, the opposite may be true as more situational constraints led to an increase in the relation of extraversion, emotional stability, and SJT performance. The results add to an increasing body of research about psychological processes in SJTs. Limitations and implications for research and practice are discussed.

KEYWORDS

personality, selection, situational judgment tests, situational strength

Practitioner points

- Prior research has shown that situational strength moderates the relation of personality and job performance.
- In the current study, we found that Situational Judgment Tests (SJTs) do not fully mimic this aspect of real-life situations.
- Thus, the debate about the underlying psychological processes of SJT performance is not yet resolved.
- More theory-driven SJT developments are needed that provide verifiable assumptions about underlying psychological processes.

1 | INTRODUCTION

In line with general assumptions about underlying processes of individual behavior, research on person–situation interactions contributed to an increased understanding of job performance (e.g., Judge & Zapata, 2015; Meyer et al., 2009; Tett & Burnett, 2003). In particular, situational strength theory offered insights into the relation between personality and behavior at work: Strong situations will lead to more homogeneous behavior across individuals, which mitigates the link between personality and job performance (e.g., Judge & Zapata, 2015; Meyer et al., 2009). Accordingly, similar underlying processes have been incorporated into the development of personnel selection tools, such as assessment centers or Situational Judgment Tests (SJTs; e.g., Lievens et al., 2009; Melchers et al., 2012; Schollaert & Lievens, 2012; Weekley et al., 2015).

SJTs are popular selection tools that rest on the principle of person–situation interaction and behavioral consistency (Lievens & De Soete, 2012; Motowidlo et al., 1990; Weekley et al., 2015). That is, SJTs typically comprise various work-related situation descriptions and several short behavioral response options (McDaniel & Nguyen, 2001). Despite recent efforts to include an interactionist perspective into SJT theory (e.g., Campion & Ployhart, 2013; Harris et al., 2016; Martin-Raugh & Kell, 2021), the underlying processes of SJT performance remain subject to debate. Recent studies demonstrated that the situation in SJTs may often be less relevant for SJT performance than conceptualized (Jackson et al., 2017; Krumm et al., 2015; Rockstuhl et al., 2015; Schäpers, Freudenstein, et al., 2020; Schäpers, Lievens, et al., 2020; Schäpers, Mussel, et al., 2020). However, other studies upheld the view of person–situation interactions as underlying processes of SJTs (Freudenstein et al., 2020; Lievens et al., 2018; Westring et al., 2009). For example, Freudenstein et al. demonstrated that an individual's perception of situations in SJT items is relevant for response processes. The results further showed that situation descriptions and response options jointly constitute psychologically relevant situations and that stripping off situation descriptions does not transform SJTs into completely context-independent measures (see also Harris et al., 2016; Melchers & Kleinmann, 2016).

Thus far, previous research did not provide conclusive evidence on which processes determine SJT performance and ultimately explain why SJTs predict job performance. Specifically, the situational strength of SJT items has—to the best of our knowledge—not yet been empirically tested as a relevant situational influence on SJT performance. This is despite previous research identifying situational strength as a relevant process with regard to job performance (Judge & Zapata, 2015; Meyer et al., 2010; Tett & Burnett, 2003). In the current research, we further shed light on the person–situation interplay that underlies SJT performance by examining whether the situational strength of SJT items moderates the relation between personality and SJT performance (see Harris et al., 2016). By doing so, we contribute to an understanding of SJTs' functioning and why SJTs predict relevant criteria.

2 | THEORETICAL BACKGROUND

One inference from the person–situation debate was that behavior is driven by both persons and situations (e.g., Fleeson & Noftle, 2008). Situational strength represents one type of such situational influences on behavior (e.g., Dalal et al., 2014; Judge & Zapata, 2015; Meyer et al., 2009, 2010; Mischel, 1973, 1977). Situational strength is defined "as implicit or explicit cues provided by external entities regarding the desirability of potential behaviors" (Meyer et al., 2010, p. 122). Accordingly, stronger situations should lead to more similar perceptions of situations and thus more similar behavior (Meyer et al., 2009; Mischel, 1977). To clarify the conceptual framework of situational strength, Meyer et al. (2010) proposed four facets, namely, clarity, consistency, constraints, and consequences. Each facet describes a group of situational cues that restrict the range of possible behaviors (i.e., clarity of responsibilities, consistency of different situational demands, constraints to behavior, and consequences of behavior). For instance, "clear instructions and support from one's supervisor" should increase the awareness of the expected behavior of all employees (Meyer et al., 2010, p. 125).

It is assumed that with increasing situational strength, the situation becomes more relevant as a determinant of behavior, which is accompanied by a decrease in the relevance of personality as a predictor of behavior (Meyer et al., 2009). Several studies supported this role of situational strength as a negative moderator on the relation between personality traits and job performance (Judge & Zapata, 2015; Lee & Dalal, 2016; Meyer et al., 2009). Meyer et al. demonstrated in their meta-analysis that conscientiousness exhibited smaller correlations with performance criteria in occupations with stronger situations. Accordingly, research adopted the concept of situational strength to simulation-based tools in personnel selection (e.g., assessment center; Christiansen et al., 2013; Herde & Lievens, 2020; Lievens et al., 2009, 2015; Melchers et al., 2012; Oliver et al., 2016; Schollaert & Lievens, 2012). From this, it follows that simulation-based selection measures should refrain from incorporating strong situations to maximize the assessment of relevant traits or dimensions that predict job performance (e.g., Lievens et al., 2009; Melchers et al., 2012).

In line with this proposition, Harris et al. (2016) argued that the situational strength of SJT items—and specifically the clarity of SJT items—should moderate the relation between personality and SJT performance (see also, Campion & Ployhart, 2013; Martin-Raugh & Kell, 2021; Rockstuhl & Lievens, 2021). That is, if no clear directive for appropriate behavior is given in ambiguous situations, test-takers would rely completely on their trait dispositions to respond to SJT items (Harris et al., 2016). In fact, personality traits have been demonstrated as one of the main antecedents of SJT performance (see McDaniel et al., 2007). However, this relation may vary depending on the situational strength of situations in SJT items. That is, SJT items that represent strong situations may show a lower correlation with personality traits than SJT items that represent weaker situations. This assumption relies on a conceptualization of SJT items as simulations of work-related scenarios, similar to assessment center exercises, in which situation descriptions are essential for underlying psychological processes (e.g., McDaniel & Nguyen, 2001; Motowidlo et al., 1990; Weekley et al., 2015).

Despite these arguments, the situational influences on SJT performance are subject to debate. Lievens and Motowidlo (2016) reconceptualized SJTs as measures of general domain knowledge. They argued that responses to SJTs are not situation-specific and that test-takers use general beliefs about the utility of trait-related behaviors to respond to SJT items. This reconceptualization of SJTs specifically builds on results that showed, for a large number of SJT items, no differences in item difficulty when situation descriptions were omitted (Krumm et al., 2015). Krumm and colleagues argued that situation descriptions may be less relevant for SJT performance than conceptually defined and that some other processes may take place. In fact, a series of recent findings further showed no or only very little relevance of situation descriptions for the construct-related validity of SJTs (Schäpers, Freudenstein, et al., 2020; Schäpers, Mussel, et al., 2020) and similar effects on item difficulties for a video-based SJT (Schäpers, Lievens, et al., 2020).

Two recent studies took a closer look at key positions of the debate about the underlying processes of SJT performance. First, Freudenstein et al. (2023) demonstrated that typical operationalizations of SJTs are not in line with the conceptualization as measures of general domain knowledge. Specifically, results showed that reliable variance in SJT scores is to a larger degree test-specific rather than context-independent, even when SJTs measured the same construct. Similar conclusions may also be drawn from studies identifying that even within SJTs, large variance components are item-(or situation-)specific (Catano et al., 2012; Westring et al., 2009). Second, within the debate, complex relations between situation cues in SJT items and psychologically relevant situations may have been neglected (Brown et al., 2016; Freudenstein et al., 2020). Freudenstein et al. demonstrated that situation construal of SJT items predicted response behavior in SJT items regardless of whether the situation description was presented or not. They concluded that response options contain sufficient situation cues to construe psychologically relevant situations (see also Harris et al., 2016; Melchers & Kleinmann, 2016). Overall, these results speak in favor of an interactionist perspective of persons and situations in SJT items similar to psychological processes in other simulation-based selection methods (e.g., assessment centers; Jansen et al., 2013).

When considering SJTs as simulation-based selection methods, these tests should specifically reflect situational processes that have been demonstrated to be relevant in real-life job situations. Therefore, assessing the relevance of SJT items' situational strength for the relation between personality traits and SJT performance is needed to further understand person–situation processes that underly SJT test scores. Such insights would contribute to an understanding of why SJTs predict job performance criteria (see Christian et al., 2010; McDaniel & Morgeson, Finnegan, et al., 2001, 2007). However, to the best of our knowledge, the influence of situational strength on the relation between personality and SJT performance has not been directly tested. In this study, we do so by assessing the situational strength of several SJT items. Given the current debate on the relevance of situation cues for SJT responses (e.g., Freudenstein et al., 2020; Krumm et al., 2015; Lievens & Motowidlo, 2016), we refrain from postulating a specific hypothesis. Instead, we pose the following research question:

RQ: *Does the situational strength of SJT items negatively moderate the relation of broad personality traits and SJT responses?*

3 | METHODS

In this study, we reanalyzed three data sets that were previously reported by Freudenstein et al. (2020; data are available on the Open Science Framework; osf.io/6kd9h) and Schäpers, Freudenstein et al. (2020; data are available on the Open Science Framework; osf.io/zdsqt). These data contain four different work-related SJTs with a total of 44 SJT items and self-reported Big Five personality. Each SJT item contained a job-related situation description and four to five response options. Depending on the SJT, the subjects were asked to assess how they should behave (knowledge instruction) or how they would behave (behavioral-tendency instruction). For some studies, SJT items were experimentally manipulated between subjects (e.g., omitting situation descriptions; Schäpers, Mussel, et al., 2020). Hence, we only included data from participants who completed the unmanipulated versions of the SJT items. We further considered participants eligible for this study if complete personality data were available. Importantly, we directly assessed the situational strength of SJT items in addition to these data. This was done by asking subject matter experts to rate the situational strength of all SJT items. We decided to assess situational strength as subject matter expert ratings as the consistent variance component across raters (i.e., average score) is in line with the definition of situational strength as an attribute of the objective situation (see Meyer et al., 2010). Data and code for this study are available on the Open Science Framework: https://osf.io/hck3j/

3.1 | Sample

Overall, 718 participants from the previous studies were eligible for this study ($n_1 = 104$; $n_2 = 315$; see Schäpers, Mussel, et al., 2020; $n_3 = 299$; see Freudenstein et al., 2020). We tested the data for careless responding by computing Mahalanobis distances for the self-reported personality data (Meade & Craig, 2012). On the basis of an $\alpha = .001$ criterion, we excluded $n = 11$ participants from further analyses. Thus, we analyzed a total sample of $N = 707$ ($n_1 = 101$; $n_2 = 313$; $n_3 = 293$; 451 female). On average, the sample was 32.87 ($SD = 13.38$; range: 18–78) years old. For detailed descriptions of sample characteristics and data collection see Schäpers, Mussel et al. (2020) and Freudenstein et al. (2020).

3.2 | Measures

3.2.1 | Team role test (TRT)

The TRT (Mumford et al., 2008) is a 10-item SJT that assesses knowledge about suitable team roles in specific situations. The data set provided by Schäpers, Mussel et al. (2020) included a total of 313 participants, who responded to a modified version of the TRT, which

comprised four response options instead of 10 for each item. Test-takers were asked what they should do in each situation. The test was administered in a pick-the-best format. The most effective response option for each situation was scored as "1." All other response options were scored as "−1." The reliability of this test was low but within the meta-analytical range of SJT's reliability (ω = .27, α = .41; Catano et al., 2012; Kasten & Freund, 2016).

3.2.2 | Situational judgment questionnaire (SJQ)

The SJQ (Motowidlo et al., 2006) consists of 22 items with four response options. The test assesses work-related behavior in the presence of other people, such as supervisors or coworkers. In this SJT, all response options are designed to express agreeableness (Motowidlo et al., 2006). The data set by Schäpers, Mussel et al. (2020) comprised 10 items of this SJT, which asked participants how they should behave in each situation (n = 313). The test was administered in a pick-the-best format. Effective response options were scored as "1" and ineffective response options were scored as "−1." Reliability of this SJT was ω = .42 and α = .59.

3.2.3 | Personal initiative SJT

The Personal Initiative SJT (PI-SJT; Bledow & Frese, 2009) consists of 12 situation descriptions with four to five response options. It assesses personal initiative in work-related settings. This SJT was applied to two out of the three samples (n = 394). Schäpers, Mussel et al. (2020) asked participants how they would behave in each situation (i.e., pick the most likely response option). However, Freudenstein et al. (2020) additionally asked participants how they would not act in each situation (i.e., pick the most likely and the least likely response options). For consistency, only responses to the question "what would you do" were considered for these analyses. Effective response options were scored as "1," ineffective response options as "−1," and all remaining response options as "0." Reliability of this SJT was ω = .62 and α = .68. As we used data from two samples for this SJT, we tested for measurement invariance between the two samples. The general factor model had good model fit, $\chi^2(54) = 69.393$, p = .077; CFI = .957; RMSEA = .027; SRMR = .060 (see Schweizer, 2010). Furthermore, factor loadings could be restrained to equality between samples without a decrease in model fit (i.e., metric invariance), $\Delta\chi^2(11) = 6.022$, p = .872; ΔCFI = −.021; ΔRMSEA = .006.

3.2.4 | Teamwork SJT

The Teamwork SJT (TW-SJT Gatzka & Volmer, 2017) measures effective behavior in teamwork situations. It consists of 12 situation descriptions with four response options. Depending on the situation descriptions, participants (n = 104; Schäpers, Mussel, et al., 2020) were asked how they should behave or what their team should do. Effective response options were scored as "1," ineffective response options as "−1," and all remaining response options as "0." Reliability of this SJT was ω = .53 and α = .62.

3.2.5 | Self-reported personality

All 718 participants responded to the German short version of the Big Five Inventory (BFI-K; Rammstedt & John, 2005). This inventory consists of 21 items assessing the broad personality traits emotional stability, extraversion, openness, agreeableness, and conscientiousness. Participants were asked to indicate on a five-point rating scale (1 = disagree strongly; 5 = agree strongly) whether each item described themselves appropriately. A correlated five factor model did not yield acceptable fit (see Schweizer, 2010); $\chi^2(179) = 843.850$, p < .001; CFI = .851; RMSEA = .072; SRMR = .074. We used Ant-Colony-Optimization (Olaru et al., 2015; Schultze, 2017) to develop a well-fitting five factor model with three indicators for each Big Five trait. The model yielded acceptable fit: $\chi^2(80) = 264.641$, p < .001; CFI = .938; RMSEA = .057; SRMR = .051. This scale also yielded metric measurement invariance (i.e., identical factor loadings) across the three different samples: $\Delta\chi^2(20) = 23.901$, p = .247; ΔCFI = .001; ΔRMSEA = .002. Reliabilities for the five factors ranged from ω = .66 to .82 and α = .65 to .82 (see Table 1). The average correlation between manifest mean scores of the short scale and the BFI-K was .96 (range: .92–.98).

3.2.6 | Situational strength

Two authors of this study and one research assistant with particular expertise in SJT research independently evaluated the situational strength of all 44 SJT items. To do so, we used three items with the highest item-total correlation of each factor from the job-related situational strength questionnaire (Meyer et al., 2014). This measurement comprises four factors, namely, clarity (e.g., "specific information about work-related responsibilities is provided"), consistency ("different sources of work information are always consistent with each other"), constraints (e.g., "procedures prevent an employee from working in his/her own way"), and consequences (e.g., "mistakes are more harmful than they are for almost all other situations"). Since SJT items typically do not contain enough situational context to assess the situation's consistency, we assessed situational strength only on the remaining three scales. For example, when an SJT is used for personnel selection, a test-taker cannot judge whether a situation from the test is consistent with other typical situations at the job as the specific work environment is unknown to the applicant. Importantly, we instructed raters to take not only the situation description but also response options into consideration, as previous research suggested that relevant situation cues may also be present in response options (Freudenstein et al., 2020; Harris et al., 2016; Melchers & Kleinmann, 2016). Internal consistency for the three factors ranged from α = .84 to .90. We computed mean scores for each factor within raters. Initial interitem correlation coefficient (ICC3,k) for these scores ranged from .63 to .79 thus indicating moderate to strong interrater agreement (LeBreton & Senter, 2008). Hence, we collapsed ratings from

TABLE 1 Descriptive statistics, bivariate correlations, and internal consistencies.

	M (*SD*)	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. ES	3.14 (0.99)	*(.78)*								
2. E	3.49 (0.94)	.26*	*(.82)*							
3. O	3.88 (0.95)	−.02	.11*	*(.82)*						
4. A	3.08 (0.91)	.21*	.16*	.06	*(.70)*					
5. C	3.98 (0.66)	.18*	.10*	−.01	.00	*(.66)*				
6. PI	1.40 (3.74)	.18*	.23*	.16*	.07	.14*	*(.62)*			
7. TRT	1.75 (3.07)	−.01	.05	.05	.11	.11*	–	*(.27)*		
8. TW	5.23 (2.88)	.22*	.00	−.03	.13	.17	.40*	–	*(.53)*	
9. SJQ	3.64 (3.35)	.02	.03	.01	.08	.04	–	.31*	–	*(.42)*

Notes: $n = 87–707$. Coefficient omega in parentheses on diagonal. As not all participants responded to all SJTs, some bivariate correlations among SJT scores could not be computed.

Abbreviations: A, agreeableness; C, conscientiousness; E, extraversion; ES, emotional stability; O, openness; PI, personal initiative; SJQ, situational judgment questionnaire; SJT, Situational Judgment Test; TRT, team role test; TW, teamwork.

*$p < .05$.

Your team coordinates its work in meetings on a regular basis. However, again and again, some of your colleagues do not stick to the decisions. This makes your work difficult. However, your colleagues insist on being able to be flexible and to sometimes deviate from the agreements.

What would you do?

A. I constantly get information about the work of my colleagues to avoid having problems with my work.
B. I won't let the behavior of my colleagues bother me.
C. I ask my colleagues to sit down with me to think of a long-term solution.
D. I accept the wish of my colleagues to be flexible. However, I make it clear that I am not responsible for the problems, I have in my work.

FIGURE 1 Example SJT item. *Notes*: This item was taken from the Personal Initiative SJT (Bledow & Frese, 2009). The average rating for the facet Clarity of this item was $M = 6.11$, which reflects the explicit information that doing one's job is difficult due to colleagues not sticking to decisions. The average rating for the facet Constraints was $M = 5.67$. This score reflects the test-takers constraint to do their work due to colleagues insisting on being flexible. The average rating for the facet Consequences was $M = 3.33$. In this situation, behavior has direct consequences for the task or team performance. However, no information is available that suggests that the consequences in this situation are more severe than in other situations. SJT, Situational Judgment Test.

all three raters. Figure 1 depicts an example SJT item and explanations of the average situational strength ratings.

3.3 | Data analyses

To test our research question, we examined an interaction effect of situational strength and personality on SJT responses. To analyze all data simultaneously, we combined all three data sets and fitted ordinal mixed-effects models with crossed random effects (Baayen et al., 2008; Tutz & Hennevogl, 1996) using the R package *ordinal* (Christensen, 2018). This combination of data was possible, as missing data only occurred on the dependent variable (i.e., SJT responses), which should not lead to biases in regression coefficients (Little, 1992). We included random intercepts for SJT items and persons to appropriately account for systematic variance components in SJT responses due to individuals and item content. We fitted separate models for each Big Five trait by stepwise including personality as a person-level predictor of SJT response, random slopes for personality across SJT items, the three facets of situational strength as an item-level predictor, and the interaction of the item-level and person-level predictors. Following guidelines by Enders and Tofighi (2007), we scaled situational strength variables within SJT items and personality variables within individuals.

4 | RESULTS

Descriptive statistics of SJT performance and personality are depicted in Table 1. The Shapiro–Wilk test for normality showed that all situational strength facets were approximately normally distributed ($W = 0.976$, $p = .470$). Importantly, situational strength in SJT items ranged from weak to strong. That is, we found a large range of situational strength values for all three facets (see Table 2; $range_{Clarity}$, 1.78–6.67; $range_{Constraints}$, 1.78–6.00; $range_{Consequences}$, 1.33–6.22), thus enabling us to observe possible moderating effects. Moreover, clarity, as well as consequences, were positively correlated with SJT item difficulty (*rs* = .15 and .24) and with the relative frequency of the most chosen (but not necessarily the correct) response option (*rs* = .19 and .24). This indicates that the situational strength had a restricting effect on individual responses in that SJT items became easier with increasing situational strength. However, due to the small sample of items ($n = 44$), no statistical significance was reached (*ps* = .110–.322). Conversely, constraints were negatively correlated with SJT item difficulty and the relative frequency of the most chosen response option (*rs* = −.10 and −.13, *ps* = .534 and .402), indicating a slight tendency that SJT items with more situational constraints were more difficult.

In line with previous attempts to explain variance in SJT scores (see Jackson et al., 2017), we conducted a variance decomposition based on generalizability theory (Brennan, 2001).[1] Our research design enabled us to differentiate between explained variance in SJT scores due to individuals (10%), individual–construct interactions (3%), situations (49%), constructs assessed with the SJT (<.01%), and unexplained variance (38%).

In the next step, we fitted the mixed-effects models. We found that all Big Five traits significantly predicted SJT performance, although effect sizes were small (*Bs* = .04–.05, *ps* = .009–.041; see Table 3). When we included random slopes in the model, the fit only increased for conscientiousness. Thus, the prediction of SJT performance across items differed only for conscientiousness: $B = .05$; $SD = 0.07$; $\Delta\chi^2(2) = 7.522$, $p = .023$. The inclusion of clarity, constraints, and consequences as fixed effects did not improve model fit. Finally, tests of interactions between facets of situational strength and personality produced mixed findings. For models with extraversion, openness, agreeableness, and conscientiousness as predictors, likelihood-ratio tests indicated that the inclusion of interaction terms between personality and situational strength did not improve model fits: $\Delta\chi^2(3) = 1.192–6.156$, *ps* = .104–.755. However, the inclusion of interaction terms did improve model fit for the emotional stability model: $\Delta\chi^2(3) = 8.295$, $p = .040$. In this model, the facet constraints interacted positively with emotional stability ($B = .04$, $p = .015$). In addition, the facet constraints significantly interacted with extraversion in the prediction of SJT performance ($B = .04$, $p = .003$), although overall model fit did not improve. These interaction terms indicate that more situational constraints led to an increased relation between emotional stability and extraversion with SJT performance.[2]

Finally, we took a closer look at the difference between SJT items with knowledge or behavioral-tendency instruction. The link between personality and SJT performance has been demonstrated to be more pronounced for behavioral-tendency instructions (McDaniel et al., 2007), so these SJTs may closer resemble simulation-based assessments whereas SJTs with knowledge instruction may rather reflect measures of general domain knowledge. We repeated all analyses with data subsets including only SJT items with knowledge or behavioral-tendency instruction. For SJT items with knowledge instruction, we found no direct effect for the relation of personality and SJT performance (*Bs* = .01–.05, *ps* = .053–.735) except for agreeableness ($B = .05$, $p = .019$). The situational strength facet constraints significantly interacted with emotional stability ($B = .04$, $p = .028$) but not with extraversion ($B = .03$, $p = .123$). Similar to the

TABLE 2 Descriptive statistics, bivariate correlations, and internal consistencies of situational strength variables.

	M (*SD*)	1.	2.	3.
1. Clarity	4.65 (1.14)	*(.63)*		
2. Constraints	4.10 (1.21)	.26	*(.65)*	
3. Consequences	3.12 (1.15)	.56*	.11	*(.79)*

Notes: $n = 44$. Intraclass correlation coefficients in parentheses on diagonal.
*$p < .05$.

TABLE 3 Summary of fixed effects from ordinal mixed-effects models.

	ES		E		O		A		C	
	B	*p*	*B*	*p*	*B*	*p*	*B*	*p*	*B*	*p*
Trait	.04	.041*	.05	.013*	.04	.016*	.05	.009*	.05	.021*
Clarity	−.06	.633	−.06	.628	−.06	.630	−.06	.639	−.06	.632
Constraints	−.10	.301	−.10	.331	−.10	.309	−.10	.308	−.10	.306
Consequences	.12	.272	.12	.282	.12	.280	.12	.275	.12	.274
Trait:Clarity	−.02	.245	−.00	.792	.01	.423	−.02	.108	.01	.460
Trait:Constraints	.04	.015*	.04	.003*	.00	.853	.00	.866	.02	.201
Trait:Consequences	.01	.519	−.01	.742	.00	.934	.02	.296	−.01	.682

Abbreviations: ES, emotional stability; E, extraversion; O, openness; A, agreeableness; C, conscientiousness; B, probit regression weight.
*$p < .05$.

analyses with complete data, we found no other significant interactions between situational strength and personality. For SJT items with behavioral-tendency instruction, we found direct links between SJT performance and emotional stability ($B = .09$, $p = .031$), extraversion ($B = .11$, $p = .029$), and openness ($B = .08$, $p = .017$) but not for agreeableness and conscientiousness ($Bs = .02–.04$, $ps = .232–.592$). We further found no significant interaction between personality traits and any situational strength facet.

5 | DISCUSSION

This study sought to examine the relevance of SJT items' situational strength as a moderator on the relation between personality and SJT performance. First, we found that all Big Five traits significantly predicted SJT performance on the item level. On the level of test scores, a more heterogeneous correlation pattern with the Big Five traits emerged. Overall, the results are consistent with meta-analytical findings revealing personality is related to SJT performance (McDaniel et al., 2007). In accordance with these meta-analytical results, personality was stronger related to SJT performance for items with behavioral-tendency instructions—at least for three out of the five tested traits. Second, the situational strength of SJT items had no significant direct effect on SJT performance. That is, test-takers did not score significantly higher on items that were high in situational strength, although our analyses suggest a tendency for a decreased variability in responses and increased item difficulty for "stronger" SJT items. Third, we found no moderating effect for two out of three situational strength facets on the relationship between personality traits and SJT item responses. Notably, we did not find a moderated link between conscientiousness and SJT performance. This is surprising as effects of situational strength are well established for the link between conscientiousness and job performance (e.g., Meyer et al., 2009), as well as conscientiousness being the most relevant antecedent of SJT performance (McDaniel et al., 2007). Neither did we find moderation effects for the situational strength facet clarity. Clarity has been proposed as the most relevant situational strength facet in the context of SJTs (Harris et al., 2016). In sum, we conclude that the situational strength of SJT items does not moderate the relationship between personality and SJT performance—at least in the way theory would suggest.

Overall, these results support the emerging notion of SJTs as context-independent measures (e.g., Krumm et al., 2015; Lievens & Motowidlo, 2016). Most relevant to our results, situation descriptions did not influence the correlation between SJT responses and personality traits (Schäpers, Freudenstein, et al., 2020; Schäpers, Mussel, et al., 2020). The authors argued that situation descriptions in SJT items may only contribute little to situational processes in SJT items and that response options may be sufficient to evoke trait-activating processes. For example, Grand (2020) argued that weighing response options against each other is one of the most relevant response processes for SJT items. Due to the limited availability of response options—typically four in most SJTs—test-takers are restricted in their response choice. It is likely that no response option will exactly resemble the behavior a test-taker may prefer. This may lead to a loss of information as the response score to an SJT cannot differentiate between fine-grained differences in behavior that would otherwise occur in an open response format or a real-life situation. However, these fine-grained differences in response behavior may be needed to observe interactions between personality and situational strength that predict performance. This argument is supported by the typically small effects observed for interaction of personality and situational strength as predictors of job performance (e.g., Meyer et al., 2014). It is important to note, that other response formats exist for SJTs and that future research should carefully examine differences in response processes due to different response formats.

However, we did find that the facet constraints positively moderated the relation of emotional stability and extraversion with SJT performance. Surprisingly, these effects were in the opposite direction of situational strength theory as stronger situations (i.e., more situational constraints) were associated with a stronger link between personality and SJT performance. Even more surprisingly, these interaction effects did not occur for SJT items with would-do instructions, although these items should more closely evoke trait-related processes whereas SJTs with knowledge instruction may tap more closely into the domain of general domain knowledge and general mental ability (Lievens & Motowidlo, 2016; McDaniel et al., 2007). In addition, our results showed a tendency that more situational constraints in SJT items did not reduce variability in response behavior although this reflects an essential condition of situational strength theory (see Keeler et al., 2019). Such mixed effects with regard to the influence of situational strength on the response variability among individuals are not unique to this study. For instance, Meyer et al. (2014) found that the negative link between conscientiousness and counterproductive work behavior was more pronounced in strong situations. These authors attributed their findings to complex processes of how individuals perceive situational strength as psychologically relevant.

Psychologically relevant perceptions of situations have also been shown to predict SJT performance (Freudenstein et al., 2020). This speaks for a situation-dependent view on SJTs, which—in the current study—was only supported by two significant moderation effects. Notably, Freudenstein et al. (2020) also found that relevant situation perceptions can be evoked by response options and that situation descriptions are negligible for these processes. Overall, the format of fixed behavioral response options in SJT items may hinder the full potential of underlying situational processes of SJT responses. This may also be true for situational strength, as behavioral response options in SJT items may not reflect a broad range of trait-related behaviors and, as a result, test-takers may be unable to respond consistently with their personality for weak or moderately strong situations. Moreover, behavioral response options may not always reflect those behaviors that are demanded by stronger situations so variability in responses may emerge even though situations per se are strong.

Given these results, more research is needed to further examine the processes of person–situation interactions underlying SJT performance. Situational strength has been particularly useful for an understanding of when and how personality predicts job

performance (e.g., Meyer et al., 2009). These processes should be taken into consideration when developing SJTs and other tools for personnel selection. However, different approaches are needed to enhance our understanding of situational processes in simulation-based assessments. Experimental test validation (Krumm et al., 2019) may be particularly useful to examine specific processes. For instance, researchers may design specific situation descriptions that vary in situational strength and align response options to reflect trait-related behavior that corresponds with the situation descriptions. In addition, recent research on situation taxonomies may be useful to understand how short situation descriptions may evoke meaningful perceptions (e.g., Rauthmann et al., 2014; Saucier et al., 2007; see also Lievens, 2017a). In this regard, we recommend investigating the moderating role of situational strength on the validity of personality traits for predicting external criteria (e.g., job performance) via SJTs (see Meyer et al., 2009). Schäpers, Mussel et al. (2020) found only preliminary evidence that criterion-related outcomes (e.g., job performance) are influenced by the availability of situational descriptions, and our findings provide further support for the context-independent view of SJTs. Future research should focus on principles for developing SJT items that may help improve the situation-dependency of SJT performance to increase the similarity of these processes to real-life processes that explain job performance.

Furthermore, we examined the interaction between personality and situational strength only for low-fidelity simulations (i.e., text-based SJTs). In terms of the distinction between constructs and methods when comparing predictors (see Arthur & Villado, 2008), we recommend to extend the examination of interactions between personality and situational strength to other simulations (e.g., high-fidelity simulations, like, assessment centers or video-based SJTs) and compare them along the lines of a multitrait–multimethod matrix.

Given these uncertainties regarding the relevance of person–situation interactions for SJT performance, we urge practitioners to rely on construct-driven SJTs (Guenole et al., 2017; Lievens, 2017b). Although the underlying principles with regard to the role of situation descriptions do not differ from traditional SJTs (Schäpers, Freudenstein, et al., 2020), these tests validly assess unidimensional constructs (e.g., Mussel et al., 2018; Olaru et al., 2019; Oostrom et al., 2018). Situation descriptions in construct-driven SJTs may function as a highly specific frame-of-reference, which may enhance the criterion-related validity (see Shaffer & Postlethwaite, 2012).

6 | LIMITATIONS

We assessed situational strength on the item level. This limits the generalizability of our results in two ways. First, we analyzed only 44 SJT items. This number is slightly lower than the typical recommendation of at least 50 level two units for unbiased estimates of standard errors (see Maas & Hox, 2005). This particularly limits the generalizability of our findings, and future research should examine larger numbers of SJTs and SJT items. Second, our results are based upon the assumption that an objective situational strength of SJT items exists. Definitions of situational strength propose that existing cues of situations determine whether certain behaviors are desirable (see Meyer et al., 2010) and that these objective situation entities may determine whether individuals are constrained in their behaviors (see Judge & Zapata, 2015). This perspective is specifically relevant in the context of SJT development as typically only general—and in that sense—objective aspects of situations descriptions are considered. Nevertheless, situational strength may also be understood as test-takers' situation construal (see Meyer et al., 2014). That is, the individual perception of situational strength may matter to influence behavior. Specifically, an individual may perceive strong constraints in a situation that determines their behavior although some agreement among test-takers or test-developers exists that the given situation contains only a few constraints. This perspective may also explain the differences we observed in ratings of situation strength. For example, interrater reliabilities were mostly moderate which leaves room for more agreement between raters. We also found differences on average clarity ratings between raters that were blind versus nonblind to the research question and slight differences in results of the mixed-effects model. Future research should incorporate this perspective on situational strength and ask individuals about their perception of situational strength in SJT items. In addition, future research may link situational strength to the low reliability of SJTs (e.g., Catano et al., 2012; Kasten & Freund, 2016) to examine whether intraindividual variability in perceptions of situational strength and SJT responses is meaningful or due to a lack of measurement precision within SJT items.

We also specifically tested work-related SJTs. These SJTs are designed to reflect multidimensional behavioral tendencies rather than specific constructs (McDaniel et al., 2007, 2016; Weekley et al., 2015). Accordingly, we assumed that all SJT items had at least some relevance for a broad range of personality traits. Although we found significant effects for all personality traits, nuances in the trait-activating potential of SJT items may exist (see Tett & Burnett, 2003). For instance, we found significant random slopes for the prediction of SJT performance by conscientiousness. This may reflect that some SJT items were more trait-activating for conscientiousness than others. Moreover, fixed effects of personality traits on SJT performance were rather small, which may further add to the notion that not all items were equally relevant to all personality traits. This may also be due to the fact that we tested broad personality traits, although some SJTs were designed to reflect specific personality facets (e.g., proactivity for the PI-SJT; Bledow & Frese, 2009) Therefore, we propose that further examinations of situational strength in the context of SJTs specifically take the trait-activation potential into account. Construct-driven SJTs may pose as a good starting point for such an undertaking (see Guenole et al., 2017).

7 | CONCLUSION

In this study, we built on situational strength theory to examine whether SJT performance reflected person–situation processes. Similar to the influence of situational strength on the relation of conscientiousness and job performance, we argued that situations in SJT items may have the

same underlying mechanism (Harris et al., 2016). However, our results demonstrated that this may not be the case. This study shows that the debate about the underlying psychological processes of SJT performance is not yet resolved. Whereas some person–situation processes may be relevant to SJT responses (i.e., situation construal; Freudenstein et al., 2020), others (i.e., situational strength) may not. Overall, we call for more theory-driven SJT developments that provide clear and verifiable assumptions about underlying psychological processes (see Guenole et al., 2017).

ACKNOWLEDGMENT

This research was part of a grant funded by the German Research Foundation (DFG; Grant KR-3457/2-1). Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data supporting the results of this study are openly available in the Open Science Framework at https://osf.io/hck3j/.

ORCID

Jan-Philipp Freudenstein https://orcid.org/0000-0002-9029-5003
Philipp Schäpers http://orcid.org/0000-0002-8270-5105
Nomi Reznik https://orcid.org/0000-0001-7294-5431
Stefan Krumm https://orcid.org/0000-0002-0840-0864

ENDNOTES

[1] This analysis was suggested by an anonymous reviewer.

[2] We followed the suggestion of an anonymous reviewer and included sensitivity analyses to differentiate between raters of situation strength that were blind to the research question and those who were not. Results differed only slightly in that we found no interaction effect between emotional stability and constraints and a significant fixed effect for constraints when only ratings of the blind rater were included. On a descriptive level, mean score ratings for clarity differed between the two groups of raters, $t(85.046) = -3.445$, $p < .001$, $d = -.73$

REFERENCES

Arthur, Jr., W., & Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology*, *93*(2), 435–442. https://doi.org/10.1037/0021-9010.93.2.435

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bledow, R., & Frese, M. (2009). A situational judgment test of personal initiative and its relationship to performance. *Personnel Psychology*, *62*(2), 229–258. https://doi.org/10.1111/j.1744-6570.2009.01137.x

Brennan, R. L. (2001). *Generalizability theory*. Springer.

Brown, N. A., Jones, A. B., Serfass, D. G., & Sherman, R. A. (2016). Reinvigorating the concept of a situation in situational judgment tests. *Industrial and Organizational Psychology*, *9*(1), 38–42. https://doi.org/10.1017/iop.2015.113

Campion, M. C., & Ployhart, R. E. (2013). Assessing personality with situational judgment measures. In N. D. Christiansen, & R. P. Tett (Eds.), *Handbook of personality at work* (pp. 439–456). Routledge. https://doi.org/10.4324/9780203526910.ch19

Catano, V. M., Brochu, A., & Lamerson, C. D. (2012). Assessing the reliability of situational judgment tests used in high-stakes situations. *International Journal of Selection and Assessment*, *20*(3), 333–346. https://doi.org/10.1111/j.1468-2389.2012.00604.x

Christensen, R. H. B. (2018). *Cumulative link models for ordinal regression with the R package ordinal*. Submitted for publication. https://cran.r-project.org/web/packages/ordinal/vignettes/clm_article.pdf

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

Christiansen, N. D., Hoffman, B. J., Lievens, F., & Speer, A. B. (2013). Assessment centers and the measurement of personality. In N. D. Christiansen, & R. P. Tett (Eds.), *Handbook of personality at work* (pp. 477–497). Routledge. https://doi.org/10.13140/2.1.3105.9847

Dalal, R. S., Meyer, R. D., Bradshaw, R. P., Green, J. P., Kelly, E. D., & Zhu, M. (2014). Personality strength and situational influences on behavior. *Journal of Management*, *41*(1), 261–287. https://doi.org/10.1177/0149206314557524

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*(2), 121–138. https://doi.org/10.1037/1082-989x.12.2.121

Fleeson, W., & Noftle, E. (2008). The end of the person–situation debate: An emerging synthesis in the answer to the consistency question. *Social and Personality Psychology Compass*, *2*(4), 1667–1684. https://doi.org/10.1111/j.1751-9004.2008.00122.x

Freudenstein, J.-P., Mussel, P., & Krumm, S. (2023). On the construct-related validity of implicit trait policies. *European Journal of Personality*, *37*(1), 113–127. https://doi.org/10.1177/08902070211056901

Freudenstein, J.-P., Schäpers, P., Roemer, L., Mussel, P., & Krumm, S. (2020). Is it all in the eye of the beholder? The importance of situation construal for situational judgment test performance. *Personnel Psychology*, *73*(4), 669–700. https://doi.org/10.1111/peps.12385

Gatzka, T., & Volmer, J. (2017). Situational judgment test für Teamarbeit (SJT-TA) [Situational judgment test for teamwork]. *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. https://doi.org/10.6102/zis249

Grand, J. A. (2020). A general response process theory for situational judgment tests. *Journal of Applied Psychology*, *105*(8), 819–862. https://doi.org/10.1037/apl0000468

Guenole, N., Chernyshenko, O. S., & Weekly, J. (2017). On designing construct driven situational judgment tests: Some preliminary recommendations. *International Journal of Testing*, *17*(3), 234–252. https://doi.org/10.1080/15305058.2017.1297817

Harris, A. M., Siedor, L. E., Fan, Y., Listyg, B., & Carter, N. T. (2016). In defense of the situation: An interactionist explanation for performance on situational judgment tests. *Industrial and Organizational Psychology*, *9*(1), 23–28. https://doi.org/10.1017/iop.2015.110

Herde, C. N., & Lievens, F. (2020). Multiple speed assessments. *European Journal of Psychological Assessment*, *36*(2), 237–249. https://doi.org/10.1027/1015-5759/a000512

Jackson, D. J. R., LoPilato, A. C., Hughes, D., Guenole, N., & Shalfrooshan, A. (2017). The internal structure of situational judgement tests reflects candidate main effects: Not dimensions or situations. *Journal of Occupational and Organizational Psychology*, *90*(1), 1–27. https://doi.org/10.1111/joop.12151

Jansen, A., Melchers, K. G., Lievens, F., Kleinmann, M., Brändli, M., Fraefel, L., & König, C. J. (2013). Situation assessment as an ignored factor in the behavioral consistency paradigm underlying the validity of personnel selection procedures. *Journal of Applied Psychology*, *98*(2), 326–341. https://doi.org/10.1037/a0031257

Judge, T. A., & Zapata, C. P. (2015). The person–situation debate revisited: Effect of situation strength and trait activation on the validity of the Big Five personality traits in predicting job performance. *Academy of Management Journal*, *58*(4), 1149–1179. https://doi.org/10.5465/amj.2010.0837

Kasten, N., & Freund, P. A. (2016). A meta-analytical multilevel reliability generalization of situational judgment tests (SJTs). *European Journal of Psychological Assessment*, *32*(3), 230–240. https://doi.org/10.1027/1015-5759/a000250

Keeler, K. R., Kong, W., Dalal, R. S., & Cortina, J. M. (2019). Situational strength interactions: Are variance patterns consistent with the theory? *Journal of Applied Psychology*, *104*(12), 1487–1513. https://doi.org/10.1037/apl0000416

Krumm, S., Hüffmeier, J., & Lievens, F. (2019). Experimental test validation: Examining the path from test elements to test performance. *European Journal of Psychological Assessment*, *35*(2), 225–232. https://doi.org/10.1027/1015-5759/a000393

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

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, *11*(4), 815–852. https://doi.org/10.1177/1094428106296642

Lee, S., & Dalal, R. S. (2016). Climate as situational strength: Safety climate strength as a cross-level moderator of the relationship between conscientiousness and safety behaviour. *European Journal of Work and Organizational Psychology*, *25*(1), 120–132. https://doi.org/10.1080/1359432x.2014.987231

Lievens, F. (2017a). Assessing personality-situation interplay in personnel selection: Toward more integration into personality research. *European Journal of Personality*, *31*(5), 424–440. https://doi.org/10.1002/per.2111

Lievens, F. (2017b). Construct-driven SJTs: Toward an agenda for future research. *International Journal of Testing*, *17*(3), 269–276. https://doi.org/10.1080/15305058.2017.1309857

Lievens, F., & De Soete, B. (2012). Simulations. In N. Schmitt (Ed.), *The Oxford handbook of personnel assessment and selection* (pp. 383–410). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199732579.013.0017

Lievens, F., Lang, J. W. B., De Fruyt, F., Corstjens, J., Van de Vijver, M., & Bledow, R. (2018). The predictive power of people's intraindividual variability across situations: Implementing whole trait theory in assessment. *Journal of Applied Psychology*, *103*(7), 753–771. https://doi.org/10.1037/apl0000280

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., Schollaert, E., & Keen, G. (2015). The interplay of elicitation and evaluation of trait-expressive behavior: Evidence in assessment center exercises. *Journal of Applied Psychology*, *100*(4), 1169–1188. https://doi.org/10.1037/apl0000004

Lievens, F., Tett, R. P., & Schleicher, D. J. (2009). Assessment centers at the crossroads: Toward a reconceptualization of assessment center exercises. In J. J. Martocchio, & H. Liao (Eds.), *Research in personnel and human resources management* (pp. 99–152). Emerald Group Publishing Limited. https://doi.org/10.1108/S0742-7301(2009)0000028006

Little, R. J. A. (1992). Regression with missing X's: A review. *Journal of the American Statistical Association*, *87*(420), 1227–1237.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*(3), 86–92. https://doi.org/10.1027/1614-1881.1.3.86

Martin-Raugh, M. P., & Kell, H. J. (2021). A process model of situational judgment test responding. *Human Resource Management Review*, *31*(2), 100731. https://doi.org/10.1016/j.hrmr.2019.100731

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. L. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*, *60*(1), 63–91. https://doi.org/10.1111/j.1744-6570.2007.00065.x

McDaniel, M. A., List, S. K., & Kepes, S. (2016). The "hot mess" of situational judgment test construct validity and other issues. *Industrial and Organizational Psychology*, *9*(1), 47–51. https://doi.org/10.1017/iop.2015.115

McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology*, *86*(4), 730–740. https://doi.org/10.1037//0021-9010.86.4.730

McDaniel, M. A., & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment*, *9*(1–2), 103–113. https://doi.org/10.1111/1468-2389.00167

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437–455. https://doi.org/10.1037/a0028085

Melchers, K. G., & Kleinmann, M. (2016). Why situational judgment is a missing component in the theory of SJTs. *Industrial and Organizational Psychology*, *9*(1), 29–34. https://doi.org/10.1017/iop.2015.111

Melchers, K. G., Wirz, A., & Kleinmann, M. (2012). Dimensions AND exercises: Theoretical background of mixed-model assessment centers. In D. J. R. Jackson, C. E., Lance, & B. J., Hoffman (Eds.), *The Psychology of Assessment Centers*, (pp. 237–254). Routledge.

Meyer, R. D., Dalal, R. S., & Bonaccio, S. (2009). A meta-analytic investigation into the moderating effects of situational strength on the conscientiousness–performance relationship. *Journal of Organizational Behavior*, *30*(8), 1077–1102. https://doi.org/10.1002/job.602

Meyer, R. D., Dalal, R. S., & Hermida, R. (2010). A review and synthesis of situational strength in the organizational sciences. *Journal of Management*, *36*(1), 121–140. https://doi.org/10.1177/0149206309349309

Meyer, R. D., Dalal, R. S., José, I. J., Hermida, R., Chen, T. R., Vega, R. P., Brooks, C. K., & Khare, V. P. (2014). Measuring job-related situational strength and assessing its interactive effects with personality on voluntary work behavior. *Journal of Management*, *40*(4), 1010–1041. https://doi.org/10.1177/0149206311425613

Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychological Review*, *80*, 252–283. https://doi.org/10.1037/h0035002

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

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

Motowidlo, S. J., Hooper, A. C., & Jackson, H. L. (2006). Implicit policies about relations between personality traits and behavioral effectiveness in situational judgment items. *Journal of Applied Psychology*, *91*(4), 749–761. https://doi.org/10.1037/0021-9010.91.4.749

Mumford, T. V., Van Iddekinge, C. H., Morgeson, F. P., & Campion, M. A. (2008). The team role test: Development and validation of a team role knowledge situational judgment test. *Journal of Applied Psychology*, *93*(2), 250–267. https://doi.org/10.1037/0021-9010.93.2.250

Mussel, P., Gatzka, T., & Hewig, J. (2018). Situational judgment tests as an alternative measure for personality assessment. *European Journal of Psychological Assessment*, *34*(5), 328–335. https://doi.org/10.1027/1015-5759/a000346

Olaru, G., Burrus, J., Maccann, C., Zaromb, F. M., Wilhelm, O., & Roberts, R. D. (2019). Situational judgment tests as a method for measuring personality: Development and validity evidence for a test of dependability. *PLoS ONE*, *14*(2), e0211884. https://doi.org/10.1371/journal.pone.0211884

Olaru, G., Witthöft, M., & Wilhelm, O. (2015). Methods matter: Testing competing models for designing short-scale big-five assessments. *Journal of Research in Personality*, *59*, 56–68. https://doi.org/10.1016/j.jrp.2015.09.001

Oliver, T., Hausdorf, P., Lievens, F., & Conlon, P. (2016). Interpersonal dynamics in assessment center exercises: Effects of role player portrayed disposition. *Journal of Management*, *42*(7), 1992–2017. https://doi.org/10.1177/0149206314525207

Oostrom, J. K., de Vries, R. E., & de Wit, M. (2018). Development and validation of a HEXACO situational judgment test. *Human Performance*, *32*(1), 1–29. https://doi.org/10.1080/08959285.2018.1539856

Rammstedt, B., & John, O. P. (2005). Kurzversion des Big Five Inventory (BFI-K) [Short version of the Big Five inventory]. *Diagnostica*, *51*(4), 195–206. https://doi.org/10.1026/0012-1924.51.4.195

Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A., Ziegler, M., Jones, A. B., & Funder, D. C. (2014). The situational eight DIAMONDS: A taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology*, *107*(4), 677–718. https://doi.org/10.1037/a0037250

Rockstuhl, T., Ang, S., Ng, K.-Y., Lievens, F., & Van Dyne, L. (2015). Putting judging situations into situational judgment tests: Evidence from intercultural multimedia SJTs. *Journal of Applied Psychology*, *100*(2), 464–480. https://doi.org/10.1037/a0038098

Rockstuhl, T., & Lievens, F. (2021). Prompt-specificity in scenario-based assessments: Associations with personality versus knowledge and effects on predictive validity. *Journal of Applied Psychology*, *106*(1), 122–139. https://doi.org/10.1037/apl0000498

Saucier, G., Bel-Bahar, T., & Fernandez, C. (2007). What modifies the expression of personality tendencies? Defining basic domains of situation variables. *Journal of Personality*, *75*(3), 479–504. https://doi.org/10.1111/j.1467-6494.2007.00446.x

Schäpers, P., Freudenstein, J.-P., Mussel, P., Lievens, F., & Krumm, S. (2020). Effects of situation descriptions on the construct-related validity of construct-driven situational judgment tests. *Journal of Research in Personality*, *87*, 103963. https://doi.org/10.1016/j.jrp.2020.103963

Schäpers, P., Lievens, F., Freudenstein, J.-P., Hüffmeier, J., König, C. J., & Krumm, S. (2020). Removing situation descriptions from situational judgment test items: Does the impact differ for video-based versus text-based formats? *Journal of Occupational and Organizational Psychology*, *93*(2), 472–494. https://doi.org/10.1111/joop.12297

Schäpers, P., Mussel, P., Lievens, F., König, C. J., Freudenstein, J.-P., & Krumm, S. (2020). The role of situations in situational judgment tests: Effects on construct saturation, predictive validity, and applicant perceptions. *Journal of Applied Psychology*, *105*(8), 800–818. https://doi.org/10.1037/apl0000457

Schollaert, E., & Lievens, F. (2012). Building situational stimuli in assessment center exercises: Do specific exercise instructions and role-player prompts increase the observability of behavior? *Human Performance*, *25*(3), 255–271. https://doi.org/10.1080/08959285.2012.683907

Schultze, M. (2017). *Constructing subtests using ant colony optimization* [Doctoral dissertation, Freie Universität Berlin]. https://doi.org/10.17169/refubium-622

Schweizer, K. (2010). Some guidelines concerning the modeling of traits and abilities in test construction. *European Journal of Psychological Assessment*, *26*(1), 1–2. https://doi.org/10.1027/1015-5759/a000001

Shaffer, J. A., & Postlethwaite, B. E. (2012). A matter of context: A meta-analytic investigation of the relative validity of contextualized and noncontextualized personality measures. *Personnel Psychology*, *65*(3), 445–494. https://doi.org/10.1111/j.1744-6570.2012.01250.x

Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology*, *88*(3), 500–517. https://doi.org/10.1037/0021-9010.88.3.500

Tutz, G., & Hennevogl, W. (1996). Random effects in ordinal regression models. *Computational Statistics & Data Analysis*, *22*(5), 537–557. https://doi.org/10.1016/0167-9473(96)00004-7

Weekley, J. A., Hawkes, B., Guenole, N., & Ployhart, R. E. (2015). Low-fidelity simulations. *Annual Review of Organizational Psychology and Organizational Behavior*, *2*(1), 295–322. https://doi.org/10.1146/annurev-orgpsych-032414-111304

Westring, A. J. F., Oswald, F. L., Schmitt, N., Drzakowski, S., Imus, A., Kim, B., & Shivpuri, S. (2009). Estimating trait and situational variance in a situational judgment test. *Human Performance*, *22*(1), 44–63. https://doi.org/10.1080/08959280802540999

How to cite this article: Freudenstein, J.-P., Schäpers, P., Reznik, N., Stolte, T., & Krumm, S. (2024). The influence of situational strength on the relation of personality and situational judgment test performance. *International Journal of Selection and Assessment*, *32*, 1–11. https://doi.org/10.1111/ijsa.12444