1-2021

Prompt-specificity in scenario-based assessments: Associations with personality versus knowledge and effects on predictive validity

Thomas ROCKSTUHL
Nanyang Technological University

Filip LIEVENS
Singapore Management University, filiplievens@smu.edu.sg

Running Head: PROMPT-SPECIFICITY

Prompt-Specificity in Scenario-based Assessments: Associations with Personality vs. Knowledge and Effects on Predictive Validity

Thomas Rockstuhl

Nanyang Technological University, Singapore

Filip Lievens

Singapore Management University, Singapore

Author Note

Thomas Rockstuhl, Nanyang Business School, Nanyang Technological University, Singapore. Filip Lievens, Lee Kong Chian School of Business, Singapore Management University, Singapore.

Correspondence concerning this article should be addressed to Thomas Rockstuhl, Division of Leadership, Management, Organization, Nanyang Technological University, Singapore 639798. E-mail: TRockstuhl@ntu.edu.sg

This is the final author version (before journal's typesetting and copyediting) It is not the copy of record and may not exactly replicate the final, authoritative version of the paper.
DOI: 10.1037/apl0000498

Abstract

Many scenario-based assessments (e.g., interviews, assessment center exercises, work samples, simulations, situational judgment tests) use prompts (i.e., cues provided to respondents to increase the likelihood that the information received from them is clear, sufficient, and job-related). However, a dilemma for practitioners and researchers is how general or specific one should prompt people's answers. We posit that such differences in prompt specificity (i.e., extent to which prompts cue performance criteria) have important implications for the predictive validity of scenario-based assessment scores. Drawing on the interplay of situation construal and situational strength theory, we propose that prompt-specificity leads to differential relationships between scenario-based scores and external constructs (personality traits vs. knowledge), which in turn affects the predictive validity of scenario-based assessments. We tested this general hypothesis using intercultural scenarios for predicting effectiveness in multicultural teams. Using a randomized predictive validation design, we contrast scores on these scenarios with general (N=157) versus specific (N=158) prompts. As a general conclusion, prompt-specificity mattered: Lesser prompt-specificity augmented the role of perspective taking and openness-to-experience in the intercultural scenario scores and their validity for predicting intercultural performance, whereas greater prompt-specificity increased the role of knowledge in these scores and their validity for predicting in-role performance. This study's theoretical and practical implications go beyond a specific assessment procedure and apply to a broad array of assessment and training approaches that rely on scenarios.

KEYWORDS: Simulations, situational judgment tests, prompts, scenarios, intercultural performance

Prompt-Specificity in Scenario-based Assessments: Associations with Personality vs. Knowledge and Effects on Predictive Validity

Scenario-based assessments are ubiquitous in research, selection, and training contexts. Examples of scenario-based measurements can be found in experimental vignettes (Aguinis & Bradley, 2014), situational interviews (Latham, Saari, Pursell, & Campion, 1980), situational judgment tests (Motowidlo, Dunnette, & Carter, 1990), training evaluation (Bhawuk & Brislin, 2000; Hauenstein, Findlay, & McDonald, 2010; Ostroff, 1991) as well as assessment center exercises, work samples, and simulations (Thornton & Cleveland, 1990). On the basis of the notion of behavioral consistency, scenario-based measurements assume that people's responses to the scenario will mirror their behavior in actual job situations (Lievens & De Soete, 2015; Schmitt & Ostroff, 1986; Wernimont & Campbell, 1968).

To ensure that scenario-based assessments elicit sufficient and job-relevant behavior and thus lead to reliable and valid scores (Levashina, Hartwell, Morgeson, & Campion, 2014; Lievens, Schollaert, & Keen, 2015), they often include prompts (i.e., cues provided to respondents). Messick (1998) identified prompt-specificity (degree to which prompts provide cues about the criteria used to evaluate performance) as a critical dimension on which prompts might vary. Indeed, a dilemma often faced by researchers and practitioners is whether one should use vague/general prompts (e.g., How do/did you solve this issue?) or more specific prompts (e.g., How do/did you solve this issue to ensure that people are motivated in the long run?). If one uses a general and vague prompt and does not provide cues about the criteria used to evaluate performance, one might get insight in how people would spontaneously express their trait(s) in responding to the scenario (Blackman & Funder, 2002). Yet, if people fail to give a response after a general prompt, it remains unclear whether they did not know how to adequately

respond to the scenario or whether they simply did not know what was expected from them. Conversely, if one uses a specific prompt and thus clarifies what is expected, one might get information about people's capabilities to respond in line with the performance criteria but one risks not getting insight in whether people also spontaneously respond like this in "unprompted" real life, or worse yet, one might give away the right answer by suggesting these criteria.

These two perspectives underscore how differences in prompt-specificity might have important implications for the constructs assessed and the predictive power of scenario-based assessments. In this paper, we shift the attention from contrasting these perspectives to better understanding "when" and "why" they might be valid. Specifically, we posit and explain that the above variations in prompt-specificity lead to valid predictions but for different outcomes. Our study's central premise is that variations in prompt-specificity increase or attenuate the relationship of individual differences constructs (personality traits vs. knowledge) with scenario scores and that these differential relationships in turn influence the predictive validity of scenario-based assessments.

This study contributes to both theory and practice on scenario-based assessments. Theoretically, we advance understanding of how prompt-specificity impacts the predictive validity of scenario-based assessments. In particular, drawing on the interplay of situation construal (Funder, 2016) and situational strength theory (Meyer, Dalal, & Hermida, 2010; Mischel, 1968; 1973) we highlight the key importance of the kind of individual differences being activated for explaining the effects of prompt-specificity. For practice, our study sheds light on when and why different prompts are associated with the outcomes predicted by scenario-based assessments. Thus, our findings can inform choices about desired levels of prompt-specificity

based on intended criteria. Importantly, our findings go beyond a specific selection procedure because scenarios are used in a variety of instruments, contexts, and domains.

Our examination of prompt-specificity takes place in the context of intercultural scenarios. Intercultural scenarios refer to examples of culture clashes between individuals from different cultural backgrounds (Brislin, 2009; Cushner & Landis, 1996). These intercultural scenarios show participants a short video-clip of a challenging intercultural interaction in the workplace and ask them to suggest solutions for the dilemma depicted in the scenario (see Table 1 for a description of an example scenario and potential open-ended responses). Conceptually, the complex and multidimensional nature of responses to intercultural scenarios provides an ideal context for studying how variations in prompt-specificity increase or dampen the relationship of individual differences constructs with scenario scores and the predictive validity of these scores. Moreover, with the growing diversity in the workplace, intercultural scenarios may increasingly form the basis of selection and training procedures.

Theory and Hypotheses

Prompts: Definition and Underlying Rationale

In experimental psychology, a prompt is defined as "an instruction to attend to a particular aspect of a multidimensional stimulus or to carry out a particular operation on it" (Hartley, Kieley, & Slabach, 1990, p. 524), thus mentioning that prompts inform participants about the basis for a correct response (Hartley et al., 1990; Schwartz, 1999). Recently, scholars studying different scenario-based methods have put forth more specific definitions of prompts. For example, in assessment centers, Schollaert and Lievens (2012) defined them as "predetermined verbal and nonverbal cues that a role-player consistently provides during the AC exercise across candidates to elicit job-related behavior" (p. 258). In interviews, Levashina et al. (2014) referred

to a prompt as a "follow-up question that is intended to augment an inadequate or incomplete response provided by the applicant, or to seek additional or clarifying information" (p. 271).

In scenario-based measurements, prompts are typically used to overcome a lack of clarity on the information already received or to increase the amount/availability of job-relevant information. For example, Lievens et al. (2015) recommended planting prompts in assessment center exercises "as a systematic and efficient tool for increasing the frequency of behavior relevant to focal constructs" (p.1183). Similarly, Levashina et al. (2014) suggested that prompts may "help applicants who might be shy or speak in succinct ways to clarify their answers and to provide more detailed job-related information" (p. 272). So, by using prompts one aims to enhance the clarity, quantity, and/or quality (higher job relevance) of the information gathered. In turn, the availability of such more job-relevant information should permit evaluators (interviewers, assessors) to increase the reliability and validity of their ratings for predicting people's future performance in real world settings.

In short, on the basis of these definitions and rationales underlying prompts, we refer to prompts as cues (i.e., hints or indications about how to behave) provided to increase the likelihood that the information received from respondents is clear, sufficient, and job-related.

Hypotheses About Effects of Prompt Specificity

As noted above, a key decision when dealing with prompts relates to their level of specificity. As posited by Messick (1998), general prompts provide no or ambiguous cues about evaluation criteria, whereas specific prompts cue behavioral demands and performance expectations. A study by Zaccaro, Mumford, Connelly, Marks, and Gilbert (2000) exemplified this distinction. In one condition, planning was more generally prompted, whereas in the other condition respondents received cues about the evaluation criteria on which their plan would be

assessed (e.g., “How would you like to see your plan carried out?” or “How would you be sure that your plan was carried out correctly?”). Although Messick highlighted the role of prompt-specificity 20 years ago, our understanding of the effects of prompt-specificity for embedding prompts in scenario-based measurements is still limited.

Situation construal theory and scenario-based assessment. To contribute to our understanding of prompt-specificity in scenario-based assessment, we start by grounding scenario-based responses in situation construal theory (Funder, 2016). In general terms, situation construal theory states that person as well as situation variables influence the way in which people perceive situations. The situation side refers to the “objective” situation (also referred to as the canonico-consensual situation; Block & Block, 1981). This is the situation (in this case the scenario) as agreed upon by many people. The person side refers to individual differences variables such as people’s personality, knowledge, cognitive ability or emotional intelligence. Although all these individual differences variables might influence how people construe situations, personality has been identified as a particularly potent driver of situation construal. Sherman, Nave, and Funder (2013) found evidence for a link between personality and situation construal of everyday situations for all Big Five factors (e.g., people high on agreeableness construed a situation such as cycling with others as an opportunity to chat and get along, whereas people high on achievement striving perceived such a situation as competitive).

So, we propose that one construes the scenario in a unique way depending on the objective situation and one’s standing on individual differences variables (Allport, 1961; Reis, 2008; Sherman et al., 2013). Situation construal is then defined as a person’s distinctive perception of the situation (i.e., the psychological situation, Block & Block, 1981; Mischel & Shoda, 1995;

Rauthman, Sherman, & Funder, 2015). Apart from main person and situation effects, such situation construal influences subsequent behavioral actions and responses.

Situational strength theory and prompt-specificity. How does prompt-specificity come into play in the process we have just put forward? To better understand prompt-specificity, we link it to situational strength theory (Meyer et al., 2010; Mischel, 1968; 1973). Situational strength is inherently related to prompt-specificity because situational strength refers to "implicit or explicit cues provided by external entities regarding the desirability of potential behaviors" (Meyer et al., 2010, p. 122). In this study, we contrast general prompts (low prompt-specificity) to specific prompts (high prompt-specificity). In the specific prompt condition and consistent with Messick (1998), we embed specificity by providing cues about performance criteria when prompting responses to intercultural scenarios (e.g., "What would you do *to both complete the task and maintain the relationship? Consider both parties' perspectives.*").

When prompts include cues about performance criteria, respondents have a clearer understanding of what is to be expected (e.g., Klehe, König, Richter, Kleinmann, & Melchers, 2008). Hence, the behavioral demands of the scenarios become less ambiguous. Cueing respondents about the criteria used for evaluating their responses to the scenarios thus increases the situational strength of the scenarios. Conversely, in the general prompt condition (e.g., "What would you do?"), cues about performance criteria remain absent. So, the response demands remain ambiguous to respondents (Tourangeau, Rips, & Rusinski, 2000), thereby decreasing the situational strength.

Effects of prompt-specificity on the relationship between individual differences constructs and scenario scores. A basic tenet of situational strength theory is that situational characteristics like prompts systematically influence behavior by either reducing (in case of

strong situations) or increasing (in case of weak situations) the expression of personality differences (Cooper & Withey, 2009). As detailed below, this forms the basis for our premise that prompt-specificity will affect the relevance of personality (vs. knowledge), which in turn will impact the validity of scenario-based assessments.

According to situational strength theory, the condition with general prompts (weak situations) leaves situational demands ambiguous. By increasing the ambiguity of situational demands, general prompts underscore the importance of adequate situation construal, of personality's impact on situation construal, and of greater expression of personality trait differences in responses. In the context of intercultural scenarios, we expect two specific traits, namely openness to experience and perspective taking, to play an important role in determining people's situation construal and responses.

First, openness to experience describes people in terms of their being cultured, broadminded, creative (McCrae, 1987), flexible (Ones & Viswesvaran, 1997), and more willing to embrace novel ideas (Barrick & Mount, 1991). Openness is relevant to responding to intercultural scenarios because it relates to a person's tendencies when faced with different cultural preferences (Ang, Van Dyne, & Koh, 2006; Shaffer, Harrison, Gregersen, Black, & Ferzandi, 2006). Individuals high in openness to experience are known to respond to different cultural preferences with greater tolerance and less ethnocentrism. Three lines of research support this perspective. First, people high in openness tend to be more accepting of both similarities and differences among people (Albrecht, Dilchert, Deller, & Paulus, 2014). Second, people high in openness typically score low on right-wing authoritarianism (Cohrs, Kämpfe-Hargrave, & Riemann, 2012) and other conservative values (Jost, Glaser, Kruglanski, & Sulloway, 2003) that have been linked to ethnocentrism (Sibley & Duckitt, 2008). Third, people

high in openness are likely to seek out new information and experiences that challenge the status quo, whereas people low in openness are seem to freeze on their opinions (McCrae, 1987).

In sum, with general prompts, people low in openness to experience may be more likely to construe the situational demands of intercultural scenarios as allowing ethnocentric solutions that do not benefit intercultural relationships, whereas people high in openness to experience are less likely to do so. Compare this to the specific prompt situation with explicit instructions ("to complete the task and maintain the relationship"). In that condition, we hypothesize individual differences in openness to experience to be less relevant. So, in that condition, the relationship between openness to experience and intercultural scenario-performance will be attenuated as compared to the general prompt condition. Thus, lower prompt-specificity should increase the relation between openness-to-experience and intercultural scenario-performance.

Second, perspective taking refers to a person's tendency to adopt spontaneously others' psychological point of view (Davis, 1983). Perspective taking is relevant to intercultural scenarios because it relates to a person's tendencies when faced with intercultural differences (Triandis, 2006). As people high in perspective taking believe that there are two sides to every conflict and try to look at both of them (Davis, 1983), they should be more likely to consider both parties' perspective in an intercultural conflict than people low in perspective taking, even when not explicitly told to do so. Compare this to the specific prompt situation with explicit instructions to consider both parties' perspectives. In that condition, we hypothesize individual differences in perspective taking to be less relevant.

In addition, a large body of research suggests that perspective taking, like openness to experience, should be negatively associated with ethnocentric responses to intercultural scenarios (Ku, Wang, & Galinsky, 2015; Pettigrew & Tropp, 2008). Perspective taking reduces

ethnocentric responding by increasing (a) liking for perspective taking targets (Davis, Conklin, Smith, & Luce, 1996), (b) a sense of psychological closeness with diverse others (Cialdini, Brown, Lewis, Luce, & Neuberg, 1997), and (c) approach behavior (Todd & Galinski, 2014).

As with openness to experience, we expect individual differences in perspective taking to affect the construal of situational demands primarily in the general prompt condition. In particular, without specific prompts, people high in perspective taking may be more likely to construe the situational demands of intercultural scenarios as requiring solutions for both parties and as avoiding ethnocentric solutions. By contrast, individual differences in perspective taking should be less relevant in the specific prompt condition with explicit instructions to consider both parties' perspectives and to complete the task and maintain the relationship.

Conversely, the condition with specific prompts (strong situations) makes the situational demands considerably less ambiguous. By reducing the ambiguity of situational demands, specific prompts restrict the range of behavior and attenuate the relevance of personality differences for behavior. Thus, we hypothesize that the effects of personality on situational construal and subsequent behavior will be reduced. So, greater prompt-specificity should decrease the relation between personality traits and scenario-scores.

Against the backdrop of the above, we posit the following hypothesis:

> *Hypothesis 1: Prompt specificity moderates the positive relationships of openness to experience (H1a) and perspective taking (H1b) with performance in intercultural scenarios, such that these relationships are weaker in the specific compared to the general prompt condition.*

Due to the strong situation invoked by specific prompts as well as the reduced roles of personality in affecting situation construal, other individual differences will play a more important role. In particular, the provision of specific cues about performance criteria enables people to rely on their performance-related knowledge to respond to the scenario. Indeed,

theories of knowledge-based inferences (Graesser, Singer, & Trabasso, 1994) posit that respondents' background knowledge is elicited by situational demands (in this case prompts) via pattern recognition processes and that they rely on this knowledge to construct possible responses to scenarios in accordance with the situational demands. Such background knowledge (e.g., schemata, scripts or stereotypes about this or similar situations that are grounded in experience) then provides contextually rich content needed to guide their possible responses. For example, Ployhart (2006, p.88) notes that once respondents understand the situational demands of a scenario, they engage in "the retrieval of information from long-term memory that is relevant to the question." This activation of knowledge that is *relevant to the situation demands* of the scenario provides the basis for hypothesizing that background knowledge will be related to scenario-scores to a greater extent in the specific than the general prompt condition.

In this study, specific prompts make explicit to respondents that the situational demands involve a need to resolve the cultural dilemma (i.e., complete the task *and* maintain the relationship) and to consider both parties' perspective. Knowledge about cultural value differences is particularly relevant to resolving cultural dilemmas from the perspective of both parties (Bhawuk, 2017, 2001). Such knowledge is relevant because resolving cultural dilemmas requires integrating information about why certain cultures exhibit specific behaviors (Triandis, 2006). By making the performance criteria explicit, the specific prompt condition cues people about the need to match observed cultural behaviors to knowledge about how cultural value differences influence such behavior. Thus, in our study, we expect that specific prompts will amplify the relevance of knowledge about cultural value differences for scenario responses.

By contrast, when performance criteria in the general prompt condition are vague, people may construe situational demands in more varied ways. When people construe the situation to

have different demands, they will in turn activate different types of knowledge, thus diminishing the relevance of cultural knowledge across respondents in this condition. For example, general prompts ("What would you do?") may cue some people to retrieve knowledge about how they responded in the past, rather than knowledge about how cultural value differences affect the behavior of the parties in the intercultural scenario. Thus, general prompts will attenuate the relevance of knowledge on cultural value differences for scenario responses.

> *Hypothesis 2: Prompt specificity moderates the positive relationship between cultural knowledge and performance in intercultural scenarios, such that the relationship is stronger in the specific compared to the general prompt condition.*

Effects on performance and validity. In the previous section, we formulated hypotheses about how variations in prompt-specificity lead to differential relations between external constructs (personality vs. knowledge) and scenario-based scores. The next question then becomes whether these effects also matter. That is: Do they translate into differences in the predictive validity of performance? As noted above, this study goes beyond considering which perspective is better and shows that both perspectives can be predictive, albeit for different outcomes. So, the premise of our study is that different outcomes can be predicted when either general or specific prompts are used because of the kind of individual differences being amplified or dampened in solving scenario-based assessments.

Conceptually, our hypotheses about the effects of prompt specificity on performance components and about the role of specific individual differences (personality vs. knowledge) in this link draw upon Ajzen's (2005) matching principle and on the predictor-criterion matching logic (Lievens, Buyse, & Sackett, 2005). Both notions posit that better prediction can be obtained when predictors and criteria are conceptually matched (Lievens et al., 2005; see also Bartram, 2005; Hogan & Holland, 2003). As noted above, more specific prompts increase the role of

cultural knowledge in completing intercultural scenarios, whereas more general prompts increase the role of two traits (openness-to-experience and perspective-taking) in solving intercultural scenarios. So, drawing on the predictor-criterion matching logic, we expect scenario-scores under specific prompts to be more related to criteria that have a record of being predicted by knowledge constructs, whereas we anticipate scenario-scores under general prompts to be more associated with criteria that are predicted by personality traits (in this case openness to experience and perspective taking).

To test all of this, we focused on three criteria. First, given the intercultural context, we investigated the effects on intercultural performance. With the continuing globalization of the workplace, intercultural performance has gained increasing attention among management scholars (Caliguiri, 2006). Intercultural performance can be defined as effectiveness in completing tasks and maintaining strong intercultural relationships (Caliguiri, 2006; Leung, Ang, & Tan, 2014). Evidence suggests that intercultural performance is conceptually distinct from general or domestic performance (e.g., Chen, Liu, & Portnoy, 2012). Intercultural performance is challenging because culture influences expectations about the type of behavior in contexts (Triandis, 2006) as diverse as leadership (House, Hanges, Javidan, Dorfman, & Gupta, 2004), relationships (Yeung & Ready, 1995) or teams (Gibson & Zellmer-Bruhn, 2001). To function effectively across cultures, it is therefore imperative to adjust to culturally diverse others' expectations (Stone-Romero, Stone, & Salas, 2003). This requires not only a tendency to suspend judgment, but also a tendency to seek out additional cues to make sense of culturally ambiguous behavior (Triandis, 2006).

If intercultural performance serves as prediction focus, we anticipate that scenario-scores in the general prompt condition will lead to better predictions of intercultural performance than

scores in the specific prompt condition because scenario-scores based on general prompts are expected to be more related to two personality traits that have been linked to intercultural performance, namely openness to experience (Leung et al., 2014; Lievens, Harris, Van Keer, & Bisqueret, 2003) and perspective taking (Molinsky, 2013; Triandis, 2006). Both perspective taking and openness to experience facilitate suspending judgment and seeking out such additional cues. For example, Parker, Atkins, and Axtell (2008) suggest that perspective taking involves an observer who tries to perceive "*in a nonjudgmental way* (emphasis added), the thoughts, motives, and/or feelings of a target, as well as why they think and/or feel the way they do'' (p. 151). Thus, people high in perspective taking are likely to suspend judgment and adjust to culturally diverse others' expectations in intercultural encounters. In a similar vein, Shaffer and colleagues noted that people high in openness to experience are curious and eager to learn about cultural differences, and therefore will be likely to suspend premature judgments (Shaffer et al., 2006; see also Leiba-O'Sullivan, 1999). By contrast, cultural knowledge may be less facilitative to intercultural performance because research suggests that cultural knowledge can be associated with premature or sophisticated stereotyping (Osland & Bird, 2000; Rockstuhl & Van Dyne, 2018), rather than the suspending of judgment and seeking out of additional cues. In sum, both openness to experience and perspective taking are crucial to intercultural performance because they facilitate adaptation to culturally diverse others' expectations. Thus, we propose:

Hypothesis 3a: Prompt specificity moderates the positive relationship between intercultural scenario performance and intercultural performance, such that the relationship is weaker in the specific compared to the general prompt condition.

Hypothesis 3b: The indirect effects of openness to experience and perspective taking on intercultural performance via intercultural scenario performance will be stronger in the general prompt condition than the specific prompt condition.

Apart from intercultural performance, this study also includes two important traditional performance components, namely in-role (i.e., task) and extra-role (i.e., citizenship) performance. We included these criteria because they enable us to test our premise that depending on the criterion that one aims to predict, one might adjust prompt-specificity accordingly. In-role behaviors refer to behaviors that are recognized by formal reward systems and are part of the requirements as described in job descriptions (Williams & Anderson, 1991). That is, in-role performance requires one to carry out one's duties and to deliver work of good quantity and quality. Job-specific knowledge is a more proximal antecedent of in-role behaviors than personality and meta-analytic evidence supports the stronger predictive validity of job-specific knowledge than personality as a predictor of in-role performance (Schmidt & Hunter, 2004). On the basis of the greater relation of knowledge with prompt-specific scenario scores and the track record of knowledge as a predictor of in-role performance (Schmidt & Hunter, 2004), we expect that scenario-scores in the specific prompt condition will lead to better predictions of in-role behaviors than scores in the general prompt condition.

Hypothesis 4a: Prompt specificity moderates the positive relationship between intercultural scenario performance and in-role performance, such that the relationship is stronger in the specific compared to the general prompt condition.

Hypothesis 4b: The indirect effect of cultural knowledge on in-role performance via intercultural scenario performance will be stronger in the specific prompt condition than the general prompt condition.

We do not posit differences for predicting extra-role performance via scenario-scores in either the general or specific prompt condition. Extra-role performance, like in-role performance, is facilitated by job-specific knowledge (Motowidlo, Borman, & Schmit, 1997; e.g., to help a colleague with a work task one needs to know how to perform the task) and thus extra-role performance in our context should be predicted by cultural knowledge. As extra-role behaviors

are more voluntary than in-role behaviors, personality traits play a greater role in predicting extra-role than in-role behaviors. This is consistent with meta-analytic evidence that personality is a stronger predictor of extra-role than in-role behaviors (Gonzales-Mulé, Mount, & Oh, 2014). Research has identified perspective taking in particular as an antecedent to extra-role behaviors (Parker & Axtell, 2001) because perspective taking increases feeling for and liking of others (Ku et al., 2015). Thus, we explored as a research question whether intercultural scenario-scores in the general or specific prompt conditions will be more predictive of extra-role performance:

Research Question 1: Will performance in intercultural scenarios in the specific or general prompt conditions be a stronger predictor of extra-role performance?

Method

Participants and Procedure

We collected data from 315 university seniors drawn from an international organizational behavior course at a large business school in Singapore. Participants included 114 males and 201 females (mean age=22.0 years; *SD*=1.85 years) and represented 32 countries. One hundred forty participants had previously lived in one foreign country for more than six months. Another 114 had lived in two, and 61 had lived in three or more foreign countries. Twenty spoke only one language, 135 spoke at least two languages, and 160 spoke three or more languages.

Participants worked in 50 culturally diverse teams (six to eight members per team; average Blau's [1977] index of nationality heterogeneity = .69, *SD* = .10, range: .38 - .83) on a three-month project. The goal of the team project was to develop a video-based dramatization of a challenging intercultural interaction. Teams were assigned a pair of countries on which to base their video-case. Teams then had to analyze differences in cultural values between the two countries, develop scripts for a challenging intercultural interaction based on the cultural value differences, and enact the script to create the video case. Teams were self-managed and were not

assigned formal leaders. The team task presents an intercultural task because (a) teams had to manage team conflicts arising from the cultural diversity of team members (e.g., Asian team members often struggled with how directly Western team mates criticized their ideas, whereas Western team members struggled to make sense of indirect feedback from their Asian team mates), and (b) successful task completion required knowledge about and the ability to enact cultural value differences.

Before starting their projects, participants randomly completed seven intercultural scenarios (for details on scenario development, see [citation removed]), while being randomly assigned to either a general (N = 157) or specific (N = 158) prompts condition. Participants also provided data on cultural knowledge, Big-5 personality, perspective taking, international experience, number of languages spoken, and sex that we used as control variables. Finally, we collected peer ratings of in-role performance, extra-role performance, and intercultural performance at the end of the group project (i.e., after 3 months).

Measures

Intercultural Scenarios. We worked with multimedia-based intercultural scenarios that had been used in various contexts and for different purposes. For example, they were used in intercultural training, open-ended situational judgment tests (SJTs), interviews or coaching (see [citation removed] for an example in the context of SJTs and the coding scheme used).

The intercultural scenario test consists of seven short video vignettes (about 3 minutes each) depicting challenging intercultural interactions at work. The scenarios are similar to video-based SJT items such as those developed by Lievens and Sackett (2006) or Olson-Buchanan et al. (1998) but also differ in three important ways in light of this study's context: (1) scenarios focus specifically on interactions between individuals from two different cultural backgrounds

including North America, South America, Europe, Asia, and the Middle East; (2) each scenario centers around a conflict caused by cultural value differences; and (3) participants respond to scenarios using a constructed-response rather than a selected-response format.

Participants responded to seven intercultural scenarios in randomized order. As noted earlier, we randomly assigned participants to two conditions. In the two conditions, the intercultural scenarios had the same content (i.e., the multimedia scenarios and scoring key were held constant), but the scenarios had different prompts: The prompt in the general prompt condition was "What action(s) would you take to continue this meeting, based on how the video ended?", whereas the prompt in the specific prompt condition was: "What actions would you take to both complete the task and maintain the relationship, based on how the video ended? Consider both parties' perspectives. Be as specific as possible."

Two raters (research assistants blind to the conditions and hypotheses) scored responses to the intercultural scenarios in both conditions using the following single-item rating scale: "To what extent does this response effectively resolve the situation depicted in the vignette? (1 = *not at all effective*; 2 = *slightly effective*; 3 = *somewhat effective*; 4 = *effective*; 5 = *very effective*)." Beforehand we provided frame-of-reference training to both raters according to procedures outlined by Pulakos (1984; see also Woehr & Huffcutt, 1994). We provided raters with definitions and scale anchors; background information on the scenario and its underlying cause of conflict; examples of solution types; and behavioral examples of effective and ineffective responses to each scenario. Next, raters discussed the information. The first author then presented and discussed example responses that represented different levels of performance. Raters then practised making ratings in response to 10 practice responses, and we provided them with feedback. Each rater then independently began rating actual responses. As the same two

raters assessed all responses on a quantitative scale, we assessed inter-rater agreement using the ICC2.1 formula introduced by Shrout and Fleiss (1979). ICC2.1 measures the absolute agreement between raters and reflects the reliability of a single rater. Inter-rater agreement (ICC2.1 = .78) was satisfactory and we averaged both coders' ratings for each video. The internal consistency coefficients of the mean ratings across all seven intercultural scenario scores were similar across conditions (α = .75 for general prompts; α = .76 for specific prompts).

Further, to explore the behavioral effects of prompting, the same two raters also coded (a) whether or not participants suggested solutions from the perspectives of both parties, and (b) whether or not the suggested solutions included an ethnocentric response. We focused on solutions for both parties and on ethnocentric solutions because our theorizing about personality-effects on intercultural scenario responses suggested these as possible behavioral manifestations of high perspective taking and low openness to experience. We rated the presence/absence of these two aspects to focus on the restriction of behavioral manifestation rather than the effectiveness of the associated behavior. Both aspects of the solution are also meaningful within the context of our intercultural scenarios[1]. As the two raters classified responses into distinct categories (presence vs absence of either ethnocentric solutions or suggested solutions for both parties), we assessed inter-rater agreement using Cohen's kappa. Cohen's kappa was .95 for coding of whether responses suggested solutions for both parties and .81 for coding ethnocentric solutions. These agreement indices exceeded the .60 threshold (Landis & Koch, 1977). We resolved all instances of disagreements through discussion between the raters and the first author.

[1] For example, in the scenario between the German and Mexican partners, one may suggest actions for the German partner, the Mexican partner, or for both. In the latter case, raters coded responses as solutions for both parties. Similarly, ethnocentric solutions mean that one party imposes their own values on the other one (e.g., suggesting the German partner continues to behave in the way depicted in the scenario). Although overall responses may be more complex, raters coded responses as ethnocentric if they included an element that suggested one party insists on continuing to act as in the scenario.

Finally, as noted by an anonymous reviewer, our instructions in the specific prompt condition combine instructions that clarify performance criteria with instructions to provide more detailed, specific answers. This combination of instructions may lead to ambiguity about whether effects in the specific prompt condition on participants' responses can be attributed to prompt specificity, instructions to provide more detailed responses, or both. To address this concern, we conducted a follow-up study that examined the effects of response instructions in a 2 (specific vs. general prompts) x 2 (with vs. without instructions to provide detailed responses). Results, based on a sample of 208 participants recruited through Amazon's Mechanical Turk, showed a significant main effect of prompt specificity on scenario scores, perceived clarity over performance criteria, and response length – but no main effect of detailed instructions and no interaction between prompt specificity and detailed instructions. This suggests that effects in the specific prompt condition on participants' responses can be attributed primarily to prompt specificity rather than instructions to provide detailed responses. However, as detailed instructions did significantly increase response length, we also controlled for response length in the main study. We report details of this follow-up study in the supplementary online file (Online Supplement #1).

Dependent variables. Team members assessed each other's in-role performance with three items (e.g., "fulfilled responsibilities of the project"). We selected these three items from Williams and Anderson (1991) on the basis of high factor loadings in prior research and adapted their wording to the context of the group project. As different groups of raters assess different targets for peer ratings, we examined inter-rater agreement using $r_{WG(J)}$ based on a uniform distribution (LeBreton & Senter, 2008). Inter-rater agreement was satisfactory (Mean $r_{WG(J)}$ = .88; *SD* = .14) and we averaged peer-ratings of in-role performance for all analyses (α = .92).

Team members rated extra-role performance using three interpersonal helping items (e.g., "assisted other group members with their work"; Mean $r_{WG(J)}$ = .82; *SD* = .16; α = .89). We selected these three items from Van Dyne and LePine (1998) on the basis of high factor loadings in prior research and adapted their wording to the context of the group project. Both the items for in-role and extra-role performance have been used in a similar context of culturally diverse project teams in prior research (e.g., [citation removed]). For a full list of study items, see the supplementary online file (Online Supplement #2).

Finally, team members rated intercultural performance using the 20-item observer version of the cultural intelligence scale (CQS, Ang & Van Dyne, 2008; e.g., "changes his/her nonverbal behavior when a cross-cultural situation requires it"; Mean $r_{WG(J)}$ = .76; *SD* = .19; α = .86). Observer-ratings of cultural intelligence can be considered an indicator of intercultural performance because they reflect an observer's view and therefore a target person's reputation of how effectively they can function in situations characterized by cultural diversity (see Ang, Van Dyne, & Rockstuhl, 2015; Hogan & Shelton, 1998).

We conducted confirmatory factor analyses on the criterion measures. The hypothesized correlated three-factor model (in-role performance, extra-role performance, and intercultural performance) showed good fit: $\chi^2(32, N = 315) = 75.72$, *ns*, $\chi^2/df = 2.37$, *CFI* = .99, *RMSEA* = .07. All factor loadings were statistically significant (.64 - .97, $p < .01$) and the correlations between the three latent dependent variables were moderate in size ($r_{\text{in-role – extra-role}} = .53$; $r_{\text{in-role – intercultural}} = .32$; $r_{\text{extra-role – intercultural}} = .36$). This model showed significantly better fit than a two-factor model combining in-role and extra-role performance ($\Delta\chi^2(1, N = 315) = 72.22$, $p = .000$) or a one-factor model ($\Delta\chi^2(3, N = 315) = 122.98$, $p = .000$).

As peer ratings of in-role performance, extra-role performance, and intercultural performance were nested within groups, we also tested for non-independence of peer-ratings. One-way ANOVA's for each outcome indicated no significant group effects for in-role performance ($F_{(49, 265)} = 1.11$, $p = .292$), extra-role performance ($F_{(49, 265)} = .89$, $p = .680$), or intercultural performance ($F_{(49, 265)} = .85$, $p = .752$). So, peer-ratings seem to be statistically independent of group membership and can thus be analyzed using statistical procedures that assume independence.

Independent variables. We measured cultural knowledge with a test consisting of 20 multiple-choice questions. The questions tested participants on their knowledge of cultural universals (e.g., Brown, 1991) and of norms associated with major cultural value dimensions (e.g., Hofstede, 1980; House et al., 2004). An example item is "If a culture has a lot of rules and guidelines, this culture is most likely: (a) high uncertainty avoidance, (b) ascription-oriented, (c) collectivistic, or (d) high-context?" Participants completed this test as part of an in-class exercise and we scored the percent of their correct responses ($\alpha = .67$).

We measured openness to experience using Goldberg's (1999) ten-item measure ($\alpha = .82$) and perspective taking with four items adapted from Davis (1980) to reflect an intercultural context. An example item for perspective taking is "I try to understand people from other cultures better by imagining how things look from their perspective" ($\alpha = .90$).

Control variables. As prior research related personality traits to both scenario-based measurements (Levashina et al., 2014; McDaniel, Hartman, Whetzel, & Grubb, 2007; Meriac, Hoffman, Woehr, & Fleisher, 2008) and performance outcomes (Gonzales-Mulé et al., 2014; Shaffer et al., 2006), we controlled for the remaining Big 5 personality traits. We measured these personality traits using 10 items of Goldberg's (1999) IPIP-FFM per trait: extraversion ($\alpha = .78$),

agreeableness (α = .81), conscientiousness (α = .82), and emotional stability (α = .85). We also controlled for gender (0 = male, 1 = female), the total number of languages spoken, and international experience (number of countries lived in). In addition, we counted the number of words per response to measure response length. We averaged the length of responses across the seven videos for all analyses (α = .95).

Additional variables. Beyond the variables to be used for hypotheses testing, we also collected additional information to verify our manipulation and potential alternative explanations. As our prompt manipulation is intended to clarify performance criteria, we conducted a manipulation check using two items (i.e., 'I understand what is considered a good response on this test' and 'The scoring criteria for this test are pretty clear to me'; α = .87). Participants responded to these items after they had responded to all intercultural scenarios using a five-point Likert scale (1 = *strongly disagree,* 5 = *strongly agree*).

Clarifying the performance criteria could also alter participants' task in ways that are unrelated to the degree of specificity. On one hand, as noted by an anonymous reviewer, making the complexity of the task (i.e., having to resolve the tension between potentially competing outcomes) explicit to respondents may inadvertently increase the perceived difficulty of the task. When participants in the specific prompt condition experience the task as more difficult, their task-related self-efficacy and motivation to perform is likely to decrease (Bandura, 1993). To examine this, we measured participants' task-related self-efficacy with the following item after they had responded to each of the scenarios: 'How confident are you that you can solve this incident effectively?' (using a 10-point Likert scale; 1 = *not confident at all*; 10 = *very confident*; α = .88). We also assessed test-taking motivation using three items upon completion of all seven scenarios. To this end, we selected three items from Arvey, Strickland, Drauden, and Martin

(1990) on the basis of high factor loadings in prior research and adapted them to refer to the video-cases (e.g., 'I wanted to do well on these video-case analyses'; α = .81). A five-point Likert scale (1 = *strongly disagree,* 5 = *strongly agree*) was used.

On the other hand, it is also possible that not making the performance criteria explicit to participants frustrates them because they feel that they are not given a fair opportunity to succeed. Hence, participants in the general prompt condition may develop more negative perceptions about the scenarios and thus reduce their effort to respond. To explore this, we assessed respondents' test perceptions upon completion of all seven scenarios. We selected items from Bauer et al. (2001) on the basis of high factor loadings in prior research and adapted them to refer to the video-cases. Two items dealt with perceptions of the face validity (e.g., 'Doing well on this test means a person will do well on international assignments'; α = .72) and two other items with perceptions of opportunity to perform (e.g., 'I was able to show what I can do on these video-case analyses'; α = .68). A five-point Likert scale (1 = *strongly disagree,* 5 = *strongly agree*) was used.

Results

Table 2 presents the means, standard deviations; and Table 3 presents the bivariate correlations of the intercultural scenario scores and all other study variables in the two prompt conditions. Prior to creating interaction terms for the regression analyses, we centered all variables based on the mean in the combined sample.

Preliminary Analyses

Manipulation checks. We began by checking whether our manipulation was successful. Results from our manipulation check items confirmed that participants in the specific prompt condition understood the performance criteria for the intercultural scenarios better (M = 3.94, SD

= .71) than participants in the general prompt condition (M = 2.44, SD = .64; t(313) = 19.56, p = .000, d = 2.22).

As a second manipulation check, we compared respondents' performance on the intercultural scenarios across the two prompt conditions. Given that specific prompts cue behavioral demands and performance expectations, one would expect intercultural scenario performance to be better in the specific than in the general prompt condition. Results supported this expectation: Respondents in the specific prompt condition performed better on the intercultural scenarios (M = 2.51, SD = .47) than participants in the general prompt condition (M = 2.35, SD = .45; t(313) = 3.12, p = .002, d = .35). Consistent with the provided prompts, respondents in the specific prompt condition were also more likely to suggest solutions for both parties (M = .64, SD = .18) and less likely to suggest ethnocentric solutions (M = .15, SD = .16) than participants in the general prompt condition (both perspectives: M = .23, SD = .26; t(313) = 15.61, p = .000, d = 1.83; ethnocentric: M = .22, SD = .16; t(313) = 3.89, p =.001, d = .44).

As a third manipulation check, we tested whether behavioral responses are more restricted in the specific prompt than in the general prompt condition, which is a key assumption underlying situational strength theory (Meyer et al., 2010). Furr and colleagues (Furr, 2009; Furr & Funder, 2004) suggested that this behavioral restriction can be represented by an index of cross-situational behavioral normativeness (i.e., the degree to which a person's responses across scenarios are similar to the average person's responses across scenarios). We computed this index of normativeness as the correlation between a person's profile of both-perspectives and ethnocentric responses and the average profile of both-perspectives and ethnocentric responses within each prompt condition (see Furr, 2009). A greater normativeness index suggests greater behavioral restriction (or less individual variation from a normative profile). Supporting our

behavioral restriction assumption, this behavioral normativeness index was significantly higher in the specific prompt condition (M = .57, SD = .25) than in the general prompt condition (M = .40, SD = .28; t(313) = 5.44, p = .000, d = .64).

Alternative explanations. Next, we conducted analyses to rule out some alternative explanations for possible differences between the two prompt conditions. First, we compared participants' task-specific self-efficacy and test-taking motivation across the two conditions. There were no significant differences across the two conditions in participants' self-efficacy (specific prompt condition: M = 5.81, SD = 1.75; general prompt condition: M = 5.75, SD = 1.21; t(313) = .39, p = .698, d = .04) or test-taking motivation (specific prompt condition: M = 4.02, SD = .60; general prompt condition: M = 4.07, SD = .48; t(313) = .80, p = .422, d = .09). These results suggest that differences in perceived task difficulty are unlikely to account for observed differences between the two prompt conditions.

Second, we compared respondents' test perceptions across the two prompt conditions. There were no significant differences across the two conditions in terms of perceived face validity (specific prompt condition: M = 2.43, SD = .78; general prompt condition: M = 2.45, SD = .50; t(313) = .40, p = .691, d = .04) and opportunity to perform (specific prompt condition: M = 2.85, SD = .66; general prompt condition: M = 2.90, SD = .50; t(313) = .79, p = .428, d = .09). These results suggest that differences in test perceptions are unlikely to account for observed differences between the two prompt conditions.

Third, we examined the measurement equivalence of the intercultural scenario-scores across conditions. Measurement equivalence exists when the numerical values across the two groups are on the same measurement scale (Drasgow, 1984, 1987). Thus, measurement equivalence across conditions would indicate that raters apply the rating scheme consistently and

without bias across conditions (e.g., the difference between a rating of 2 and a rating of 3 has the same meaning across the two prompt conditions), thereby eliminating these possibilities as potential alternative explanations for our results. A series of multigroup confirmatory factor analyses (Byrne et al., 1989) using the lavaan package in R (Rosseel, 2012) supported the measurement equivalence of the intercultural scenario-scores across conditions. In particular, a single-factor model had good fit to the data in both conditions (general prompt: $\chi^2(14, N = 157) = 23.65$, $p = .051$, $\chi^2/df = 1.69$, $CFI = .95$, $RMSEA = .07$; specific prompt: $\chi^2(14, N = 158) = 13.23$, $p = .509$, $\chi^2/df = .95$, $CFI = 1.00$, $RMSEA = .00$). Results further showed (1) configural invariance (i.e., equal patterns of factor-indicator relationships) between both conditions: $\chi^2(28, N = 315) = 36.87$, $p = .122$, $\chi^2/df = 1.32$, $CFI = .98$, $RMSEA = .04$, (2) metric invariance (i.e., equal factor loadings) ($\Delta\chi^2(6, N = 315) = 4.47$, $p = .613$), and (3) scalar invariance (i.e., equal indicator intercepts) ($\Delta\chi^2(12, N = 315) = 16.14$, $p = .185$) of intercultural scenario-scores across the two prompt conditions.

Hypotheses Tests

Prompt-specificity and the relation of personality vs. knowledge with scenario scores. We tested our hypotheses about the differential role of personality and knowledge in intercultural scenario-scores using hierarchical OLS regression analyses. We also conducted two robustness tests using alternative analyses. First, we tested our interaction hypotheses using a series of multigroup path analysis models (Byrne et al., 1989) using the lavaan package in R (Rosseel, 2012). Second, following Bernerth and Aguinis (2016), we tested our hypotheses with and without the inclusion of control variables. The results using these alternative analyses remained substantially unchanged. Here, we report results of the hierarchical regression analyses

without control variables (see Table 4). Results with control variables are available from the first author upon request.

H1 posits that prompt specificity will weaken the positive relationships of openness to experience (H1a) and perspective taking (H1b) with intercultural scenario performance. The overall relationship between openness to experience and scenario performance was positive and significant (Table 4, Model 1: $b = .14$, $t = 2.62$, $p = .009$). However, this main effect was qualified by a significant and negative interaction between prompt specificity and openness to experience (Table 3, Model 2: $b = -.22$, $t = -2.05$, $p = .041$). Supporting H1a, the relationship between openness to experience and intercultural scenario performance was non-significant in the specific prompt condition ($b = .04$, $t = .47$, $p = .635$) but was positive and significant in the general prompt condition ($b = .26$, $t = 3.34$, $p = .001$). The differences in zero-order correlations (specific prompt condition: $r = .07$ vs general prompt condition: $r = .27$) and standardized simple slopes (specific prompt condition: $\beta = .03$ vs general prompt condition: $\beta = .24$) across the two conditions suggest that the effect of prompt specificity is also practically significant.

The overall relationship between perspective taking and scenario performance was positive and significant (Table 4, Model 1: $b = .10$, $t = 3.94$, $p = .000$). Similar to openness to experience, the interaction between prompt specificity and perspective taking was negative and significant (Table 4, Model 2: $b = -.11$, $t = -2.30$, $p = .022$). As hypothesized, the relationship between perspective taking and intercultural scenario performance was non-significant in the specific prompt condition ($b = .02$, $t = .59$, $p = .554$) but was positive and significant in the general prompt condition ($b = .14$, $t = 4.60$, $p = .000$). Thus, H1b is supported. The differences in zero-order correlations (specific prompt condition: $r = .04$ vs general prompt condition: $r = .28$) and standardized simple slopes (specific prompt condition: $\beta = .05$ vs general prompt condition:

β = .30) across the two conditions suggest that the effect of prompt specificity is also practically significant.

H2 states that prompt specificity will moderate the positive relationship between cultural knowledge and intercultural scenario performance, such that the relationship is stronger in the specific compared to the general prompt condition. The overall relationship between cultural knowledge and scenario performance was positive and significant (Table 4, Model 1: b = .01, t = 5.27, p = .000). More importantly, the interaction between prompt specificity and cultural knowledge was positive and significant (Table 4, Model 2: b = .01, t = 2.68, p = .008). In particular, cultural knowledge was more positively associated with intercultural scenario performance in the specific (b = .02, t = 4.45, p = .000) than in the general (b = .01, t = 2.56, p = .011) prompt condition. These results support H2. In addition, the differences in zero-order correlations (specific prompt condition: r = .39 vs general prompt condition: r = .20) and standardized simple slopes (specific prompt condition: β = .45 vs general prompt condition: β = .17) across the two conditions suggest that the effect of prompt specificity is also practically significant.

Prompt-specificity and validity of scenario scores. Table 5 shows the results of hierarchical OLS regression analyses that test our hypotheses about the differences in predictive validity of intercultural scenario performance in the specific versus general prompt conditions. Table 5 shows the results of these analyses without inclusion of all control variables (we obtained substantively unchanged results when conducting analyses with control variables).

Our next hypotheses state that prompt specificity will weaken the positive relationship between intercultural scenario performance and intercultural performance (H3a) and thus the indirect effects of openness to experience and perspective taking on intercultural performance

(H3b). The overall relationship between intercultural scenario performance and intercultural performance was positive and significant (Table 5, Model 1: b = .30, t = 5.69, p = .000). However, this main effect was qualified by a small negative interaction between prompt specificity and scenario performance (Table 4, Model 2: b = -.17, t = -1.66, p = .098). This result was not statistically significant, but was close to the a priori alpha level of .05. Supporting H3a, the relationship between scenario performance and intercultural performance was weaker in the specific (b = .23, t = 3.35, p = .000) than in the general (b = .40, t = 5.29, p = .000) prompt condition. The hypothesized interaction between prompt specificity and scenario performance explained an extra 1% of variance in intercultural performance above all main effects. In addition, the differences in operational validity (specific prompt condition: r_C = .30 vs general prompt condition: r_C = .47) and standardized simple slopes (specific prompt condition: β = .26 vs general prompt condition: β = .44) across the two conditions suggest that the effect of prompt specificity is also practically significant.

We further tested the proposed moderated mediation model using a bootstrapped model of conditional indirect effects (Preacher, Rucker, & Hayes, 2007). We used the SPSS Process macro for these analyses (Hayes, 2013). Results based on 5,000 bootstrap samples suggest that the indirect effects of openness to experience (.013; 95% CI [-.031, .056]) and perspective taking (.006; 95% CI [-.020, .033]) are statistically non-significant in the specific prompt condition. By contrast, the indirect effects of openness to experience (.075; 95% CI [.032, .122]) and perspective taking (.041; 95% CI [.021, .065]) are statistically significant in the general prompt condition. Following Hayes (2015), we calculated the index of moderated mediation to test the statistical significance of the moderated mediation effects. Results indicate that this index is

statistically significant for both openness to experience (-.062; 95% CI [-.126, -.007]) and perspective taking (-.035; 95% CI [-.073, -.005]). These analyses therefore support H3b.

Next, we posited that prompt specificity will strengthen the positive relationship between intercultural scenario performance and in-role performance (H4a) and thus the indirect effects of cultural knowledge on in-role performance (H4b). The overall relationship between intercultural scenario performance and in-role performance was positive and significant (Table 5, Model 3: b = .21, t = 4.82, p = .000). As predicted, this main effect was qualified by a positive and significant interaction between prompt specificity and scenario performance (Table 5, Model 4: b = .18, t = 2.14, p = .033). Supporting H4a, the relationship between scenario performance and in-role performance was stronger in the specific (b = .31, t = 5.37, p = .000) than in the general (b = .14, t = 2.14, p = .034) prompt condition. The hypothesized interaction effect between prompt specificity and scenario performance explained an additional 3% of variance in in-role performance over and above all main effects. The differences in operational validity (specific prompt condition: r_C = .43 vs general prompt condition: r_C = .28) and standardized simple slopes (specific prompt condition: β = .40 vs general prompt condition: β = .17) across the two conditions suggest that the effect of prompt specificity is also practically significant.

Testing the proposed moderated mediation model, results based on 5,000 bootstrap samples suggest that the indirect effect of cultural knowledge on in-role performance is statistically significant in both the specific (.004; 95% CI [.002, .007]) and the general (.002; 95% CI [.000, .003]) prompt conditions. In addition, Hayes' (2015) index of moderated mediation is statistically significant (.002; 95% CI [.000, .005]). Thus, H4b is supported.

Finally, we examined whether prompt specificity moderates the positive relationship between intercultural scenario performance and extra-role performance. The overall relationship

between intercultural scenario performance and extra-role performance was positive and significant (Table 5, Model 5: $b = .18$, $t = 3.22$, $p = .001$). However, this main effect was not qualified by a statistically significant interaction effect between prompt specificity and scenario performance (Table 5, Model 6: $b = .07$, $t = .63$, $p = .528$). So, prompt specificity did not moderate the relation between scenario performance and extra-role performance.

Additional Analyses of Underlying Mechanisms

A major tenet of our theorizing is that specific prompts attenuate the role of personality traits in scenario scores by dampening the behavioral expression of personality trait differences. To illuminate this mechanism, we explored two mediators of the effects of personality on scenario-scores: (a) whether respondents suggested solutions for both parties and (b) whether respondents suggested ethnocentric solutions. On the basis of our arguments that both openness to experience and perspective taking reduce ethnocentric responding, we expected both personality traits to be negatively associated with suggesting ethnocentric solutions in the general prompt condition. In addition, as perspective taking refers to the tendency to adopt spontaneously others' psychological point of view (Davis, 1983), we expected perspective taking to be positively associated with suggesting solutions for both parties in the general prompt condition. As specific prompts restrict personality differences in expressed behavior, we expected personality to be less strongly related to ethnocentric and both-perspectives solutions in the specific prompt condition.

We examined these expectations by conducting moderated mediation analyses with intercultural scenario scores as the dependent variable and ethnocentric solutions and solutions for both parties as mediators respectively. Results show that the indirect effect of openness to experience on scenario performance via ethnocentric solutions is statistically significant in the

general (.130; 95% CI [.057, .206]) but not the specific (-.000; 95% CI [-.073, .076]) prompt condition. Similarly, the indirect effect of perspective taking on scenario performance via suggesting solutions for both parties is statistically significant in the general (.015; 95% CI [.005, .029]) but not the specific (-.001; 95% CI [-.010, .007]) prompt condition. Consistent with expectations, Hayes' (2015) index of moderated mediation is statistically significant for both indirect effects (openness to experience via ethnocentric solutions: -.131; 95% CI [-.236, -.027]; perspective taking via solutions for both parties: -.016; 95% CI [-.034, -.003]). None of the other indices of moderated mediation were significant. Results are available from the first author.

Discussion

Main Conclusions

In light of the omnipresence of scenario-based assessment in research and practice, it is surprising how little consensus exists around (1) choices in prompt-specificity and (2) their effects on key outcomes. In fact, opposing views exist as to whether prompts should be designed more specifically or more generally: One perspective states that more specific prompts increase validity because they enhance the likelihood that all respondents are provided with the opportunity to display their abilities and relevant behaviors. Conversely, according to the other perspective, more general prompts increase validity because they allow for the observation of more spontaneous trait-related behavior (Blackman & Funder, 2002).

This study reconciles both views by developing and testing hypotheses which build on the interplay between situation construal and situational strength theory to suggest that the differential activation of individual differences (personality vs. knowledge) serves as the key explanatory mechanism through which prompt-specificity exerts its effects on key outcomes. Our hypotheses received general support. Lesser prompt-specificity increased the role of two

relevant personality traits (openness to experience and perspective taking) for solving intercultural scenario-scores, while greater prompt-specificity increased the role of knowledge in these scores.

Critically, these differential associations with individual differences constructs in turn explained the differences in predictive validity of scenario-scores in the specific and general prompt conditions. That is, greater prompt-specificity not only increased the predictive validity of scenario-scores for in-role performance, but also increased the indirect effect of knowledge on in-role performance via scenario-scores. By contrast, lesser prompt-specificity not only increased the predictive validity of scenario-scores for intercultural performance, but also increased the indirect effect of openness to experience and perspective taking on intercultural performance via scenario-scores. The size of these interaction effects ranged from 1% in incremental variance explained for intercultural performance to 3% for in-role performance, suggesting practically meaningful differences in the predictive validity of scenario scores across prompt conditions. More importantly, the differences in operational validity (intercultural performance – specific prompt condition: r_C = .30 vs general prompt condition: r_C = .47; in-role performance – specific prompt condition: r_C = .43 vs general prompt condition: r_C = .28) highlight the practical significance of prompt-specificity for predictions of intercultural and in-role performance. Thus, taken together, our results show that simple variations in prompt-specificity produce variations in the constructs being correlated with scenario scores and predictive validity of these scores. Below, we discuss the implications for future theoretical development, research and practice.

Theoretical Implications

One theoretical contribution of this study lies in deepening our understanding of prompt-specificity by illuminating its conceptual meaning. We isolated the effects of prompt-specificity

as method factor while keeping other things constant (multimedia stimulus, scoring, etc.), Accordingly, it became clear that conceptually different constructs are activated depending on how specifically or generally prompts are specified. Importantly, the implications of these findings go beyond one specific assessment procedure because scenarios may be included into a variety of contexts (research, selection, training) and approaches (e.g., structured interviews, assessment centers, work samples, SJTs, simulations). Therefore, our findings increase the theoretical connectivity among these different literatures that all deal with the use of scenarios.

As another contribution, the differential effects of general vs. specific prompts highlight the novelty of prompt-specificity as a method factor vis-à-vis the broader concept of "structure" in selection procedures. In the context of interviews, Campion, Palmer, and Campion (1997) defined structure as "any enhancement of the interview that is *intended to increase psychometric properties* by increasing standardization" (p. 656, emphasis added). As Levashina et al. (2014) noted, twelve meta-analyses on the effects of structure consistently found that structure increases the predictive validity of interviews. They also note that research has shown that structured interviews can be designed to measure different constructs and predict different criteria. The fact that structure improves the predictive validity of interviews regardless of constructs assessed and criteria considered suggests that structure improves prediction primarily by increasing job-relatedness and reliability (see also Schmidt & Zimmerman, 2004). By contrast, prompt specificity did not significantly affect measurement reliability, but did influence the constructs being correlated with scenario-scores and subsequent predictive validity for different criteria.

Third, our findings add insights to our knowledge on the effects of "instructions". Lievens and Sackett's (2017) modular framework of selection procedures includes instructions as one of the key factors. They made a distinction between general and specific instructions and

link this distinction to the effects of making selection procedures more transparent. We agree that transparency is an umbrella concept that refers to various interventions that divulge the constructs measured to in a selection procedure to candidates. Similar to earlier transparency instructions (e.g., Kleinmann, 1993; Ingold, Kleinmann, König, & Melchers, 2016; Smith-Jentsch, 2007), prompt specificity might also resort under this broad concept, although there are some differences in operationalization. In prior transparency studies, the focal constructs were typically revealed to participants *prior* to the assessment procedure. Conversely, prompts are tied to a specific question and/or answer and are therefore given *during* the assessment. In addition, verbal prompts are often given in the form of an additional question, which was not the case in prior transparency manipulations. A final difference is that specific prompts do not mention the concrete behaviors to be shown, whereas typical transparency manipulations mention both the focal constructs and behavioral examples (e.g., Ingold et al., 2016). Regardless of these differences in operationalization, this study shows that under specific circumstances prompt specific variations and transparency might exert similar negative effects on validity. For example, we found that the use of more specific prompts made the task easier[2] and increased the validity of scenario-scores for predicting in-role performance but it also lowered the validity for predicting intercultural performance. Apparently, the predictiveness of the scenarios for predicting intercultural performance (which put emphasis on cooperation and communication

[2] We thank an anonymous reviewer for pointing out this interpretation of mean differences in scenario performance in terms of task difficulty. While specific prompts appear to make the task easier, we also note that manipulating prompt specificity is probably not the same as manipulating task difficulty. For example, if the two were the same, one would expect that in the specific prompt condition, scenario scores would exhibit a lower correlation with cognitive ability scores. Additional analyses, available from the first author, suggest that this is not the case, and that cognitive ability scores, like cultural knowledge are more strongly associated with scenario scores in the specific, rather than the general prompt condition. Furthermore, we replicated our analyses with intercultural scenario scores that are centered within each prompt condition, thus removing mean differences in scenario performance. Results remained unchanged, suggesting that our observed interaction effects are not driven primarily by mean differences in task difficulty across conditions.

among culturally diverse people) was impaired when specific prompts made the performance requirement to "maintain the relationship" more obvious to participants in the test situation, irrelevant of whether they would show these behaviors on the job. This matches recent findings of transparency lowering the predictive validity of assessment centers (Ingold et al., 2016).

As a final contribution, our findings shift attention from the “contest” between prompt specificity and prompt generality to recognizing that both are predictive, albeit for different outcomes. A key take-away from our study is that both prompt specificity variations lead to differential correlations with external constructs – i.e., the leading to a greater role of knowledge vs. personality. In turn, these differential correlations of test scores with relevant constructs affect predictive validity. Consistent with Ajzen’s (2005) matching principle and the predictor-criterion matching logic (Lievens et al., 2005), scenario-based ratings on the basis of specific prompts are more predictive of outcomes that have a strong record of being predicted by cognitive constructs (e.g., in-role performance), whereas scenario-based ratings on the basis of general prompts are more predictive of outcomes associated with personality constructs (e.g., intercultural performance).

Directions For Future Research

We highlight several avenues for future research. First, a major finding from our study is that prompt-specificity increases the role of knowledge in scenario-based assessments. One implication of this finding refers to the potential for adverse impact of selection procedures, which are typically driven by cognitive load (Ployhart & Holtz, 2008; Dalhke & Sackett, 2017). Our findings therefore suggest that greater prompt-specificity might increase the adverse impact of selection procedures and future research should test this possibility directly.

Second, an exciting avenue for future research consists of developing a comprehensive taxonomy of prompt manipulations. For example, our findings bear some resemblance to research on SJT instructions that distinguishes between knowledge vs. behavioral tendency instructions (McDaniel et al., 2007). However, our research increased situational strength by specifying behavioral demands, whereas knowledge instructions in SJTs increase situational strength by invoking normative constraints (i.e., "what one should do"). According to Meyer et al. (2010), specificity of behavioral demands and normative constraints are different facets of situational strength. Future research could draw upon the framework by Meyer et al. to explore consistency of prompts and consequences of prompted behaviors as additional prompt manipulations. Similarly, future studies might conceptualize prompt-specificity less as a continuum and examine qualitatively different prompts (e.g., verbal vs. non-verbal).

Finally, future research could examine the generalizability of our findings to other non-scenario-based measures. Tourangeau et al.'s (2000) survey response process model suggests that our findings also extend to other measures. This model describes four basic cognitive processes in responding to survey items: comprehension, retrieval, judgment, and response. Comprehension involves cognitive processes involved with reading, interpreting, and understanding the question's purpose. This initial comprehension is followed by retrieval of question-relevant information from long-term memory, a judgment about the likelihood of question-relevant events based on this information, and finally the response. Prompt-specificity may affect the comprehension process in that higher prompt-specificity in survey items affects how constructs assessed in surveys relate to other constructs in their nomological network.

Limitations

A first limitation relates to the nature of our sample. We relied on students which may evoke concerns about the external validity of our findings. However, although being students, participants worked in teams similar to teams one may find in real-world contexts. Moreover, teams were highly diverse and had to work interdependently within their team on a fairly complex project and under time pressure. Nevertheless, future research that replicates our findings in applicant and managerial samples would strengthen their external validity.

A second limitation stems from our focus on intercultural scenarios. We noted at the outset that intercultural scenarios may play an increasing role in selection and training procedures due to the growing diversity in the workplace. We also recognize that resolving intercultural dilemmas may be particularly challenging because of lack of clear cultural norms about what constitutes an appropriate response. Although this makes intercultural scenarios an interesting context to study prompt specificity manipulations, we encourage future research to examine the generalizability of our findings to other scenario-based assessments. Such research should also examine whether our findings generalize to other knowledge or personality trait constructs that may be relevant to different types of scenarios.

Finally, we note that we had used a relatively narrow operationalization of extra-role performance that focused on interpersonal helping behaviors. Although this may limit the generalizability of our findings to broader citizenship behaviors, the meta-analyses by LePine, Erez, and Johnson (2002) reveals that different dimensions of citizenship behaviors form a latent construct and can thus be used interchangeably as indicators of broader citizenship behaviors. Nevertheless, future research should examine the generalizability of our findings to broader measures of extra-role performance.

Practical Implications

Generally, our findings identify prompt-specificity as an important building block (see modular framework of Lievens and Sackett, 2017) to consider when embedding prompts in scenario-based assessments. First, this study offers advice to practitioners when they have difficulty choosing between the use of specific or general prompts in scenario-based assessment in selection and training contexts. If one's goal is to test whether an individual, for example, can take both parties' perspectives into account, then this study suggests that one should use specific prompts so that it becomes clear whether (s)he can do so. If one wants to know whether (s)he will spontaneously tend to take multiple perspectives into account, then this study suggests one should refrain from using specific prompts.

Second, our results offer timely information on how prompts elicit not only different candidate responses but also impact critical selection outcomes. Given that the scenario-based assessment predicted different outcomes depending on the level of prompt-specificity, practitioners may consider combining scenarios with general and specific response prompts in selection procedures to maximize their criterion coverage and predictive breadth.

Third, the predictive validity of the intercultural scenario-based assessment for performance in a multicultural context is encouraging in light of the growing internationalization of the workplace. Intercultural scenario-based assessment may offer a useful complement to existing measures of intercultural skills (e.g., Ang et al., 2007) that are often based on self-reports. Such intercultural scenario-based assessment can also be fruitfully used in the context of cross-cultural training programs (Bhawuk & Brislin, 2000).

Conclusion

This study draws on the interplay of situation construal and situational strength theory to examine the effects of prompt-specificity on the validity of scenario-based assessments. At a

theoretical level, this study highlights that conceptually different constructs (personality vs. knowledge) are activated by different prompts and affect the predictive potential of the scores obtained. At a practical level, this study generates theory-based and evidence-based recommendations on the level of prompt-specificity in scenario-based assessments.

References

Aguinis, H., & Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods, 17*, 351-371.

Ajzen, I. (2005). *Attitudes, personality, and behavior* (2nd ed.). Chicago, Il: Dorsey Press.

Albrecht, A. G., Dilchert, S., Deller, J., & Paulus, F. M. (2014). Openness in cross-cultural work settings: A multicountry study of expatriates. *Journal of Personality Assessment, 96,* 64–75.

Allport, G. W. (1961). *Pattern and growth in personality*. New York: Holt, Rinehart & Wilson.

Ang, S., & Van Dyne, L. (2008) (Eds.). *Handbook of cultural intelligence*. New York: M.E. Sharpe.

Ang, S., Van Dyne, L., & Koh, C. (2006). Personality correlates of the four-factor model of cultural intelligence. *Group and Organization Management, 31*, 100-123.

Ang, S., Van Dyne, L., Koh, C. K. S., Ng, K. Y., Templer, K. J., Tay, C., & Chandrasekar, A. N. (2007). Cultural intelligence: Its measurement and effects on cultural judgment and decision making, cultural adaptation, and task performance. *Management and Organization Review, 3*, 335-371.

Ang, S., Van Dyne, L., & Rockstuhl, T. (2015). Cultural intelligence: Origins, conceptualization, evolution and methodological diversity. In Gelfand, M., Chiu, C. Y., & Hong, Y. Y. (Eds.), *Handbook of Advances in Culture and Psychology*, Vol 5 (pp. 273-323). New York: Oxford University Press.

Arvey, R. D., Strickland, W., Drauden, G., & Martin, C. (1990). Motivational components of test taking. *Personnel Psychology, 43*, 695-711.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist, 28*, 117-148.

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1-26.

Bartram, D. (2005). The Great Eight competencies: A criterion-centric approach to validation. *Journal of Applied Psychology, 90*, 1185-1203.

Bauer, T. N., Truxillo, D. M., Sanchez, R. J., Craig, J. M., Ferrara, P., & Campion, M. A. (2001). Applicant reactions to selection: Development of the selection procedural justice scale (SPJS). *Personnel Psychology, 54*, 387-419.

Bernerth, J. B., & Aguinis, H. (2016). A critical review and best-practice recommendations for control variable usage. *Personnel Psychology, 69*, 229-283.

Bhawuk, D. P. (2001). Evolution of culture assimilators: Toward theory-based assimilators. *International Journal of Intercultural Relations, 25*, 141-163.

Bhawuk, D. P. (2017). Culture assimilator. In Y. Y. Kim (Ed.) *The International Encyclopedia of Intercultural Communication* (pp. 1-9). Wiley Online Publication.

Bhawuk, D. P., & Brislin, R. (2000). Cross-cultural training: A review. *Applied Psychology: An International Review, 49*, 162–191.

Blackman, M. C., & Funder, D. C. (2002). Effective interview practices for accurately assessing counterproductive traits. *International Journal of Selection and Assessment, 10*, 109-116.

Blau, P. M. (1977). *Inequality and heterogeneity: A primitive theory of social str*ucture (Vol. 7). New York: Free Press.

Block, J., & Block, J. H. (1981). Studying situational dimensions: A grand perspective and some limited empiricism. In D. Magnusson (Ed.), *Toward a psychology of situations: An interactional perspective* (pp. 85–103). Hillsdale, NJ: Erlbaum.

Brown, D. (1991). *Human Universals*. New York: McGraw-Hill.

Brislin, R. W. (2009). Theory, critical incidents, and the preparation of people for intercultural experiences. In R. S. Wyer, C.-Y. Chiu & Y.-Y. Hong (Eds.), *Understanding culture: Theory, research, and application* (pp. 379-392). New York, NY: Psychology Press.

Bryne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of practical measurement invariance. *Psychological Bulletin, 105*, 456-466.

Caligiuri, P. M. (2006). Performance measurement in a cross-national context. In W. Bennett, C. E. Lance, & D. J. Woehr (Eds.), *Performance measurement: Current perspectives and future challenges* (pp. 227 – 244). Mahwah, NJ: Erlbaum.

Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. *Personnel Psychology, 50*, 655-702.

Chen, X. P., Liu, D., & Portnoy, R. (2012). A multilevel investigation of motivational cultural intelligence, organizational diversity climate, and cultural sales: Evidence from US real estate firms. *Journal of Applied Psychology, 97*, 93-106.

Cialdini, R. B., Brown, S. L., Lewis, B. P., Luce, C., & Neuberg, S. L. (1997). Reinterpreting the empathy–altruism relationship: When one into one equals oneness. *Journal of Personality and social Psychology, 73*, 481-494.

Cohrs, J. C., Kämpfe-Hargrave, N., & Riemann, R. (2012). Individual differences in ideological attitudes and prejudice: Evidence from peer- report data. *Journal of Personality and Social Psychology, 103,* 343– 361.

Cooper, W. H., & Withey, M. J. (2009). The strong situation hypothesis. *Personality and Social Psychology Review, 13*, 62-72.

Cushner, K., & Landis, D. (1996). The intercultural sensitizer. In D. Landis & R. S. Bhagat (Eds.), *Handbook of intercultural training* (2nd ed., pp. 185-202). Thousand Oaks, CA: Sage.

Dahlke, J. A., & Sackett, P. R. (2017). The relationship between cognitive-ability saturation and subgroup mean differences across predictors of job performance. *Journal of Applied Psychology, 102*, 1403-1420.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. JSAS *Catalog of Selected Documents in Psychology, 10*, 85.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*, 113-126.

Davis, M. H., Conklin, L., Smith, A., & Luce, C. (1996). Effect of perspective taking on the cognitive representation of persons: a merging of self and other. *Journal of Personality and Social Psychology, 70*, 713-726.

Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. J. Deary, F. De Fruyt, and F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 7–28). Tilburg, The Netherlands: Tilburg University Press.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*, 371-395.

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

Furr, R. M. (2009). Profile analysis in person–situation integration. *Journal of Research in Personality*, *43*, 196-207.

Furr, R. M., & Funder, D. C. (2004). Situational similarity and behavioral consistency: Subjective, objective, variable-centered, and person-centered approaches. *Journal of Research in Personality*, *38*, 421-447.

Gibson, C. B., & Zellmer-Bruhn, M. E. (2001). Metaphors and meaning: An intercultural analysis of the concept of teamwork. *Administrative Science Quarterly, 46*, 274–303.

Gonzales-Mulé, E., Mount, M. K., & Oh, I.-S. (2014). A meta-analysis of the relationship between general mental ability and nontask performance. *Journal of Applied Psychology, 99*, 1222-1243.

Hartley, A. A., Kieley, J. M., & Slabach, E. H. (1990). Age differences and similarities in the effects of cues and prompts. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 523-537.

Hauenstein, N. M. A., Findlay, R. A., & McDonald, D. P. (2010). Using situational judgment tests to assess training effectiveness: Lessons learned evaluating military equal opportunity advisor trainees. *Military Psychology, 22*, 262-281.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: The Guilford Press.

Hayes, A. F. (2015). An index and test of linear moderated mediation. *Multivariate Behavioral Research, 50*, 1-22.

Hofstede, G. (1980). *Culture's Consequences: International Differences in Work-Related Values*. Beverly Hills, CA: Sage.

Hogan, J., & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. *Journal of Applied Psychology, 88*, 100-112.

Hogan, R., & Shelton, D. (1998). A socioanalytic perspective on job performance. *Human Performance, 11*, 129-144.

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Palo Alto, CA: Sage.

Ingold, P. V., Kleinmann, M., König, C. J., & Melchers, K. G. (2016). Transparency of assessment centers: Lower criterion-related validity but greater opportunity to perform? *Personnel Psychology, 69*, 467-497.

Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin, 129,* 339 –375.

Klehe, U.-C., König, C. J., Richter, G. M., Kleinmann, M., & Melchers, K. G. (2008). Transparency in structured interviews: Consequences for construct and criterion-related validity. *Human Performance, 21*, 107-137.

Kleinmann, M. (1993). Are rating dimensions in assessment centers transparent for participants? Consequences for criterion and construct validity. *Journal of Applied Psychology, 78*, 988-993.

Ku, G., Wang, C. S., & Galinsky, A. D. (2015). The promise and perversity of perspective-taking in organizations. *Research in Organizational Behavior, 35*, 79-102.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174.

Latham, G. P., Saari, L. M., Pursell, E. D., & Campion, M. A. (1980) The situational interview. *Journal of Applied Psychology, 65*, 422-427.

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods, 11*, 815-852.

Leiba-O'Sullivan, S. (1999). The distinction between stable and dynamic cross-cultural competencies: Implications for expatriate trainability. *Journal of International Business Studies, 30*, 709-725.

LePine, J. A., Erez, A., & Johnson, D. E. (2002). The nature and dimensionality of organizational citizenship behavior: A critical review and meta-analysis. *Journal of Applied Psychology, 87*, 52-65.

Leung, K., Ang, S., & Tan, M. L. (2014). Intercultural competence. *Annual Review of Organizational Psychology and Organizational Behavior, 1*, 489–519.

Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology, 67*, 241-293.

Lievens, F., Buyse, T., & Sackett, P. R. (2005). The operational validity of a video-based situational judgment test for medical college admissions: illustrating the importance of matching predictor and criterion construct domains. *Journal of Applied Psychology, 90*, 442-452.

Lievens, F., & De Soete, B. (2015). Situational judgment tests. In James D. Wright (editor-in-chief), International Encyclopedia of the Social & Behavioral Sciences (Vol. 22, pp. 13-19). Oxford, UK: Elsevier.

Lievens, F., Harris, M. M., Van Keer, E., & Bisqueret, C. (2003). Predicting cross-cultural training performance: the validity of personality, cognitive ability, and dimensions measured by an assessment center and a behavior description interview. *Journal of Applied Psychology, 88*, 476-489.

Lievens, F., & Sackett, P. R. (2006). Video-based versus written situational judgment tests: A comparison in terms of predictive validity. *Journal of Applied Psychology, 91*, 1181-1188.

Lievens, F., & Sackett, P. R. (2017). The effects of predictor method factors on selection outcomes: A modular approach to personnel selection procedures. *Journal of Applied Psychology, 102*, 43-66.

Lievens, F., Schollaert, E., & Keen, G. (2015). The interplay of elicitation and evaluation of trait-expressive behavior: Evidnce in assessment center exercises. *Journal of Applied Psychology, 100*, 1169-1188.

McCrae, R. R. (1987). Creativity, divergent thinking, and openness to experience. *Journal of Personality and Social Psychology, 52,* 1258– 1265.

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*, 63-91.

Meriac, J. P., Hoffman, B. J., Woehr, D. J., & Fleisher, M. S. (2008). Further evidence for the validity of assessment center dimensions: A meta-analysis of the incremental criterion-related validity of dimension ratings. *Journal of Applied Psychology*, *93*, 1042-1052.

Messick, S. J. (1998). Alternative modes of assessment, uniform standards of validity. In M. D. Hakel (Ed.), *Beyond multiple choice* (pp. 59-74). Mahwah, NJ: Lawrence Erlbaum.

Meyer, R. D., Dalal, R. S., & Hermida, R. (2010). A review and synthesis of situational strength in the organizational sciences. *Journal of Management, 36,* 121-140.

Mischel, W. (1968). *Personality and assessment*. New York: John Wiley.

Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychological Review, 80,* 252-283.

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review, 102*, 246-268.

Molinsky, A. L. (2013). The psychological processes of cultural retooling. *Academy of Management Journal, 56*, 683–710.

Motowidlo, S. J., Borman, W. C., & Schmit, M. J. (1997). A theory off individual differences in task and contextual performance. *Human Performance, 10*, 71-83.

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990) An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology, 75*, 640-647.

Olson-Buchanan, J. B., Drasgow, F., Moberg, P. J., Mead, A. D., Keenan, P. A., & Donovan, M. A. (1998). Interactive video assessment of conflict resolution skills. *Personnel Psychology, 51*, 1-24.

Ones, D. S., & Viswesvaran, C. (1997). Personality determinants in the prediction of aspects of expatriate job success. In Z. Aycan (Ed.), *New approaches to employee management, Vol. 4. Expatriate management: Theory and research* (pp. 63-92). Bingley, UK: Emerald.

Osland, J. S., & Bird, A. (2000). Beyond sophisticated stereotyping: Cultural sensemaking in

context. *The Academy of Management Executive, 14*, 65–77.

Ostroff, C. (1991). Training effectiveness measures and scoring schemes - a comparison. *Personnel Psychology, 44*, 353-374.

Parker, S. K., Atkins, P. W. B., & Axtell, C. M. (2008). Building better work places through individual perspective taking: A fresh look at a fundamental human process. In G. Hodgkinson & K. Ford (Eds.), *International review of industrial and organizational psychology* (pp. 149–196). Chichester, England: Wiley.

Parker, S. K., & Axtell, C. M. (2001). Seeing another viewpoint: Antecedents and outcomes of employee perspective taking. *Academy of Management Journal, 44*, 1085-1100.

Pettigrew, T. F., & Tropp, L. R. (2008). How does intergroup contact reduce prejudice? Meta-analytic tests of three mediators. *European Journal of Social Psychology, 38*, 922-934.

Ployhart, R. E. (2006). The predictor response model. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and practice* (pp. 83-105). Mahwah, NJ: Erlbaum.

Ployhart, R. E., & Holtz, B. C. (2008). The diversity–validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology, 61*, 153-172.

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research, 42*, 185-227.

Pulakos, E. D. (1984). A comparison of rater training programs: Error training and accuracy training. *Journal of Applied Psychology, 69*, 581-588.

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*, 363-381.

Reis, H. T. (2008). Reinvigorating the concept of situation in social psychology. *Personality and Social Psychology Review, 12*, 311-329.

Rockstuhl, T., & Van Dyne, L. (2018). A bi-factor theory of the four-factor model of cultural intelligence: Meta-analysis and theoretical extensions. *Organizational Behavior and Human Decision Processes, 148*, 124-144

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*(2), 1-36.

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: occupational attainment and job performance. *Journal of Personality and Social Psychology, 86*, 162-173.

Schmidt, F. L., & Zimmerman, R. D. (2004). A counterintuitive hypothesis about employment interview validity and some supporting evidence. *Journal of Applied Psychology, 89*, 553-561.

Schmitt, N., & Ostroff, C. (1986). Operationalizing the "behavioral consistency" approach: Selection test development based on a content-oriented strategy. *Personnel Psychology, 39*, 91-108.

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*, 255-271.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, *54*, 93-105.

Shaffer, M. A., Harrison, D. A., Gregersen, H., Black, J. S., & Ferzandi, L. A. (2006). You can take it with you: Individual differences and expatriate effectiveness. *Journal of Applied Psychology, 91*, 109-125.

Sherman, R. A., Nave, C. S., & Funder, D. C. (2013). Situational construal is related to personality and gender. *Journal of Research in Personality, 47*, 1-14.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428.

Sibley, C. G., & Duckitt, J. (2008). Personality and prejudice: A meta- analysis and theoretical review. *Personality and Social Psychology Review, 12,* 248–279.

Smith-Jentsch, K. A. (2007). The impact of making targeted dimensions transparent on relations with typical performance predictors. *Human Performance, 20*, 187–203.

Stone-Romero, E., Stone, D. L., & Salas, E. (2003). The influence of culture on role conceptions and role behavior in organizations. *Applied Psychology: An International Review, 52*, 328–362.

Thornton, G. C., & Cleveland, J. N. (1990). Developing managerial talent through simulation. *American Psychologist, 45*, 190-199.

Todd, A. R., & Galinsky, A. D. (2014). Perspective-taking as a strategy for improving intergroup relations: Evidence, mechanisms, and qualifications. *Social and Personality Psychology Compass, 8*, 374-387.

Tourangeau, R., Rips, I. J., & Rasinski, K. (2000). *The psychology of survey response.* Cambridge: Cambridge University Press.

Triandis, H. C. (2006). Cultural intelligence in organizations. *Group & Organization Management, 31*, 20-26.

Van Dyne, L., & LePine, J. A. (1998). Helping and voice extra-role behaviors: Evidence of construct and predictive validity. *Academy of Management Journal, 41*, 108-119.

Wernimont, P. F., & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology, 52*, 372-376.

Williams, L. J., & Anderson, S. E. (1991). Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behavior. *Journal of Management, 17,* 601-618.

Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology, 6*7, 189-205.

Yeung, A., & Ready, D. (1995). Developing leadership capabilities of global corporations: A comparative study in eight nations Human Resource Management, 34, 529–547

Zaccaro, S. J., Mumford, M. D., Connelly, M. S., Marks, M. A., & Gilbert, J. A. (2000). Assessment of leader problem-solving capabilities. *The Leadership Quarterly, 11*, 37-64.

Table 1
Illustrative Example of Intercultural Scenario and Possible Responses

Scenario:	A German business partner visits his Mexican partner to discuss a business proposal. During the meeting, the Mexican deals with many interruptions as assistants continue to come into the meeting to ask the Mexican partner for signatures or advice. Just when the German partner thinks they can finally focus on the proposal, an associate comes in to announce that a visitor was just stopping by the office to see the Mexican business partner.

At this point, the scenario freezes and respondents were asked one of two questions:

1. What action(s) would you take to continue the meeting, based on how the video ended? [general prompt condition]
2. What action(s) would you take to both complete the task and maintain the relationship, based on how the video ended? Consider both parties' perspective. Be as specific as possible. [specific prompt condition]

Responses:	- The German partner should ask his Mexican partner to cut off all distractions and finish their discussion. [task-focused / ethnocentric] - The German partner should wait for the Mexican partner to come back from seeing the visitor to continue the discussion. [relationship-focused] - The German partner should go along with the Mexican partner to meet the visitor and extend his social network in Mexico. Similarly, the Mexican partner could perhaps explain to the German partner that the visitor may become a valuable resource for their joint business activities. [task- & relationship-focused / solutions from multiple perspectives]

Table 2
Means and Standard Deviations in the Specific and General Prompt Conditions.

	Specific Prompt Condition		General Prompt Condition	
Variable	*M*	*SD*	*M*	*SD*
1. Intercultural Scenario Scores	2.51	0.47	2.35	0.45
2. Suggesting Solutions for Both Parties	0.64	0.18	0.23	0.26
3. Suggesting Ethnocentric Solutions	1.03	1.10	1.53	1.10
Dependent Variables				
4. Intercultural Performance	4.97	0.42	4.99	0.44
5. In-role Performance	6.37	0.42	6.44	0.31
6. Extra-role Performance	5.39	0.52	5.46	0.35
Independent Variables				
7. Openness to Experience	3.57	0.44	3.57	0.44
8. Perspective Taking	5.48	0.86	5.18	1.14
9. Cultural Knowledge	64.35	9.76	64.12	12.31
Control Variables				
10. Conscientiousness	3.40	0.57	3.41	0.50
11. Agreeableness	3.97	0.42	4.08	0.39
12. Emotional Stability	3.26	0.71	3.31	0.68
13. Extraversion	3.48	0.55	3.65	0.61
14. Sex (0 = male, 1 = female)	0.66	0.47	0.62	0.49
15. # of Languages Spoken	2.78	1.08	2.70	1.08
16. International Experience	1.80	0.82	1.80	0.90
17. Response Length	112.18	66.37	68.02	33.60

Note. $N = 158$ (specific response prompts). $N = 157$ (general response prompts).

Table 3
Correlations in the Specific and General Prompt Conditions.

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Intercultural Scenario Scores	—	.43	-.63	.44	.27	.28	.27	.28	.20	.16	.00	-.11	.01	.03	.01	-.04	.52
2. Suggesting Solutions for Both Parties	.37	—	-.38	.17	.07	.16	.12	.30	.05	-.02	-.07	-.05	-.10	-.03	.08	-.07	.68
3. Suggesting Ethnocentric Solutions	-.70	-.35	—	-.32	-.15	-.15	-.25	-.28	-.17	-.05	.05	.02	.01	-.04	-.07	.06	-.39
Dependent Variables																	
4. Intercultural Performance	.28	.02	-.12	(.86)	.36	.40	.22	.12	.14	.17	.03	.07	.22	.04	.06	.05	.21
5. In-role Performance	.41	.09	-.33	.24	(.92)	.41	.27	.03	.21	.16	.09	.01	.18	.18	.04	-.06	.27
6. Extra-role Performance	.23	-.08	-.10	.25	.51	(.89)	.13	.11	.08	.10	.19	.02	.13	.13	-.09	-.04	.22
Independent Variables																	
7. Openness to Experience	.07	.11	-.03	.21	.19	.09	(.82)	.06	.08	.21	.13	.18	.33	-.06	.08	.02	.18
8. Perspective Taking	.04	-.01	-.11	-.04	-.03	.09	.06	(.90)	-.02	.02	.14	-.10	.00	.01	.03	-.01	.27
9. Cultural Knowledge	.39	.16	-.28	.07	.33	.14	.09	.00	(.67)	.00	-.07	-.19	.03	.07	-.04	-.13	.19
Control Variables																	
10. Conscientiousness	.08	.06	-.09	.02	.18	.06	.19	.04	.07	(.82)	.23	.16	.15	.11	.11	-.07	.14
11. Agreeableness	.09	.12	-.09	.11	-.06	.15	.20	.28	-.12	.21	(.81)	.22	.32	.18	.01	.05	.02
12. Emotional Stability	.08	.10	-.05	.21	.05	.00	.24	.01	-.16	.15	.17	(.85)	.20	-.14	.12	.07	-.02
13. Extraversion	.02	.03	-.12	.18	-.03	.00	.35	.14	.06	.17	.37	.33	(.78)	.06	.04	.25	-.11
14. Sex (0 = male, 1 = female)	-.02	-.04	-.06	-.12	.13	.06	-.16	-.01	-.02	.08	.14	-.31	-.17	—	-.11	-.25	.07
15. # of Languages Spoken	-.13	-.10	.12	.07	-.05	.03	.00	.04	-.13	.15	.09	.11	.02	.08	—	.22	.04
16. International Experience	-.16	-.02	.07	.24	-.19	-.19	-.01	-.10	-.19	-.05	.08	.11	.27	-.11	.16	—	.10
17. Response Length	.32	.57	-.33	.00	.01	-.25	.21	-.03	.13	.07	.12	.07	.12	-.08	.02	.02	(.95)

Note. $N = 158$ (specific response prompts). $N = 157$ (general response prompts). Correlations for specific intercultural scenarios below the diagonal. Correlations for general intercultural scenarios above the diagonal. Alpha reliabilities are shown in parentheses along the diagonal. Alpha reliability of the mean ratings across all seven intercultural scenario scores: .76 (specific response prompts) / .75 (general response prompts). Correlations greater than .16 significant at $p < .05$; correlations greater than .21 significant at $p < .01$.

Table 4
Regression Results for Differential Construct Relations of Intercultural Scenario Scores in General and Specific Prompt Conditions.

	DV: Intercultural Scenario Scores			
Variables	Model 1		Model 2	
Intercept	2.37**	(.04)	2.37**	(.05)
Openness to Experience	.14**	(.06)	.26**	(.08)
Perspective Taking	.10**	(.02)	.14**	(.03)
Cultural Knowledge	.01**	(.00)	.01*	(.00)
Prompt Condition (0=general, 1=specific)	.13**	(.05)	.13**	(.05)
Openness to Experience * Prompt			-.22*	(.10)
Perspective Taking * Prompt			-.11*	(.05)
Cultural Knowledge * Prompt			.01**	(.00)
F	16.15**		11.96**	
df	(4,310)		(7,307)	
adjusted R^2	.16		.20	
ΔR^2			.04**	

Note. $N = 315$. Unstandardized coefficients reported with standard errors in parentheses.

* $p < .05$. ** $p < .01$.

Table 5

Regression Results for Differential Prediction of Performance Outcomes with Intercultural Scenario Scores in General and Specific Prompt Conditions.

	DV: Intercultural Performance		DV: In-role Performance		DV: Extra-role Performance	
Variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Intercept	4.98** (.02)	5.02** (.03)	6.41** (.02)	6.45** (.03)	5.43** (.02)	5.48** (.05)
Openness to Experience	.14** (.05)	.13* (.05)	.12** (.04)	.13** (.04)	.07 (.06)	.07 (.06)
Perspective Taking	-.01 (.02)	-.01 (.02)	-.03 (.02)	-.02 (.02)	.02 (.03)	.03 (.02)
Cultural Knowledge	.00 (.00)	.00 (.00)	.01** (.00)	.01** (.00)	.00 (.00)	.00 (.00)
Intercultural Scenario Scores	.30** (.05)	.40** (.08)	.21** (.04)	.14* (.06)	.18** (.06)	.16* (.08)
Prompt Condition (0=general, 1=specific)		-.07 (.05)		-.11** (.04)		-.11* (.05)
Intercultural Scenario Scores * Prompt		-.17 (.10)		.18* (.08)		.07 (.10)
F	12.72**	9.38**	15.17**	12.46**	4.91**	4.27**
df	(4,310)	(6,308)	(4,310)	(6,308)	(4,310)	(6,308)
adjusted R^2	.13	.14	.15	.18	.05	.06
ΔR^2		.01		.03**		.01

Note. $N = 315$. Unstandardized coefficients reported with standard errors in parentheses.

* $p < .05$. ** $p < .01$.