The importance of situation evaluation and the ability to identify criteria in a construct-driven situational judgment test

Personality and Individual Differences 208 (2023) 112182

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

The importance of situation evaluation and the ability to identify criteria in a construct-driven situational judgment test

Dan Wang [a,*], Janneke K. Oostrom [b], Eveline Schollaert [c]

[a] School of Management, Xihua University, 999 Jin Zhou Road, Chengdu 610039, China
[b] Department of Social Psychology, Tilburg School of Social and Behavioral Sciences, Tilburg University, PO Box 90153, 5000 LE Tilburg, the Netherlands
[c] Department of Marketing, Innovation, and Organization, Faculty of Economics and Business Administration, Ghent University, Tweekerkenstraat 2, Gent 9000, Belgium

ARTICLE INFO

Keywords:
Situational judgment test
Situational construal
Ability to identify criteria
Criterion-related validity
Trait activation theory

ABSTRACT

Situational judgment tests (SJTs) typically focus on candidates' judgment of various response options and ignore the process of arriving at that judgment. In three studies, we developed two situational construal measures (i.e., situation evaluation and the ability to identify criteria [ATIC]) based on a construct-driven SJT measuring cooperativeness. We tested their role in the response judgment process and examined their incremental validity in predicting work performance. Study 1 ($N = 30$) provided qualitative evidence that situation evaluation and ATIC shape response judgment. Study 2 ($N_{\text{student}} = 136$, $N_{\text{employee}} = 154$) showed that situation evaluation predicts response judgment. Moreover, situation evaluation had incremental validity in predicting supervisor-rated task performance and peer-rated interpersonal performance over and above response judgment and ATIC. Furthermore, ATIC predicted response judgment and had incremental validity in predicting interpersonal performance but not task performance over and above response judgment. Finally, Study 3 ($N = 266$) replicated the findings of Study 2 using a new sample and yielded similar results regardless of the order in which situation evaluation and response judgment were measured.

1. Introduction

An individual's behavior is a function of the characteristics of the individual and the situation (e.g., Mischel, 1973; Tett & Burnett, 2003). However, it are not just objective characteristics of a situation that predict an individual's behavior; perceptions of a situation—which are in part idiosyncratic (Rauthmann et al., 2015)—also matter (Sherman et al., 2015; Ziegler et al., 2019). In fact, this so-called situational construal can reveal important and unique trait-relevant information (Golubovich et al., 2022; Lievens et al., 2021; Martin-Raugh & Kell, 2021). Unfortunately, the development of personnel selection instruments that can assess situational construal lags behind. In the present study, we argue that situational judgment tests (SJTs), which typically focus on candidates' perceptions of response options (i.e., response judgment; Lievens et al., 2021; Weekley et al., 2015), can be redesigned to measure situational construal as well. Specifically, by revealing the thought process leading up to response judgment (Rockstuhl et al., 2015), SJTs would allow for a more thorough assessment of candidates' characteristics, which can increase their criterion-related validity (e.g., Freudenstein et al., 2020; Harris et al., 2016; Martin-Raugh & Kell, 2021).

So far, researchers have proposed two approaches to measure situational construal in SJTs (Freudenstein et al., 2020; Golubovich et al., 2022; Rockstuhl et al., 2015). First, Rockstuhl et al. (2015) proposed an open-ended methodology (i.e., asking candidates to report their thoughts, feelings, and intentions of the people in each SJT situation) to measure candidates' sense-making of situations. Second, Freudenstein et al. (2020) proposed a selected-response methodology (i.e., asking candidates to rate broad dimensions of situation characteristics within each SJT item based on the DIAMONDS taxonomy; Rauthmann et al., 2014) to measure candidates' psychological representations of situations. These two situational construal approaches predicted response judgment (even when situation descriptions were absent; Freudenstein et al., 2020). Importantly, the two situational construal approaches also predicted relevant criteria over and above response judgment (Freudenstein et al., 2020; Rockstuhl et al., 2015). Although these previous studies show the added value of adding a situational construal measure to an SJT, we still know relatively little about how candidates interpret specific stimuli in SJT situations (i.e., situational cues), how these interpretations influence candidates' response judgment, and how these

* Corresponding author at: School of Management, Xihua University, 999 Jin Zhou Road, Chengdu 610039, China.
E-mail addresses: danwang@mail.xhu.edu.cn (D. Wang), J.K.Oostrom@tilburguniversity.edu (J.K. Oostrom), Eveline.Schollaert@UGent.be (E. Schollaert).

https://doi.org/10.1016/j.paid.2023.112182
Received 8 November 2022; Received in revised form 8 March 2023; Accepted 15 March 2023
Available online 29 March 2023

interpretations contribute to the criterion-related validity of SJTs (Fan et al., 2016).

We build on trait activation theory (TAT) to examine the role of situational cues in SJTs (Lievens et al., 2021; Tett & Burnett, 2003), which postulates that individuals regard situational cues as hints for the traits or attributes that should be activated and guide their work behaviors (Tett et al., 2021). Thus, when completing SJTs, candidates interpret situations by extracting cues they believe are meaningful (Golubovich et al., 2022) and infer what attributes are being assessed based on these cues (Fan et al., 2016; Harris et al., 2016). These interpretations are subjective and partly guided by candidates' traits (Lievens & Motowidlo, 2016; Ziegler et al., 2019) and should provide unique trait-relevant information. In the present research, we developed two situational construal measures that should help us understand how candidates interpret and utilize situational cues in SJTs: situation evaluation and the ability to identify criteria (i.e., ATIC). These concepts refer to different types of situational construal; situation evaluation refers to the subjective interpretation of situational cues that are aimed to elicit specific traits (Freudenstein et al., 2020), whereas ATIC refers to a more general "ability to correctly perceive performance criteria when participating in an evaluative situation" (Kleinmann et al., 2011, p. 129).

Our study contributes to the SJT literature in three ways. First, by integrating the SJT (Freudenstein et al., 2020; Rockstuhl et al., 2015) and ATIC literature (Jansen et al., 2013; Kleinmann et al., 2011; Oostrom et al., 2016) we advance current knowledge of the thought process leading up to response judgment by differentiating between situation evaluation and ATIC and examining how these two types of situational construal shape response judgment. Second, we improve the understanding of how SJTs work and offer a means to increase their criterion-related validity by examining the extent to which situation evaluation and ATIC incrementally predict work performance over and above response judgment. Third, we extend TAT concerning the main effect of situational cues on individuals' behaviors by identifying which types of situational cues are more likely to activate relevant traits and elicit trait-expressive behaviors (Tett et al., 2021).

1.1. *Situational judgment tests*

SJTs typically consist of simulated work-related situations and a range of response options to deal with situations (Lievens et al., 2021; Murano et al., 2021; Weekley et al., 2015). Originally, SJTs were seen as low-fidelity simulations that were not explicitly developed to measure specific constructs (Lievens et al., 2021; Weekley et al., 2015). This traditional development approach caused their construct-related validity to be "a 'hot mess' without much theoretical or empirical guidance" (McDaniel et al., 2016, p. 50). Consequently, the adoption of a more construct-driven approach for developing SJTs has been pushed (e.g., Guenole et al., 2017; Oostrom et al., 2019) (see supplementary materials for details on the differences between the two approaches).

Construct-driven SJTs are designed to assess "inherent beliefs about the general effectiveness of actions that express traits to varying degrees" (Lievens et al., 2021, p. 295), which partly reflect candidates' trait levels. Recent research has shown that the reliability and construct-related validity of construct-driven SJTs vary across studies and traits (Neubauer & Hofer, 2022). For example, Oostrom et al. (2019) found alphas from 0.22 to 0.66 and Mussel et al. (2018) from 0.55 to 0.73. Nevertheless, these SJTs have acceptable criterion-related validity in predicting work-related behaviors and life events (Koschmieder and Neubauer, 2021; Lievens et al., 2021; Mussel et al., 2018; Oostrom et al., 2019).

The criterion-related validity of SJTs is a function of individuals' characteristics and their situational construal. Individuals' characteristics are typically measured by response judgment, which reflects their ratings of trait-expressive behaviors in response to the situations (Mussel et al., 2018; Oostrom et al., 2019). In contrast, we know relatively little about the measurement of situational construal (Lievens et al., 2021). Besides, the existing findings about the role of situation descriptions in situational construal measures are inconsistent (Freudenstein et al., 2020; Golubovich et al., 2022; Rockstuhl et al., 2015). For example, Rockstuhl et al. (2015) found that a situational construal measure (i.e., "situational judgment") based on situation descriptions provides incremental validity over and above response judgment. In contrast, Freudenstein et al. (2020) found that a situational construal measure based on only response options (i.e., situation descriptions were absent) can also contribute to SJT performance, indicating that situation descriptions are not crucial for individuals' situational construal. Recently, Golubovich et al. (2022) substantiated situation descriptions' small but important role in SJTs and suggested using situational cues to boost item validity. Thus, more research on the role of situation descriptions (i.e., focusing on how and to what extent situational cues can boost validity) is warranted to address this ongoing debate (Krumm et al., 2015; Lievens et al., 2021; Rockstuhl et al., 2015).

1.2. *Situational construal*

TAT posits three levels of situational cues in the workplace: task-, social-, and organizational-level cues (Tett et al., 2021). These cues or situational demands can "activate" specific traits and elicit valued work-related behaviors (e.g., work performance) (Tett & Burnett, 2003). Situations vary in terms of the number and types of situational cues they provide and the opportunity to express certain traits (Tett & Burnett, 2003; Tett & Guterman, 2000). A situation is relevant to a trait to the degree that it offers the opportunity to express that trait (Tett & Guterman, 2000). Trait-relevant situations have a main effect on individuals' behavior but can also expose individual differences (Tett et al., 2021), including people's threshold and sensitivity to situational cues (Tett & Guterman, 2000). These individual differences in the interpretation of situational cues (i.e., situational construal) are an important driver of people's behavior and are (partly) guided by people's traits (Ziegler et al., 2019). For example, a person with a high level of cooperativeness is more likely to focus on situational cues that convey the demand for cooperation, attach importance to cues relevant to cooperativeness, and show cooperative behavior. As such, situational construal should reveal unique trait-relevant information. The present study aims to develop two types of situational construal based on these TAT principles: situation evaluation and ATIC. Specifically, we focus on the situational construal of situations that are aimed to elicit cooperativeness (i.e., an important aspect of agreeableness).

1.2.1. *Situation evaluation*

Drawing on TAT (Tett et al., 2021) and in line with recent process models (e.g., Grand, 2020; Martin-Raugh & Kell, 2021), we conceptualize situation evaluation as candidates' subjective interpretation of crucial situational cues that activate (implicit) traits and guide behaviors. Thus, we believe that, by including a situation evaluation measure, SJTs can measure two trait aspects: 1) beliefs about the effectiveness of the expression of the intended traits (i.e., response judgment) and 2) the subjective interpretation of the cues relevant to traits activated (i.e., situation evaluation).

1.2.1.1. Situation evaluation and response judgment. According to TAT, the subjective interpretation of situational cues guides trait-relevant behaviors (Tett & Burnett, 2003; Tett & Guterman, 2000). Recently, Brown et al. (2016) argued that this rationale applies to SJTs as well. In fact, they argued that individual differences in the perception of situational cues might be pivotal in understanding candidates' response judgment. Therefore, the situational cues that candidates deem most meaningful should subsequently guide their response judgment (Golubovich et al., 2022). Furthermore, situation evaluation and response judgment capture trait aspects that are relevant within the SJT's construct domain (Martin-Raugh & Kell, 2021). Indeed, although their operationalization of situation evaluation differed from ours, both Rockstuhl et al. (2015) and Freudenstein et al. (2020) found situation evaluation to predict response judgment. Accordingly, we postulated the following:

Hypothesis 1. Situation evaluation predicts response judgment in a construct-driven SJT.

1.2.1.2. Incremental validity of situation evaluation over and above response judgment. Although situation evaluation and response judgment are expected to be correlated, we argue that situation evaluation has incremental validity over and above response judgment in predicting work performance. Situation evaluation captures candidates' subjective interpretation of crucial situational cues (Grand, 2020), which is guided by candidates' traits (Ziegler et al., 2019) and activates the subsequent expression of these traits (Harris et al., 2016; Krumm et al., 2015; Lievens & Motowidlo, 2016). However, it is likely that the expression of these traits is only partly captured by the response options presented to the candidate (Krumm et al., 2015; Lievens et al., 2021; Motowidlo et al., 1990). Thus, a more direct measure of the situation evaluation should reveal important and unique trait-relevant information (Rockstuhl et al., 2015). Therefore, situation evaluation is likely to explain incremental variance in work performance (i.e., task and interpersonal performance) over and above response judgment. Accordingly, we postulated the following:

Hypothesis 2. Situation evaluation predicts task performance (H2a) and interpersonal performance (H2b) over and above response judgment in a construct-driven SJT.

1.2.2. The ability to identify criteria

Similar to situation evaluation, ATIC plays an important role in the thought process leading up to response judgment. However, situation evaluation and ATIC differ in three important ways. First, while situation evaluation refers to candidates' *psychological* construal of the situation and encompasses the interaction process between situational cues and noncognitive individual differences such as personality traits and states (Funder, 2016; Rauthmann et al., 2014; Ziegler et al., 2019), ATIC refers to candidates' *cognitive* construal of the situation and encompasses the interaction process between situational cues and cognitive individual differences such as cognitive ability, skills, and knowledge (Kleinmann et al., 2011). Second, there is a difference in reference point: when evaluating the situation, candidates indicate the most important situational cue for their response (Golubovich et al., 2022). In contrast, when identifying the criteria being measured, candidates consider which dimension others would use to evaluate their response (Kleinmann et al., 2011). Third, while situation evaluation contributes to measuring specific traits, ATIC refers to a more general ability to read relevant situational cues (Kleinmann et al., 2011). Indeed, research has shown that ATIC correlates with cognitive ability (Jansen et al., 2013; Melchers et al., 2020) and that ATIC is consistent across predictor methods that measure different trait dimensions (e.g., Kleinmann et al., 2011; Melchers et al., 2020). Thus, by including an ATIC measure in a construct-driven SJT, the SJT can measure 1) response judgment, 2) situation evaluation, and 3) a general ability to correctly decipher situational demands (i.e., ATIC).

1.2.2.1. The ability to identify criteria and response judgment. The theoretical assumptions underlying ATIC posit that candidates better at perceiving the situational demands in simulation-based predictor methods manage to show more relevant behaviors (i.e., behaviors that reflect the constructs being assessed) and therefore receive better scores on the targeted constructs. In other words, ATIC helps candidates figure out what traits are being assessed in the SJT based on the situational cues, thereby facilitating the expression of these traits through their response judgment (Fan et al., 2016; Tett et al., 2021). Empirical evidence suggests that ATIC may indeed predict candidates' performance in simulation-based predictor methods (Ingold et al., 2015; Kleinmann et al., 2011). The present study extends the existing ATIC research by examining how ATIC relates to response judgment in SJTs, answering calls for research revealing the cognitive process of responding to situational cues and how SJTs work (Fan et al., 2016; Harris et al., 2016). Accordingly, we postulated the following:

Hypothesis 3. ATIC predicts response judgment in a construct-driven SJT.

1.2.2.2. Incremental validity of the ability to identify criteria over and above response judgment. ATIC accounts for candidates' performance in simulation-based predictor methods (Ferris et al., 2002; Kleinmann et al., 2011) and on the job (Ingold et al., 2015; Jansen et al., 2013; Oostrom et al., 2016). Although ATIC and response judgment are expected to be correlated, we argue that ATIC has incremental validity over and above response judgment in predicting work performance (i.e., task and interpersonal performance). Specifically, ATIC helps candidates discern evaluation criteria, which affects their response judgment (Fan et al., 2016), and helps improve their work performance (Kleinmann et al., 2011; Oostrom et al., 2016). Thus, ATIC predicts work performance because it increases candidates' social effectiveness, which refers to "the ability to effectively read, understand, and control social interactions" (Ferris et al., 2002, p. 49). This suggests that ATIC captures unique abilities and skills only partly captured by response judgment. Accordingly, we postulated the following:

Hypothesis 4. ATIC predicts task performance (H4a) and interpersonal performance (H4b) over and above response judgment in a construct-driven SJT.

1.2.2.3. Incremental validity of situation evaluation over and above the ability to identify criteria. Given that situation evaluation and ATIC capture related yet distinct individual differences (Ingold et al., 2015; Krumm et al., 2015), we compared their criterion-related validity in predicting work performance. We argue that situation evaluation explains additional variance in work performance over and above ATIC for the following two reasons. First, situation evaluation refers to candidates' *psychological* construal of the situation and is most likely to reflect noncognitive individual differences (Funder, 2016; Rauthmann et al., 2014; Ziegler et al., 2019), while ATIC refers to candidates' *cognitive* construal of the situation and is most likely to reflect cognitive ability, skills, and knowledge. Generally, traits have incremental validity in predicting work performance over and above ability (e.g., Lee et al., 2019). Second, construct-driven SJTs may be relatively transparent in terms of the targeted traits that are being assessed. Such transparency may limit the variance in the cognitive construal of the situation (i.e., ATIC; Ingold et al., 2015) but does not need to reduce variance in the psychological construal of the situation (i.e., situational judgment). Accordingly, we postulated the following:

Hypothesis 5. Situation evaluation predicts task performance (H5a) and interpersonal performance (H5b) over and above ATIC in a construct-driven SJT.

1.3. Overview studies

We conducted three studies to show the value of revealing candidates' thought process leading up to response judgment in a construct-driven SJT measuring cooperativeness. In Study 1, we developed two situational construal measures (i.e., situation evaluation and ATIC) and conducted interviews to reveal participants' thoughts about situational cues. In Study 2, we examined the role of situation evaluation and ATIC in response judgment and tested their incremental validity over and above response judgment in predicting work performance (i.e., task and interpersonal performance). We focused on task and interpersonal performance because cooperativeness is seen as one of the 21st century skills that can predict both types of valued organizational behaviors across many jobs (Sackett & Walmsley, 2014). Finally, in Study 3, we strengthened the internal validity of our findings in Study 2 by testing the order effects of measuring situation evaluation and response judgment.

2. Study 1

Study 1 aimed to develop a situational judgment and ATIC measure for a construct-driven SJT, which were used in subsequent studies. Following the development of the two situational construal measures, we conducted interviews to explore whether our two new measures reveal individual differences in situational construal and to reveal why participants consider certain situational cues to be crucial.

2.1. Method

We developed two situational construal measures (i.e., situation evaluation and ATIC) and conducted the interviews using a validated construct-driven SJT that measured cooperativeness (CO-SJT; Wang, 2022). This CO-SJT consisted of eight items comprising rich situational cues (in the situation descriptions), which enabled us to investigate the validity of the situational construal measures. For each CO-SJT item, three possible response options were presented following each situation description. Participants were asked to indicate their behavioral tendency (i.e., indicate the likelihood that they would respond that way) for each response option on a 6-point Likert scale. Participants received 6 points if they assessed the highly effective option as (very) likely and 1 point if they assessed it as (very) unlikely. Similarly, participants received 6 points if they assessed the highly ineffective option as (very) unlikely and 1 point if they assessed it as (very) likely. Scores on response options were averaged to create an overall score for each item, ranging from 1 to 6.

2.1.1. Development of situation evaluation measure

Ten experts[1] (five managers and five postgraduate students; three men and seven women; $M_{age} = 26.90$ years, $SD = 3.60$; $M_{work\ experience} = 3.30$ years, $SD = 2.58$) independently read the situation descriptions of the CO-SJT (Wang, 2022). Then, together, they identified all situational cues that conveyed a unique piece of information that may elicit trait-expressive behaviors. These cues were labeled as A, B, C, and D (and if there were five situational cues also E) according to the order in which they were presented. Then, to generate the scoring key for the situation evaluation measure, nine experts (five male and four female; $M_{age} = 30.22$ years, $SD = 3.15$; $M_{work\ experience} = 4.67$ years, $SD = 2.74$) rated the importance of each labeled situational cue on a 3-point Likert-type scale, with 1 = *not crucial*, 2 = *somewhat crucial*, and 3 = *crucial* (ICC = 0.84), which then served as scoring keys. To measure situation evaluation, participants are asked after each SJT situation to indicate which situational cue they consider the most crucial for deriving their response. They receive 3 points if they choose a crucial cue, 2 points if they choose a somewhat crucial cue, and 1 point if they choose a cue that the experts did not deem crucial. The scoring keys for situation evaluation can be found in Table S1.

2.1.2. Development of the ability to identify criteria measure

ATIC is typically measured using an open-ended approach (e.g., Ingold et al., 2015; Kleinmann, 1993; Oostrom et al., 2016), in which participants are instructed to write down assumptions about targeted dimensions and example behaviors. Afterward, trained raters rate the fit with the targeted dimensions. However, Speer et al. (2014) introduced a multiple-choice approach in which participants are instructed to select the most and least important dimension from a larger list of dimensions. Following Speer et al.'s approach, we also developed a multiple-choice ATIC measure in which we compare participants' assumptions regarding the evaluated dimensions to experts' judgment of the dimensions being measured in each item (e.g., Kleinmann et al., 2011). Nine experts (five men, four women; $M_{age} = 25.33$ years, $SD = 0.50$) created a list of evaluated dimensions for each item. Cooperativeness serves as the scoring key for each item because the CO-SJT was developed to measure cooperativeness. Besides, the experts wrote down several distractors (e.g., leadership or responsibility) and discussed them to determine the final response options. To measure ATIC participants are asked, after their response judgment, to indicate which dimension is most likely to be evaluated with the item. They are presented with five options: a correct option, three distractors, and an "other skills or ability" category. They receive 1 point for choosing the correct dimension (i.e., cooperativeness) and 0 points for choosing any of the other dimensions.

2.1.3. Interviews

2.1.3.1. Participants and procedure. Given that data saturation is generally expected after 20–30 interviews (Marshall et al., 2013), we interviewed 30 Ph.D. students (17 women, 13 men) from China majoring in Business Administration. On average, they were 25.10 years old ($SD = 1.00$ years) and had 0.45 years ($SD = 0.78$) of work experience. Six experts (five women and one man, $M_{age} = 26.17$ years, $SD = 2.93$, $M_{work\ experience} = 2.83$ years, $SD = 2.64$), including two personnel managers, two doctoral students, and two master students majoring in Business Administration, acted as interviewers. We obtained participants' written informed consent before the interview.

At the start of the interview, the experts presented participants with a short questionnaire to measure their demographic characteristics, followed by the eight items of the CO-SJT (i.e., without the situational cue labels nor our situation evaluation and ATIC measure). Then, the experts asked participants for each SJT item to indicate what they would do (i.e., If you were to solve the dilemma, what would you do in the situation?), why they would do it (e.g., perception of crucial situational cues and what is being assessed), and why they considered specific situational cues crucial (e.g., elaboration on the possible meaningful information that specific cues convey). The interviews lasted, on average, 18.72 min ($SD = 8.61$). Participants received a notebook and a ballpoint pen as a token of appreciation for their efforts. After each interview, the experts recorded and labeled participants' interview answers to each SJT item, resulting in 240 unique records of *crucial situational cues, evaluated dimensions, and elaborations on* crucial situational cues (see supplementary materials for an explanation of the coding procedure).

2.2. Results and discussion

The coding of the interview answers (see Tables S2-S4) confirmed that participants' perceptions of crucial situational cues and evaluated dimensions vary from person to person. The number of participants believing a specific situational cue to be crucial in each situation ranged from 1 to 14 (see Table S2), and the number of participants considering a specific dimension to be measured in each item ranged from 1 to 19 (see Table S3). These initial findings indicate that our situational judgment and ATIC measure can capture individual differences in situational construal.

We reviewed participants' elaboration on why they considered specific cues crucial (see Table S4). Of the 240 recorded responses, 128

[1] Experts comprised human resources managers and postgraduate students. The human resources managers worked at different organizations and had more than two years of experience in personnel selection. The postgraduate students had majored in business administration, and, therefore, had good knowledge of personality traits, including cooperativeness.

times (53.3 %) participants indicated that they found a situational cue crucial because it conveyed information about specific problems and their causes and possible consequences. Furthermore, participants indicated 112 times (46.7 %) that the situational cue signified implicit principles for solving a problem. These findings suggest that participants concentrate on situational cues that offer hints about which traits to show to solve the presented dilemma.

Thus, Study 1 suggests that participants differ in terms of their situation evaluation and ATIC and that these two types of situational construal may shape response judgment. However, these insights demand a further empirical investigation. Accordingly, in Study 2, we examined the extent to which situation evaluation and ATIC predict response judgment and work performance.

3. Study 2

Study 2 aimed to examine the role of situation evaluation and ATIC in response judgment and to test the incremental validity of these two types of situational construal over and above response judgment in predicting work performance.

3.1. Method

3.1.1. Participants and procedure

Study 2 comprised two samples: a student sample and an employee sample. The student sample comprised 136 Chinese master students (53 men, 83 women; $M_{age} = 28.10$ years, $SD = 9.70$). The students voluntarily participated in this study as an optional task for a Human Resource Management course. As study participation was rated and rewarded with extra study credits, the students were motivated to perform well. Indeed, all participants passed our attention check item (i.e., “Please choose the strongly disagree option for this question”), indicating they took the test seriously. We believe these students are a relevant sample because working in teams is a required part of this course and cooperativeness is an important predictor of team outcomes (e.g., Bradley et al., 2013). A total of 116 participants were part-time MBA students, and 20 were full-time students. The average work experience was 3.10 years ($SD = 3.66$). The average grade point average (GPA) was 3.01 points ($SD = 0.83$). Data were collected online at three time points. At the beginning of the course (Time 1), participants gave their informed consent, completed a personality inventory, and provided demographic information. One month later (Time 2), participants completed the SJT. Finally, at the end of the course (Time 3), participants' team leaders evaluated their task performance, and five team members rated their interpersonal performance.

The employee sample comprised 154 employees from a public organization in eastern China (92 women, 62 men; $M_{age} = 28.90$ years, $SD = 3.02$; $M_{work\ experience} = 2.56$ years, $SD = 1.46$). The employees were told that the test was part of an upcoming training program important for their career development. They voluntarily participated without payment. The results of the attention check indicated that all participants took the test seriously. We reached these participants via the human resources manager of the organization. All employees in this organization work in teams. Participants provided their informed consent before completing a personality inventory, some basic demographic questions (Time 1), and CO-SJT (Time 2) over a one-month interval. One month later (Time 3), participants' supervisors evaluated their task performance, and three team members rated their interpersonal performance (i.e., their interactive behaviors in teamwork during a job rotation). Participants and raters were unaware of the SJT scores before the performance assessment.

3.1.2. Measures

3.1.2.1. SJT. We used the eight-item CO-SJT to measure situation evaluation, response judgment, and ATIC (see Appendix SA in supplementary materials). Each SJT situation description was followed by one item assessing situation evaluation, three response options assessing response judgment, and one item assessing ATIC (see Appendix A for a sample item).

3.1.2.2. Situation evaluation. We measured situation evaluation ($\alpha_{student} = 0.71$, $\alpha_{employee} = 0.72$) using the situation evaluation measure developed in Study 1. Following each SJT situation, participants were asked to indicate the most crucial situational cue as an indicator of their situation evaluation. We averaged the scores on the eight items to create an overall score ranging from 1 to 3.

3.1.2.3. Response judgment. We assessed response judgment ($\alpha_{student} = 0.70$, $\alpha_{employee} = 0.62$) by asking participants to indicate their behavioral tendency for each response option on a 6-point Likert-type scale. We averaged the scores on the eight items to create a response judgment score ranging from 1 to 6.

3.1.2.4. ATIC. We tested ATIC ($\alpha_{student} = 0.70$, $\alpha_{employee} = 0.70$) using the multiple-choice measure developed in Study 1. Following their response judgment, participants were asked to choose the dimension most likely to be evaluated with that item. We averaged the scores on the eight items to create an ATIC score ranging from 0 to 1.

3.1.2.5. Work performance. We measured work performance via *a three-item task performance scale and a three-item* interpersonal performance scale (see Appendix SB in supplementary materials), adapted from the Work Role Performance scale by Griffin et al. (2007). Supervisors rated participants' task performance (e.g., “Carried out the core parts of the task assigned to him/her well,” $\alpha_{student} = 0.83$, $\alpha_{employee} = 0.76$) on a 5-point Likert-type scale ranging from 1 (*very little*) to 5 (*great deal*). Five of participants' team members in the student sample and three of participants' team members in the employee sample rated participants' interpersonal performance (e.g., “Coordinated his or her work with team members,” student sample: $\alpha = 0.75$, $r_{WG(J)} = 0.88$; employee sample: $\alpha = 0.73$, $r_{WG(J)} = 0.83$) on a 5-point Likert-type scale ranging from 1 (*very little*) to 5 (*great deal*).

3.2. Results

3.2.1. Preliminary analyses

The means, standard deviations, and correlations of all variables for both samples are reported in Tables S5–S6. In both samples, situation evaluation and ATIC were significantly related to response judgment. Moreover, situation evaluation, ATIC, and response judgment were significantly associated with task and interpersonal performance.

3.2.2. Hypothesis testing

We postulated that situation evaluation would predict response judgment (H1) and have incremental validity over and above response judgment in predicting task performance (H2a) and interpersonal performance (H2b). Table 1 shows that situation evaluation was a significant predictor of response judgment ($\beta_{student} = 0.31$, $p < .001$; $\beta_{employee} = 0.39$, $p < .001$). Therefore, H1 was supported.

Hypothesis 2 was tested via hierarchical regression analyses, with response judgment in the first step and situation evaluation in the second step. Table 2 shows that situation evaluation predicted task performance ($\beta_{student} = 0.18$, $p = .03$; $\beta_{employee} = 0.17$, $p = .04$) and interpersonal performance ($\beta_{student} = 0.22$, $p < .01$; $\beta_{employee} = 0.26$, $p < .01$) over and above response judgment. Moreover, situation evaluation explained additional variance in task performance ($\Delta R^2_{student} = 0.03$, $p = .03$; $\Delta R^2_{employee} = 0.03$, $p = .04$) and interpersonal performance ($\Delta R^2_{student} = 0.05$, $p < .01$; $\Delta R^2_{employee} = 0.05$, $p < .01$) over and above response judgment. Thus, H2a and H2b were supported.

Table 1
Hierarchical linear regression results of situation evaluation and ability to identify criteria on response judgment (Study 2).

Variables	Response judgment			
	Student sample (N = 136)		Employee sample (N = 154)	
	Model 1	Model 2	Model 3	Model 4
Situation evaluation	0.31**		0.39**	
Ability to identify criteria		0.29**		0.31**
F	14.29**	12.58**	27.71**	16.08**
R^2	0.10	0.09	0.15	0.10
Adjusted R^2	0.09	0.08	0.15	0.09

Note. This table reports the standardized beta coefficients. *$p < .05$, **$p < .01$.

We hypothesized that ATIC would predict response judgment (H3) and have incremental validity over and above response judgment in predicting task performance (H4a) and interpersonal performance (H4b). Table 1 shows that ATIC was significantly related to response judgment ($\beta_{\text{student}} = 0.29, p < .01$; $r_{\text{employee}} = 0.31, p < .01$), supporting H3. Table 2 shows the results of hierarchical regression analysis with response judgment in the first step and ATIC in the second step. In contrast with H4a, ATIC incrementally predicted task performance over and above response judgment in the employee sample ($\beta_{\text{employee}} = 0.26$, $\Delta R^2_{\text{employee}} = 0.06$, $p < .01$), but not in the student sample ($\beta_{\text{student}} = 0.16$, $\Delta R^2_{\text{student}} = 0.02, p = .06$). However, ATIC incrementally predicted interpersonal performance ($\beta_{\text{student}} = 0.21$, $p = .01$, $\Delta R^2_{\text{student}} = 0.04$; $\beta_{\text{employee}} = 0.23$, $p < .01$, $\Delta R^2_{\text{employee}} = 0.05$) over and above response judgment, thus supporting H4b.

We posited that situation evaluation would have incremental validity over and above ATIC in predicting task performance (H5a) and interpersonal performance (H5b). Table 3 shows the results of hierarchical regression analysis with ATIC in the first step and situation evaluation in the second step. As predicted, situation evaluation predicted task performance ($\beta_{\text{student}} = 0.22$, $p = .01$; $\beta_{\text{employee}} = 0.21$, $p < .01$) and interpersonal performance ($\beta_{\text{student}} = 0.25, p < .01$; $\beta_{\text{employee}} = 0.28$, $p < .01$) and explained additional variance in task performance ($\Delta R^2_{\text{student}} = 0.05, p = .01$; $\Delta R^2_{\text{employee}} = 0.04, p < .01$) and interpersonal performance ($\Delta R^2_{\text{student}} = 0.06, p < .01$; $\Delta R^2_{\text{employee}} = 0.08, p < .01$) over and above ATIC. Thus, H5a and H5b were supported.

3.2.3. Additional analyses

We included control variables (i.e., work experience, several personality traits, and cognitive ability; see supplementary materials for our argumentation) in our hypothesized models to test the robustness of our findings. These analyses showed similar results, except for H4a. Specifically, ATIC explained additional variance in interpersonal performance but not in task performance over and above response judgment in the student sample ($\beta = 0.13, p = .16, \Delta R^2 = 0.01$) and employee sample ($\beta = 0.15$, $p = .07$, $\Delta R^2 = 0.02$) when the control variables were included. Besides, we found that cognitive ability ($\beta = 0.26$, $p < .01$) significantly predicted task performance in the employee sample. This finding suggests that the validity of cognitive construal in predicting task performance is partly because of its common variance with cognitive ability, confirming that "the identification of criteria is a cognitive construal" (Kleinmann et al., 2011, p.134).

3.3. Discussion

Study 2 shows that situation evaluation and ATIC are significantly and positively related to response judgment, providing empirical evidence for the findings in Study 1 that individuals' perceptions of crucial

Table 2
Hierarchical linear regression results of situation evaluation and ability to identify criteria over response judgment in predicting task performance and interpersonal performance (Study 2).

Variables	Student sample (N = 136)				Employee sample (N = 154)			
	Task performance		Interpersonal performance		Task performance		Interpersonal performance	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Response judgment	0.28**	0.29**	0.29**	0.30**	0.26**	0.25**	0.20*	0.23*
Situation evaluation	0.18*		0.22**		0.17*		0.26**	
Ability to identify criteria		0.16†		0.21*		0.26**		0.23**
F	11.33**	10.68**	14.07**	13.69**	11.92**	15.81**	13.31**	12.47**
R^2	0.15	0.14	0.17	0.17	0.14	0.17	0.15	0.14
ΔR^2	0.03*	0.02†	0.05**	0.04*	0.03*	0.06**	0.06**	0.05**
Adjusted R^2	0.13	0.13	0.16	0.16	0.12	0.16	0.14	0.13

Note. This table reports the standardized beta coefficients.
† $p < .10$.
* $p < .05$.
** $p < .01$.

Table 3
Hierarchical linear regression results of situation evaluation over the ability to identify criteria in predicting task performance and interpersonal performance (Study 2).

Variables	Student sample (N = 136)				Employee sample (N = 154)			
	Task performance		Interpersonal performance		Task performance Interpersonal performance			
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Ability to identify criteria	0.24**	0.19*	0.30**	0.23**	0.34**	0.29	0.30**	0.24**
Situation evaluation		0.22*		0.25**		0.21**		0.28**
F	8.41**	7.66**	13.29**	11.73**	19.94**	14.19**	15.40**	15.27**
R^2	0.06	0.10	0.09	0.15	0.12	0.16	0.09	0.17
ΔR^2	0.06**	0.05*	0.09**	0.06**	0.12**	0.04**	0.09**	0.08**
Adjusted R^2	0.05	0.09	0.08	0.14	0.11	0.15	0.09	0.16

Note. This table reports the standardized beta coefficients.
* $p < .05$.
** $p < .01$.

situational cues and evaluated dimensions shape response judgment. Moreover, situation evaluation significantly predicts task performance and interpersonal performance over and above response judgment and ATIC, implying that situation evaluation taps into individual differences not captured by response judgment and ATIC. It is important to note that the criterion-related validity of our SJT measuring cooperativeness for predicting task performance was almost as high as for predicting interpersonal performance. One possible explanation is that cooperativeness is an important factor contributing to high task performance when working in teams, which is increasingly common in today's work environment (Griffin et al., 2007; Lee et al., 2019).

4. Study 3

Study 3 aimed to strengthen the internal validity of our findings by comparing the incremental validity of situation evaluation over and above response judgment and ATIC in predicting work performance with two different test orders of situation evaluation and response judgment.

4.1. Method

4.1.1. Participants and procedure

Participants were 266 master students (101 men, 165 women; 123 part-time MBA students, 143 full-time students), majoring in Business Administration. As in Study 2, students voluntarily participated in this study as an optional task (for extra study credits) for a Human Resource Management course. Consistent with Study 2, our attention check showed that participants took the test seriously. Working in teams was a required part of this course. Their mean age was 28.8 years ($SD = 7.19$), and they had an average of 5.18 years ($SD = 3.79$) of internship or work experience with an average GPA of 3.07 ($SD = 0.77$). We randomly divided the participants into two groups. The first group of participants (i.e., situation evaluation first; $n = 133$) completed the CO-SJT with the *situation evaluation* item presented before response judgments. In contrast, the second group of participants (i.e., response judgment first; $n = 133$) completed the CO-SJT with the response judgments presented before the *situation evaluation* item. Data were collected online at two time points. At Time 1, participants provided their informed consent and completed a personality inventory, demographic information items, and the SJT (i.e., *situation evaluation*, response judgment, and ATIC). One month later (Time 2), without being aware of the SJT scores, the participants' team leader evaluated their task performance. Three team members rated their interpersonal performance. The two groups did not differ in demographic and personality traits, except for Openness, $t(264) = 2.57$, $p = .01$, $d = 0.32$.

4.1.2. Measures

4.1.2.1. SJT. We used the same eight-item CO-SJT as in Study 2 and generated a new version of the CO-SJT by switching the order of response judgment and situation evaluation.

4.1.2.2. Situation evaluation. We measured situation evaluation ($\alpha_{\text{situation evaluation first}} = 0.74$, $\alpha_{\text{response judgment first}} = 0.68$) using the same measure as in Study 2.

4.1.2.3. Response judgment. We assessed response judgment ($\alpha_{\text{situation evaluation first}} = 0.71$, $\alpha_{\text{response judgment first}} = 0.73$) using the same measure as in Study 2.

4.1.2.4. ATIC. We used the same *ATIC measure* ($\alpha_{\text{situation evaluation first}} = 0.74$, $\alpha_{\text{response judgment first}} = 0.72$) as in Study 2.

4.1.2.5. Work performance. We measured task performance ($\alpha_{\text{situation evaluation first}} = 0.74$, $\alpha_{\text{situation evaluation first}} = 0.71$) and interpersonal performance ($\alpha_{\text{situation evaluation first}} = 0.87$, $r_{WG(J)} = 0.88$; $\alpha_{\text{response judgment first}} = 0.88$, $r_{WG(J)} = 0.89$) using the same scales as in Study 2.

4.2. Results and discussion

4.2.1. Preliminary analyses

The means, standard deviations, and correlations of all variables in the two groups are reported in Tables S7–S8. Situation evaluation was significantly related to response judgment and ATIC in both groups. Furthermore, situation evaluation, ATIC, and response judgment were significantly associated with both task and interpersonal performance in both groups.

4.2.2. Hypothesis testing

Following the same analysis strategy as in Study 2, we found a consistent pattern of results (see Tables S9-S12 in supplementary materials) as in Study 2, supporting all the hypotheses. Moreover, we compared the regression coefficients using z-tests to test the order effect (Paternoster et al., 1998). Results showed no significant difference in incremental validity of situation evaluation over and above response judgment in predicting task performance ($z = 0.72$, $p = .47$) and interpersonal performance ($z = -0.07$, $p = .94$) between the two groups. These results provide further support for H2a and H2b. Similarly, test order did not affect the incremental validity of situation evaluation over and above ATIC in predicting task ($z = 0.46$, $p = .64$) and interpersonal performance ($z = -0.17$, $p = .86$). These results provide further support for H5a and H5b.

4.3. Additional analyses

As in Study 2, we included work experience, several personality traits, and cognitive ability as control variables in our hypothesized models to test the robustness of our findings. These analyses showed similar results, except for H4a. Specifically, ATIC explained additional variance in interpersonal performance but not in task performance over and above response judgment for the situation evaluation first group ($\beta = 0.16$, $p = .06$, $\Delta R^2 = 0.02$) and the response judgment first group ($\beta = 0.11$, $p = .16$, $\Delta R^2 = 0.01$) when the control variables were included. Furthermore, cognitive ability significantly predicted task performance ($\beta = 0.31$, $p < .01$) for the situation evaluation first group.

4.4. Discussion

Study 3 shows that situation evaluation and ATIC are significantly and positively related to response judgment, regardless of the order in which situation evaluation and response judgment are measured. Similarly, the criterion-related and incremental validity of situation evaluation, ATIC, and response judgment seems independent of the measurement order. Thus, adding construal measures does not seem to harm the criterion-related validity of response judgment but instead appears to increase the overall predictive power of a construct-driven SJT.

5. General discussion

We developed two types of situational construal: situation evaluation and ATIC based on the CO-SJT. This research shed light on the role of situational construal in arriving at a response judgment: situation evaluation and ATIC predicted response judgment and unique variance in work performance. Thus, by adding measures of situation evaluation and ATIC to an SJT, the validity of response judgment remains the same, whereas the overall criterion-related validity of the SJT increases. These findings offer several meaningful theoretical and practical implications.

5.1. Theoretical implications

This research contributes to the SJT literature in several ways. First, by differentiating between psychological construal (i.e., situation evaluation) and cognitive construal (i.e., ATIC) and showing that both types of situational construal shape response judgment, we integrate the SJT (Freudenstein et al., 2020; Rockstuhl et al., 2015) and the ATIC literature (Jansen et al., 2013; Kleinmann et al., 2011; Oostrom et al., 2016), extend TAT (Tett et al., 2021), and advance current knowledge of the thought process leading up to response judgment (e.g., Grand, 2020; Martin-Raugh & Kell, 2021). A host of research building on TAT has established that situational cues activate traits and allow candidates to express these traits through response judgment (e.g., Harris et al., 2016; Krumm et al., 2015; Lievens et al., 2021). However, which specific situational cues are most likely to activate relevant traits expressed through response judgments remains unclear. Our study suggests that cues that convey information about specific problems and their causes and possible consequences or signify implicit principles for solving a problem play an important role in the above-mentioned trait activation effect. In line with TAT, these cues serve as stimuli and provide opportunities for people to express their traits. For example, cues that signify implicit principles of cooperation (like inviting other members to join your team) serve as stimuli for cooperative activities of people with high cooperativeness because those people inherently tend to cooperate with others to solve a problem. Given that these cues seem to activate specific traits (in this case, cooperativeness), we advise SJT developers to include these situational cues in future construct-driven SJTs. Although our situational construal measures did not build on existing situational taxonomies, some of the identified situational cues showed significant overlap with the Situational Eight DIAMONDS (Rauthmann et al., 2014) and Reeder's (2013) framework (see Appendix SC in supplementary materials).

Our findings align with previous studies (Freudenstein et al., 2020; Rockstuhl et al., 2015), demonstrating that situational construal affects response judgment. However, we also extend previous studies in two important ways. First, we provide two new multiple-response approaches to measure situational construal in SJTs. These multiple-response approaches answer repeated calls for selected-response (or objective) approaches to assess situational construal (Lievens et al., 2021; Rockstuhl et al., 2015). Second, we expand existing selected-response approaches to assess situational construal by focusing on how candidates value specific trait-relevant cues in a situation. Thus far, researchers have proposed approaches based on validated situational taxonomies. For example, Freudenstein et al. (2020) proposed a selected-response measurement based on the Situational Eight DIAMONDS (e.g., duty and intellect; Rauthmann et al., 2014), and Golubovich et al. (2022) used Reeder's (2013) framework (e.g., task demands and competition/power). However, these taxonomies are rather broad and do not always apply to typical work-related SJTs (e.g., mating in the DIAMONDS taxonomy; Rauthmann et al., 2014). Furthermore, adopting Reeder's (2013) list of situational cues would mean that candidates need to complete a 43-item situational-cues inventory following each SJT item. Thus, Reeder's taxonomy seems more useful for SJT development (see Golubovich et al., 2022) than situational construal measurement. Thus, our situational construal measures provide a fine-grained understanding of the thought process leading up to response judgment while being easily applicable.

Second, by showing that situation evaluation and ATIC incrementally predict work performance over and above response judgment, we improve the understanding of how SJTs work and offer a potential means to increase their criterion-related validity. Although a range of factors can contribute to performance in work-related situations (Lievens & Motowidlo, 2016; Rockstuhl et al., 2015), this study zoomed in on an often neglected predictor: situational construal. We showed that situation evaluation and ATIC explain unique variances in work performance. Thus, adding a measure of these two types of situational construal to a construct-driven SJT allows for a more thorough assessment of job-relevant characteristics, which can increase the overall criterion-related validity of the SJT. As such, the present study also contributes to the current debate on the role of situation descriptions in construct-driven SJTs by providing further evidence on the importance of situational construal in SJTs (Krumm et al., 2015; Schäpers, Mussel, et al., 2020). Dropping situation descriptions to shorten items could seem appealing. However, the validity gains we observed for our SJT by adding situational construal measures suggest that capturing candidates' perceptions of situation descriptions might be a better way to improve SJTs than removing these descriptions.

Finally, we contribute to the ATIC literature (Jansen et al., 2013; Kleinmann et al., 2011; Oostrom et al., 2016) by showing that ATIC has more incremental validity over and above response judgment in predicting interpersonal than task performances. Considering that ATIC captures cognitive construal, it shares common variance with cognitive ability (Melchers et al., 2020), which may explain why the incremental validity of ATIC in predicting task performance was further reduced when controlling for a cognitive ability measure (Kleinmann et al., 2011). It could also be that our construct-driven SJT was relatively transparent, which may have limited the variance in ATIC. Thus, replicating these findings using less transparent simulation-based predictor methods is important.

5.2. Practical implications

This research offers several practical implications. First, we advise practitioners to incorporate situation evaluation and ATIC when using construct-driven SJTs in personnel selection contexts. Thus, the SJT incorporates the psychological interpretation of situational cues for specific traits and a general ability to identify evaluated dimensions and may therefore have higher validity for predicting work performance. Furthermore, incorporating situation evaluation items and ATIC items in an SJT requires candidates to report their feelings about situational cues, which could make candidates react more favorably to SJTs because they have more opportunities to demonstrate their skills (Rolland & Steiner, 2007). However, some test-takers may dislike longer assessments, leading to higher attrition rates (Hardy et al., 2017). This implication should be tested in future research.

Second, this research provides some evidence-based advice for practitioners on generating situational construal measures (see Study 1). Specifically, practitioners can develop situational construal measures for each item by asking experts to a) identify situational cues that convey unique information as possible response options for situation evaluation, b) determine targeted dimensions and distractors as possible response options for ATIC, and c) generate scoring keys for these options. Such situational construal measures are easy to score and implement in large-scale testing programs, making them attractive options for early recruitment and selection stages (Rockstuhl et al., 2015). We also recommend the application of situation evaluation and ATIC to other types of construct-driven SJTs (e.g., video-based SJTs) as well as other simulation-based predictors, such as assessment centers (Schollaert & Lievens, 2012) and situational interviews (Oostrom et al., 2016), to assess candidates' qualities better. However, this implication demands further investigation in future research because the validity of our situational construal measures may vary owing to differences between these simulated-based instruments (e.g., fidelity and test length).

5.3. Limitations and suggestions for future research

Despite the abovementioned implications, several limitations should be noted and addressed in future research. First, our research echoes Lievens et al.'s (2021) recommendation that “SJTs can be extended for use as a novel approach for assessing situation construal” (p. 298). However, we proposed situational construal measures for a construct-driven SJT measuring a single personality facet. We focused on a construct-driven SJT because these types of SJTs are often designed based on TAT (e.g., Golubovich et al., 2022; Oostrom et al., 2019). However, given that all SJTs, including traditional SJTs, contain situational cues, we suggest that the two types of situational construal could predict response judgment and work performance in various SJTs. However, we expected that the validity of situational construal might be contingent on the extent to which the SJTs contain situational cues that activate individual differences. Other potential boundary conditions may be the number of situational cues and the format of the response options. Indeed, Krumm et al. (2015) argued that highly contextualized items might be more suited to measure situational construal than generic items. Future research is needed to test these predictions and strengthen our findings' generalizability to other SJTs.

Second, we showed the value of measuring situation evaluation and ATIC in work-related and academic contexts but not in selection contexts. Previous has research suggested that findings based on incumbent and academic samples provide theoretical and practical implications for personnel selection practices (e.g., Krumm et al., 2015; Rockstuhl et al., 2015). However, just like would-do SJTs are answered by candidates as should-do SJTs in high-stakes settings (Weekley et al., 2015), cognitive construal (ATIC) and psychological construal (situation evaluation) may also show more overlap when candidates have a strong motivation to perform well (Ingold et al., 2015). Thus, to generalize our findings further, specifically concerning the incremental validity of situation evaluation and ATIC, we suggest that future research apply these situational construal measures in real selection contexts.

CRediT authorship contribution statement

Dan Wang: Conceptualization, methodology, data collection and analysis, writing – original draft.

Janneke K. Oostrom: Conceptualization, supervision, writing – original draft, review, and editing.

Eveline Schollaert: Conceptualization, Writing – review and editing.

Dan Wang, Janneke K. Oostrom, and Eveline Schollaert: Reviewed and approved the final version of the manuscript for submission.

Declaration of competing interest

None.

Data availability statement

The data set is available from the corresponding author on reasonable request via e-mail.

Appendix A. Sample item measuring situation evaluation, response judgment, and ATIC

Situational stem	A) Your superior has assigned you an urgent task. B) You come up with a preliminary plan, but feel that the plan may have potential problems. C) However, you don't know where the problem is. D) You are worried that these potential problems may cause trouble. E) You can organize a meeting to discuss your plan, but that takes time. What would you do?
Situation evaluation	Which situational cue do you consider to be the most crucial in the above situation? A B C D E (1 2 2 3 1)
Response judgment	Please rate the following options to indicate the likelihood that you would respond that way in the above situation on a 6-point type scale (1 = *very unlikely*, 6 = *very likely*). For all three options, we ask you to indicate the extent to which you would respond that way to the given situation. A. *Carry out the task according to the preliminary plan and promptly deal with any problems found.* (Reverse scored) B. Organize a meeting to discuss and eliminate potential problems of the preliminary plan before carrying out the task. C. Encourage any plans or suggestions from team members and carry out the task according to a fairly completed plan.
ATIC	Which dimension do you think is most likely to be evaluated in the above item? A. Cooperativeness (1) B. Leadership (0) C. Problem Solving Skills (0) D. Responsibility (0) E. Other Skills or Ability (0)

Note. Scoring keys are in brackets and the response option in italic was reverse scored.

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.paid.2023.112182.

References

Bradley, B. H., Baur, J. E., Banford, C. G., & Postlethwaite, B. E. (2013). Team players and collective performance: How agreeableness affects team performance over time. *Small Group Research, 44*(6), 680–711. https://doi.org/10.1177/1046496413507609

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-Perspectives on Science and Practice, 9*(1), 38–42. https://doi.org/10.1017/iop.2015.113

Fan, J., Stuhlman, M., Chen, L., & Weng, Q. (2016). Both general domain knowledge and situation assessment are needed to better understand how SJTs work. *Industrial and Organizational Psychology, 9*(1), 43–47. https://doi.org/10.1017/iop.2015.114

Ferris, G. R., Perrewé, P. L., & Douglas, C. (2002). Social effectiveness in organizations: Construct validity and research directions. *Journal of Leadership & Organizational Studies, 9*(1), 49–63. https://doi.org/10.1177/107179190200900104

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

Funder, D. C. (2016). Taking situations seriously: The situation construal model and the Riverside Situational Q-Sort. *Current Directions in Psychological Science, 25*(3), 203–208. https://doi.org/10.1177/0963721416635552

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

Griffin, M. A., Neal, A., & Parker, S. K. (2007). A new model of work role performance: Positive behavior in uncertain and interdependent contexts. *Academy of Management Journal, 50*(2), 327–347. https://doi.org/10.5465/amj.2007.24634438

Golubovich, J., Lake, C. J., Seybert, J., & Anguiano-Carrasco, C. (2022). Do situational characteristics affect the validity of personality situational judgment items? *International Journal of Selection and Assessment, 30*(3), 427–447. https://doi.org/10.1111/ijsa.12390

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

Hardy, J. H., III, Gibson, C., Sloan, M., & Carr, A. (2017). Are applicants more likely to quit longer assessments? Examining the effect of assessment length on applicant attrition behavior. *Journal of Applied Psychology, 102*(7), 1148–1158. https://doi.org/10.1037/apl0000213

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

Ingold, P. V., Kleinmann, M., König, C. J., Melchers, K. G., & van Iddekinge, C. H. (2015). Why do situational interviews predict job performance? The role of interviewees' ability to identify criteria. *Journal of Business and Psychology, 30*(2), 387–398. https://doi.org/10.1007/s10869-014-9368-3

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

Kleinmann, M. (1993). Are rating dimensions in assessment centers transparent for participants? Consequences for criterion and construct validity. *Journal of Applied Psychology, 78*(6), 988–993. https://doi.org/10.1037/0021-9010.78.6.988

Kleinmann, M., Ingold, P. V., Lievens, F., Jansen, A., Melchers, K. G., & Konig, C. J. (2011). A different look at why selection procedures work: The role of candidates' ability to identify criteria. *Organizational Psychology Review, 1*(2), 128–146. https://doi.org/10.1177/2041386610387000

Koschmieder, C., & Neubauer, A. C. (2021). Measuring emotion regulation for preservice teacher selection: A theory-driven development of a situational judgment test. *Personality and Individual Differences, 168*(3), 1–10. https://doi.org/10.1016/j.paid.2020.110363

Krumm, S., Lievens, F., Huffmeier, 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

Lee, Y., Berry, C. M., & Gonzalez-Mulé, E. (2019). The importance of being humble: A meta-analysis and incremental validity analysis of the relationship between honesty-humility and job performance. *Journal of Applied Psychology, 104*(12), 1535–1546. https://doi.org/10.1037/apl0000421

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., Schäpers, P., & Herde, C. N. (2021). Situational judgment tests: From low-fidelity simulations to alternative measures of personality and the person-situation interplay. In D. Wood, P. Harms, S. Read, & A. Slaughter (Eds.), *Emerging approaches to measuring and modeling the person and situation* (pp. 285–311). Elsevier.

Marshall, B., Cardon, P., Poddar, A., & Fontenot, R. (2013). Does sample size matter in qualitative research?: A review of qualitative interviews in is research. *Journal of Computer Information Systems, 54*(1), 11–22. https://doi.org/10.1080/08874417.2013.11645667

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

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

Melchers, K. G., Bill, B., Buehl, A.-K., Rybczynski, K., & Kühnel, J. (2020). Identification of the targeted performance dimensions and self-promotion in interviews: Investigations of uncharted waters. *European Journal of Work and Organizational Psychology, 29*(5), 728–739. https://doi.org/10.1080/1359432X.2020.1746828

Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychological Review, 80*(4), 252–283. https://doi.org/10.1037/h0035002

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

Murano, D., Lipnevich, A. A., Walton, K. E., Burrus, J., Way, J. D., & Anguiano-Carrasco, C. (2021). Measuring social and emotional skills in elementary students: Development of self-report Likert, situational judgment test, and forced choice items. *Personality and Individual Differences, 169*, Article 110012. https://doi.org/10.1016/j.paid.2020.110012

Mussel, P., Rodrigues, J., Krumm, S., & Hewig, J. (2018). The convergent validity of five dispositional greed scales. *Personality and Individual Differences, 131*(2), 249–253. https://doi.org/10.1016/j.paid.2018.05.006

Neubauer, A. C., & Hofer, G. (2022). (Retest-) reliable and valid despite low alphas? An example from a typical performance situational judgment test of emotional management. *Personality and Individual Differences, 189*, Article 111511. https://doi.org/10.1016/j.paid.2022.111511

Oostrom, J. K., Melchers, K. G., Ingold, P. V., & Kleinmann, M. (2016). Why do situational interviews predict performance? Is it saying how you would behave or knowing how you should behave? *Journal of Business and Psychology, 31*(2), 279–291. https://doi.org/10.1007/s10869-015-9410-0

Oostrom, J. K., de Vries, R. E., & de Wit, M. (2019). Development and validation of a HEXACO situational judgment test. *Human Performance, 32*(1), 1–29. https://doi.org/10.1080/08959285.2018.1539856

Paternoster, R., Brame, R., Mazerolle, P., & Piquero, A. (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology, 36*(4), 859–866. https://doi.org/10.1111/j.1745-9125.1998.tb01268.x

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

Rauthmann, J. F., Sherman, R. A., & Funder, D. C. (2015). Principles of situation research: Towards a better understanding of psychological situations. *European Journal of Personality, 29*(3), 363–381. https://doi.org/10.1002/per.1994

Reeder, M. C. (2013). *Situational judgment tests and psychologically active characteristics ofsituations: A dimensional approach to analyzing situational judgment test content and itspsychometric implications [Unpublished doctoral dissertation]*. Michigan State University.

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

Rolland, F., & Steiner, D. D. (2007). Test-taker reactions to the selection process: Effects of outcome favorability, explanations, and voice on fairness perceptions. *Journal of Applied Social Psychology, 37*(12), 2800–2826. https://doi.org/10.1111/j.1559-1816.2007.00282.x

Sackett, P. R., & Walmsley, P. T. (2014). Which personality attributes are most important in the workplace? *Perspectives on Psychological Science, 9*(5), 538–551. https://doi.org/10.1177/1745691614543972

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

Sherman, R. A., Rauthmann, J. F., Brown, N. A., Serfass, D. G., & Jones, A. B. (2015). The independent effects of personality and situations on real-time expressions of behavior and emotion. *Journal of Personality and Social Psychology, 109*(5), 872–888. https://doi.org/10.1037/pspp0000036

Speer, A. B., Christiansen, N. D., Melchers, K. G., Konig, C. J., & Kleinmann, M. (2014). Establishing the cross-situational convergence of the ability to identify criteria: Consistency and prediction across similar and dissimilar assessment center exercises. *Human Performance, 27*(1), 44–60. https://doi.org/10.1080/08959285.2013.854364

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

Tett, R. P., & Guterman, H. A. (2000). Situation trait relevance, trait expression, and cross-situational consistency: Testing a principle of trait activation. *Journal of Research in Personality, 34*(4), 397–423. https://doi.org/10.1006/jrpe.2000.2292

Tett, R. P., Toich, M. J., & Ozkum, S. B. (2021). Trait activation theory: A review of the literature and applications to five lines of personality dynamics research. *Annual Review of Organizational Psychology and Organizational Behavior, 8*(1), 199–233. https://doi.org/10.1146/annurev-orgpsych-012420-062228

Wang, D. (2022, November 19-20). *The development of a simulation-based measure of cooperativeness: A construct-driven situational judgment test [Paper presentation]*. Changsha, Hunan, China: Human Resource Development of China Seventh Annual Meeting.

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

Ziegler, M., Horstmann, K. T., & Ziegler, J. (2019). Personality in situations: Going beyond the OCEAN and introducing the situation five. *Psychological Assessment, 31*(4), 567–580. https://doi.org/10.1037/pas0000654