Journal of Applied Psychology
2015, Vol. 100, No. 2, 464–480

© 2014 American Psychological Association
0021-9010/15/$12.00 http://dx.doi.org/10.1037/a0038098

Putting Judging Situations Into Situational Judgment Tests: Evidence From Intercultural Multimedia SJTs

Thomas Rockstuhl, Soon Ang, and Kok-Yee Ng
Nanyang Technological University

Filip Lievens
Ghent University

Linn Van Dyne
Michigan State University

Although the term situational judgment test (SJT) implies judging situations, existing SJTs focus more on judging the effectiveness of different response options (i.e., response judgment) and less on how people perceive and interpret situations (i.e., situational judgment). We expand the traditional SJT paradigm and propose that adding explicit assessments of situational judgment to SJTs will provide incremental information beyond that provided by response judgment. We test this hypothesis across 4 studies using intercultural multimedia SJTs. Study 1 uses verbal protocol analysis to discover the situational judgments people make when responding to SJT items. Study 2 shows situational judgment predicts time-lagged, peer-rated task performance and interpersonal citizenship among undergraduate seniors over and above response judgment and other established predictors. Study 3 shows providing situational judgment did not affect the predictive validity of response judgment. Study 4 replicates Study 2 in a working adult sample. We discuss implications for SJT theory as well as the practical implications of putting judging situations back into SJTs.

Keywords: situational judgment test, intercultural skills, performance, verbal protocol analysis

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Most people remember having to solve story problems on math tests. Solving these problems requires complex intermediate judgments to arrive at the final response (Newell & Simon, 1972). For example, students first make judgments about the situation in the story (e.g., representing the situation as a set of equations) before applying mathematical reasoning to solve the equations. Nevertheless, grades typically depend only on whether final responses are correct. This is problematic because feedback that focuses only on final responses provides little diagnostic information about intermediate judgments. Tracing intermediate judgments illuminates individual differences in judgment and suggests interventions to improve such judgment (Weber & Johnson, 2009).

Similar to math problems in school, selection and assessment procedures such as situational judgment tests (SJTs) typically focus on final responses. SJTs ask test-takers to rate the quality of multiple response options or identify the best option to written or video-based work-related situations (Weekley & Ployhart, 2006). Although rating response options requires complex judgments about both the situation and possible responses (Ployhart, 2006; Schmitt & Chan, 2006), SJTs typically capture information about the final response only. As a result, we do not know much about the judgments that people make of the situation that lead to their final response.

Noting this limitation, SJT scholars have repeatedly called for research to examine the intermediate judgments that people make about the situation when responding to SJT items (Ployhart, 2006; Whetzel & McDaniel, 2009). For example, Ployhart (2006) called for research to open the black box of situational judgment in SJTs. Whetzel and McDaniel (2009) noted that "research into the factors considered by a respondent in evaluating an item may prove useful in understanding the constructs assessed by the item and may lead to novel ways of scoring SJT items" (p. 199). Ployhart (2006) further suggested that to open the black box of situational judgment, "it would be informative to conduct a protocol tracing analysis as respondents complete SJTs. Test takers could describe orally what they are doing mentally as they complete the SJT" (p. 102). In sum, SJT research stands to benefit substantially from a better understanding of how people subjectively interpret situations in SJTs. In fact, one may even argue that the term *situational judgment test* is a misnomer without an explicit assessment of judgments about the situation.

In this article, we build on the strong foundation of existing SJT research and propose that adding an assessment of how people perceive and interpret situations (situational judgment) to SJTs will provide incremental information beyond the typical focus on choosing the best response (response judgment). We demonstrate the value of this novel approach with four studies and multiple samples.

This article was published Online First October 6, 2014.

Thomas Rockstuhl, Soon Ang, and Kok-Yee Ng, Division of Strategy, Management, and Organization, Nanyang Business School, Nanyang Technological University; Filip Lievens, Department of Personnel Management, Work, and Organizational Psychology, Ghent University; Linn Van Dyne, Department of Management, Michigan State University.

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

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

In Study 1, we follow the recommendations of Ployhart (2006) and use verbal protocol analysis to uncover the kinds of situational judgments that people make when responding to SJTs. We draw upon the results from Study 1 to operationalize situational judgment in subsequent studies. In Study 2, we examine the incremental validity of situational judgment over and above response judgment in predicting task performance and interpersonal citizenship (organizational citizenship behavior [OCB]). We focus on task performance and interpersonal OCB for two reasons. First, meta-analyses show that SJTs are important predictors of both types of performance (Christian, Edwards, & Bradley, 2010). Second, assessing the contributions of situational judgment and response judgment to predicting task performance and interpersonal OCB may deepen our understanding of why SJTs predict both types of performance. In Study 3, we assess whether adding situational judgment in SJTs could inadvertently affect the predictive validity of response judgment, for example, due to cognitive fatigue, learning, context-, or accessibility effects (e.g., Feldman & Lynch, 1988). In Study 4, we replicate and extend Study 2 by testing hypotheses with a different population and controlling for additional individual characteristics.

Overall, this set of four studies should contribute to existing SJT research in three key ways. First, we expand the SJT paradigm by putting situational judgment back into situational judgment tests. We do this by asking people explicitly to make judgments about the situation, in addition to making response judgments. Second, we test and show incremental predictive validity of situational judgment, over and above response judgment. Third, we compare and show the relative importance of situational judgment and response judgment as predictors of two key performance outcomes: task performance and interpersonal OCB.

The nature and type of situations tested by any SJT represent boundary conditions (Schmitt & Chan, 2006). Most SJTs focus on interpersonal situations (Christian et al., 2010). In this research, we focus specifically on intercultural interpersonal situations. This is because intercultural interactions present especially challenging interpersonal situations (Ang & Van Dyne, 2008; Gelfand, Erez, & Aycan, 2007) and are prone to misjudgments (Earley & Ang, 2003; Triandis, 2006). The difficulty of judging intercultural situations makes such situations an appropriate context to uncover types of situational judgments. Moreover, with the growing diversity in the workplace (Shore et al., 2011), having an intercultural SJT that predicts performance outcomes in culturally diverse contexts should make a significant practical contribution to the selection literature (Lievens, 2006).

Theory and Hypotheses

What Is Situational Judgment?

Motowidlo and colleagues (Motowidlo & Beier, 2010; Motowidlo, Hooper, & Jackson, 2006) proposed that the typical assessment of response judgment in SJTs measures "procedural knowledge about effective action in work situations described by the SJT" (Motowidlo & Beier, 2010, p. 321). Thus, response judgments in SJTs focus on assessments of the appropriateness of various response options presented for a specific situation.

Recent theorizing about SJTs suggests that response judgment represents only one aspect of judgment in SJTs (Ployhart, 2006). For instance, Ployhart (2006) suggested that before respondents can arrive at a response judgment, they must first comprehend the situation, especially when the situation in the SJT is filled with ambiguous cues and incomplete information. These initial attempts to comprehend the situation represent situational judgments, which according to attribution theory (Heider, 1958; Weiner, 1995) are distinct from response judgments.

Drawing upon attribution theory, we refer to situational judgments as individuals' sense-making of a situation, which enables them to comprehend, explain, attribute, extrapolate, and predict situations. Response judgments, on the other hand, refer to judgments about the most appropriate response after evaluating the costs and benefits of available response options. Although distinct, the two types of judgment are related. As Heider (1958) noted, "a person reacts to *what he thinks the other person is perceiving, feeling, and thinking*" (p. 1; emphasis added). This highlights that a person's situational judgments (e.g., what I think the other person is perceiving, feeling, and thinking) play a critical role in shaping the person's response judgments (e.g., what is an appropriate action to take; Jansen et al., 2013).

The distinction between situational judgment and response judgment suggests that both types of judgment provide unique and complementary information about individuals. In support of this argument, Magnusson and Ekehammar (1975) empirically found two forms of judgments—judgments of the situation and judgments of responses to the situation. In some instances, people judge the situation differently but respond similarly; in other instances, people may judge the situation similarly but respond differently. Accordingly, a central hypothesis in our research is that people's judgments of the situation provide incremental explanatory power, over and above their response judgments, in predicting job performance. Below, we elaborate on the relationships between situational judgment and response judgments in predicting task performance and interpersonal OCB in intercultural situations, which provide the context to our study.

Situational Judgment and Task Performance

As noted above, Heider (1958) suggested that judgments about interpersonal situations involve judgments about what others are perceiving, feeling, and thinking. People make these judgments to understand others' expectations about appropriate behaviors, which helps them to weigh the benefits and costs of different response options (Jansen et al., 2013).

Understanding others' expectations is particularly important in intercultural situations because culture influences expectations of what is an appropriate behavior in different contexts (Triandis, 2006). For example, research shows that culture influences expectations about appropriate leadership (House, Hanges, Javidan, Dorfman, & Gupta, 2004) and team (Gibson & Zellmer-Bruhn, 2001) behaviors. At the same time, people often do not directly communicate their expectations of effective behaviors, thus making intercultural interactions particularly challenging (Molinsky, 2013). To effectively fulfill role expectations related to task performance, individuals first need to infer accurately what culturally diverse others consider as appropriate behaviors.

Hence, we expect that keen situational judgment will predict task performance in intercultural contexts, above and beyond response judgment. The more accurately individuals can observe,

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

interpret and explain what is happening in the situation (i.e., situational judgment), the more they can adjust their behaviors according to role expectations to achieve their work outcomes (Stone-Romero, Stone, & Salas, 2003). For example, a Western manager with keen situational judgment might be sensitive to the appropriate time to speak up during meetings with Asians. Likewise, an Asian manager with keen situational judgment might deduce when Western managers prefer to make decisions on the spot and minimize the time spent involving others in discussion. In sum, we expect individuals with better situational judgment to be more effective in their tasks in intercultural situations, after taking into account response judgment.

> *Hypothesis 1:* Situational judgment predicts task performance over and above response judgment.

Situational Judgment and Interpersonal OCB

Situational judgments about what others are perceiving, feeling, and thinking should also complement response judgment as a predictor of interpersonal OCB. In their meta-analysis of the predictive validity of SJTs, Christian et al. (2010) suggested that response judgments in interpersonal SJTs relate to interpersonal OCB to the extent that they “reflect the ability to perceive and interpret social dynamics in such a way that facilitates judgments regarding the timing and appropriateness of contextual behaviors” (p. 92). This reasoning implies that response judgments in interpersonal SJTs inherently involve situational judgments, and that situational judgments are critical in helping individuals know when and how to engage in interpersonal citizenship behaviors such as helping.

Knowing when and how to help others is particularly relevant in intercultural contexts because cultural differences in emotional expressions (Elfenbein & Ambady, 2002) and communication directness (Spencer-Oatey, 2008) can make it hard to judge when others need help. Cultural differences also affect expectations about what constitutes appropriate helping (Farh, Zhong, & Organ, 2004).

Building on Christian et al.’s (2000) arguments, we expect situational judgment to predict interpersonal OCB in intercultural contexts, even after controlling for response judgment. Individuals who can more accurately observe, interpret, and explain behaviors in intercultural situations are more likely to discern when culturally diverse others would appreciate help and what kinds of help would be appropriate.

> *Hypothesis 2:* Situational judgment predicts interpersonal OCB over and above response judgment.

Study 1

As highlighted in the introduction, we do not know much about the nature of situational judgment in SJTs or how to operationalize situational judgment. To better understand the nature of situational judgment in SJTs, we conducted a verbal protocol analysis (Ericsson & Simon, 1993) of people’s thought processes as they completed the SJT. Specifically, we asked individuals to “think out loud” as they decided how to respond to various SJT items. This verbal protocol approach allowed us to access the thought processes of SJT respondents in real time. We then content-analyzed transcriptions of these verbal protocols to uncover the kinds of situational judgments that respondents make. Our goal was to identify which aspects of the situation in the SJT items respondents used to help them arrive at a possible response to the situation. We used the results from Study 1 to guide our operationalization of situational judgment in subsequent studies.

Method

Participants. Twelve international managers with professional links to the university’s research center provided data for Study 1. Managers (50% female; mean age = 30.6 years; *SD* = 5.88 years) came from a variety of Asian and Western countries: China (two), France (one), Germany (two), Japan (one), Malaysia (one), Philippines (one), Singapore (two), and the United States (two). On average, managers had 5.6 years (*SD* = 5.30) of previous international work experience.

Procedure. The 12 managers came to the experimental labs at the university and learned to produce verbal protocols. Each session consisted of 20-min practice verbal protocol exercises adapted from Ericsson and Simon (1993). Next, they completed verbal protocols on four randomly ordered, multimedia SJT items depicting intercultural situations (the Appendix describes the development of the intercultural SJT). Managers first watched an intercultural SJT item. We then asked them to “think aloud: What would you do next in the situation you have just seen?” We gave managers as much time as they needed. Verbal protocols lasted between 2 and 8 min (*M* = 4.27 min; *SD* = 2.43 min) for each SJT item.

Coding. The 12 managers generated 48 verbal protocols (12 managers × 4 SJT items). We audio-recorded the verbal protocols. A professional transcription service then transcribed the verbal protocols. The first two authors randomly selected one transcribed verbal protocol at a time and identified categories of situational judgment types. We first independently read each verbal protocol and then discussed types of situational judgment from that protocol. By the end of the sixth protocol, we had identified 11 situational judgments (see Table 1 for definitions and examples) and reached theoretical saturation, as we could no longer find additional situational judgment categories (Strauss & Corbin, 1990).

We then recruited and trained two research assistants (Raters A and B) with the six protocols used to generate the 11 categories. Each rater then independently coded the remaining 42 transcribed protocols. The first author met with the two raters regularly to assess interrater agreement and resolve differences. All 11 situational judgment categories yielded acceptable interrater agreement using Cohen’s kappa (range: .74–.93; see Table 1). All agreement indices exceed the .60 threshold (Landis & Koch, 1977). The overall Cohen’s kappa across all 48 verbal protocols and 11 situational judgment types was .81.

Results and Discussion

Table 1 summarizes the coding results. Three hundred twelve, or approximately 82%, of the 382 situational judgments reflect three dominant categories:

1. intentions (*n* = 158; 41.4%);

2. emotions (*n* = 90, 23.6%); and

Table 1
Interrater Agreement and Frequencies of Situational Judgment Types in Verbal Protocol Analysis (Study 1)

Type of situational judgment	Definition	Example from verbal protocols	Cohen's κ	Overall frequency
1. Intentions	What someone plans to do or achieve; an aim or purpose	"X wants to get things done"; "X wants to build relationship"	.81	158 (41.4%)
2. Emotions	A strong feeling (such as love, anger, joy, hate, or fear)	"X feels annoyed"; "X is relaxed"	.93	90 (23.6%)
3. Thoughts	An idea, plan, opinion, picture, etc., that is formed in someone's mind	"X thought that everything was worked out"; "X is thinking about how to let Y know"	.78	64 (16.8%)
4. Relationship	The way in which two or more people or things are connected	"Because both are peers"; "Because X is the client"	.87	20 (5.2%)
5. Situational constraints	An aspect of the situation that limits or restricts someone's actions or behavior	"X is under time pressure"; "Because of their budget constraints"	.85	18 (4.7%)
6. Traits	An enduring quality that makes one person different from another	"X is straight-forward"; "X is messy"	.83	14 (3.7%)
7. Knowledge	Awareness of something; the state of being aware of something	"X had no idea that the contact had changed"	.79	5 (1.3%)
8. Abilities	The power or skill to do something	"X is not yet able to make such decisions"	.74	4 (1.0%)
9. Effort	A serious attempt to do something	"X is not trying hard enough"	.74	4 (1.0%)
10. Prior behavior	Inferred actions or behaviors that occurred prior to the episode	"X did not even inform Y before the meeting"	.80	3 (0.8%)
11. Situational forecasting	Predictions about what will happen in the future	"X is not going to send out the e-mail"	.80	2 (0.5%)
Total				382 (100.0%)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

3. thoughts ($n = 64$, 16.8%) of the parties in the intercultural SJT item.

The remaining 70 (approximately 18%) situational judgments clustered into eight other categories reflecting

4. the relationship between the parties ($n = 20$, 5.2%);
5. situational constraints faced by the parties ($n = 18$, 4.7%);
6. parties' traits ($n = 14$, 3.7%);
7. parties' knowledge about the situation of their counterpart ($n = 5$, 1.3%);
8. parties' abilities ($n = 4$, 1.0%);
9. parties' effort ($n = 4$, 1.0%);
10. prior behaviors ($n = 3$, 0.8%); and
11. situational forecasting ($n = 2$, 0.5%).

Given that intentions, emotions, and thoughts composed 82% of the situation judgments made by respondents in Study 1, we assessed situational judgment in our subsequent studies by asking participants to describe the thoughts, emotions, and intentions of the parties in the intercultural SJT items. In the next study, we test whether situational judgment predicts task performance and interpersonal OCBs beyond response judgment typically used in SJTs.

Study 2

Method

Participants. One hundred thirty-two ($n = 132$) university seniors participated in this study. They were drawn from an international organizational behavior course at a large business school in Singapore. Participant's mean age was 22 years (*SD* = 1.4 years), and 67% were female. Participants were diverse, representing 24 countries across five continents. On average, each participant had 3.9 years (*SD* = 1.28 years) of previous work experience and had traveled to 10 countries (*SD* = 7.67); 94% of them spoke at least two languages.

Procedures. We randomly assigned students to 19 culturally diverse teams (six to eight members per team). Teams were highly diverse in national cultures (average Blau's [1977] index of heterogeneity = .82, *SD* = 0.12). Teams were self-managed and were not assigned formal leaders. Teams worked on an intensive 3-month team project. The goal of the project was to produce a 10-min multimedia dramatization of a challenging intercultural interaction.

We collected data at three points during the course. At Time 1 (beginning of the course), team members completed online surveys of personality, cognitive ability, and demographic characteristics. At Time 2 (2 weeks into the course, at the beginning of their project), they completed the intercultural SJT. At Time 3 (at the end of their project), peers rated task performance and interpersonal OCB of team members using a full round-robin design (i.e.,

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

every team member rated all other team members; Kenny & La Voie, 1984).

Measures. Unless otherwise indicated, we measured all variables using Likert-type scales anchored at 1 = *strongly disagree* and 7 = *strongly agree*.

Criterion measures: Task performance and interpersonal OCB. Team members assessed task performance with three in-role behavior items (e.g., fulfilled responsibilities of the project; $r_{WG(J)} = .90$; $\alpha = .89$) adapted from Williams and Anderson (1991). Team members rated interpersonal OCB using three items (e.g., assisted other group members with their work; $r_{WG(J)} = .85$; $\alpha = .80$) from Van Dyne and LePine (1998).

We conducted confirmatory factor analyses of both criterion measures using LISREL 8.8 (Jöreskog & Sörbom, 1996). The hypothesized two-factor model (task performance and interpersonal OCB) showed excellent fit: $\chi^2(8df) = 3.97$, *ns*; $\chi^2/df = .50$, *IFI* = .99, *RMSEA* = .01. All factor loadings were statistically significant (.72 – .96, $p < .01$). The two-factor model showed significantly better fit than a single-factor model, $\Delta\chi^2(1df) = 141.12$, $p < .001$.

Predictor measures: Situational judgment and response judgment. The intercultural SJT (see Appendix) contained seven multimedia items that participants watched in randomized order. After each SJT item, they answered two constructed-response (open-ended) questions—one assessing situational judgment and the other, response judgment. On the basis of the results from Study 1, we elicited situational judgment with the following question: "What are the thoughts, feelings, and intentions of the people in the video?" We elicited response judgment with the following question: "What would you do next in the situation you have just seen?"

We employed four research assistants, all blind to the hypotheses, for Study 2. The original two raters, A and B, from Study 1 rated the quality of situational judgment. Raters scored situational judgment using the following holistic rating scale (e.g., Byham, 1977): "To what extent does this response demonstrate understanding of the thoughts, feelings, and intentions of the parties in the vignette? (1 = *not at all*; 2 = *little*; 3 = *moderate*; 4 = *well*; 5 = *very well*)." Raters C and D rated response judgments, using the following holistic 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*)."

We provided frame-of-reference training to all raters according to procedures outlined by Pulakos (1984). We provided raters with definitions and scale anchors, as well as annotated scripts and behavioral examples (Smith & Kendall, 1963) of effective and ineffective situational and response judgments for each SJT item. For situational judgment, we generated behavioral examples of effective responses based on point of view interpretations of each culture depicted in each item. For response judgment, we generated behavioral examples of effective and ineffective responses based on the dual concern model of conflict management (Thomas, 1976). Next, raters discussed the information. We then presented and discussed example responses that represented different levels of performance (i.e., good situational and response judgments vs. poor situational and response judgments). Raters then practiced making ratings in response to practice responses, and we provided them with feedback. Each rater then independently began rating actual responses. We met with all raters after the first 10 responses to assess interrater agreement and resolve differences.

We assessed interrater agreement between Raters A and B as well as between Raters C and D using intraclass correlations (ICC2.1; Shrout & Fleiss, 1979). The average intraclass correlation was .88 for situational judgment and .77 for response judgment. These ICC2.1 statistics met LeBreton and Senter's (2007) cutoff of .70. We averaged ratings of situational judgment across Raters A and B ($\alpha = .92$), and ratings of response judgment across Raters C and D ($\alpha = .79$).

Table 2 shows that the hypothesized two-factor (situational judgment and response judgment) measurement model with correlated uniqueness factors between indicators from the same SJT items (Lance, Woehr, & Meade, 2007) showed good fit to the data: $\chi^2(69df) = 88.73$, *ns*; $\chi^2/df = 1.29$, *IFI* = .98, *RMSEA* = .05. All factor loadings were statistically significant (.46–.86, $p < .01$).

We compared the hypothesized model with alternative models. The hypothesized two-factor model was a significantly better fit than (a) a two-factor model that did not allow for correlated uniqueness factors (Model 2: $\Delta\chi^2(7df) = 71.14$, $p < .001$); (b) a single-factor model with correlated uniqueness factors (Model 3: $\Delta\chi^2(1df) = 28.65$, $p < .001$); and (c) a single-factor model without correlated uniqueness factors (Model 4: $\Delta\chi^2[8df] = 101.96$, $p < .001$). Together, these results support the discriminant validity of situational judgment and response judgment.

Control measures. We included cognitive ability, Big Five personality, and previous work experience (in years) as in prior

Table 2
CFA Comparisons of Alternative Nested Model Fit for Situational and Response Judgments (Study 2)

Model	Description	χ^2	*df*	χ^2/df	IFI	RMSEA	$\Delta\chi^2$	Δdf
Model 1	Hypothesized two-factor model with correlated uniqueness factors for the same SJT items	88.73	69	1.29	.98	.047		
	Alternate nested models compared to the hypothesized two-factor model							
Model 2	Two-factor model without correlated uniqueness factors for the same SJT items	159.87***	76	2.10	.94	.092	71.14***	7
Model 3	Single-factor model with correlated uniqueness factors for the same SJT items	117.38***	70	1.68	.94	.094	28.65***	1
Model 4	Single-factor model without correlated uniqueness factors for the same SJT items	190.69***	77	2.48	.90	.110	101.96***	8

Note. *N* = 132. IFI = incremental fit index; RMSEA = root-mean-square error of approximation.
*** $p < .001$.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

SJT research (Lievens, Peeters, & Schollaert, 2008; McDaniel, Hartman, Whetzel, & Grubb, 2007). We measured cognitive ability with the Wonderlic Personnel Test (Wonderlic, 1999); Big Five personality traits with Goldberg's (1999) 50-item IPIP-FFM: extraversion (10 items, α = .90), agreeableness (10 items, α = .76), conscientiousness (10 items, α = .81), emotional stability (10 items, α = .88), and openness to experience (10 items, α = .84). We also controlled for gender (0 = male, 1 = female), language proficiency (total number of languages spoken), and international experience (number of countries visited).

Analytic strategy. We used the software program SOREMO (Kenny, 1995) for social relations model analyses of the round-robin (i.e., peer) data on task performance and interpersonal OCB. SOREMO calculates target scores, which are an index of how that individual was perceived by others in the group, for each participant. In calculating target scores, SOREMO removed group differences (i.e., subtracting the group mean from the average of all peer ratings for a target person), making target scores statistically independent of group membership and thus appropriate for statistical procedures that assume independence (see Kenny & La Voie, 1984).

To support the estimation of individual target scores, SOREMO partitions the overall variance in the dependent variable into variance attributable to different sources. The proportion of the total variance in peer ratings that is attributable to group effects is referred to as group variance. The proportion of the total variance in peer ratings that is attributable to characteristics of targets (i.e., focal team members) is referred to as target variance (Kenny & La Voie, 1984). SOREMO also produces a significance test for these variances based on a *Z* test (Kenny, 1995). Nonsignificant group variances support the removal of group differences, whereas significant target variances indicate substantial interrater agreement and support the estimation of individual target scores (Kenny & La Voie, 1984).

Group variances were not significant for either task performance (group variance = .12, $Z = 1.60$, *ns*) or interpersonal OCB (group variance = .11, $Z = 1.60$, *ns*). At the same time, target variance was statistically significant for task performance (target variance = .13, $Z = 3.67$, $p < .01$) and interpersonal OCB (target variance = .09, $Z = 3.13$, $p < .01$). We therefore estimated individual target scores for both task performance and interpersonal OCB.

We tested our hypotheses using hierarchical regression analyses. We entered control variables (i.e., gender, language proficiency, international experience, work experience, cognitive ability, and Big Five personality) in the first step, response judgment in the second, and situational judgment in the third. We conducted relative weights analysis (Johnson & LeBreton, 2004) to assess the relative importance of situational judgment vis-à-vis response judgment.

Results and Discussion

Descriptive statistics. Means, standard deviations, and correlations appear in Table 3. Situational judgment was positively related to task performance ($r = .40$, $p < .01$) and interpersonal OCB ($r = .38$, $p < .01$), providing preliminary support for the role of situational judgment in predicting performance outcomes.

Response judgment correlated positively with task performance ($r = .44$, $p < .01$) and interpersonal OCB ($r = .28$, $p < .01$). These findings are consistent with prior SJT research. For instance,

Table 3
Means, Standard Deviations, and Correlations (Study 2)

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Criterion measures																
1. Task Performance	.01	.41	(.89)													
2. Interpersonal OCB	.00	.39	.54**	(.80)												
Predictor measures																
3. Situational Judgment	3.02	.67	.40**	.38**	(.92)											
4. Response Judgment	2.28	.48	.44**	.28**	.48**	(.79)										
Control measures																
5. Extraversion	3.58	.66	.17*	.07	−.03	.09	(.90)									
6. Agreeableness	4.03	.41	.13	.31**	.16	.05	.28**	(.76)								
7. Conscientiousness	3.42	.66	.26**	.16	.02	.27**	.06	.12	(.81)							
8. Emotional Stability	3.41	.71	−.03	−.04	.04	.01	.13	.06	.19*	(.88)						
9. Openness to Experience	3.64	.57	.15	.15	.23**	.29**	.27**	.15	.19*	.17	(.84)					
10. Cognitive Ability	23.14	5.22	.26**	.20*	.30**	.42**	−.03	.24**	.24**	.11	.09	—				
11. Work experience (in years)	3.89	1.28	−.15	−.06	−.12	−.01	.32**	−.07	.18*	.08	.19*	−.30**	—			
12. International experience	10.20	7.67	.10	.19*	.22*	.10	.36**	.10	−.02	.07	.14	−.17	.47**	—		
13. Number of languages spoken	2.75	1.13	.16	.03	.07	.10	.10	.00	.11	.03	.09	−.02	.13	.15	—	
14. Sex (0 = male, 1 = female)	.67	.47	.10	.17	.07	.06	.00	.18*	.18*	−.13	−.18*	.16	−.27**	−.11	−.02	—

Note. $N = 132$. Alpha reliabilities are shown in parentheses along the diagonal.
* $p < .05$. ** $p < .01$.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

meta-analyses of SJT research (based on interpersonal SJTs and supervisor ratings of performance) reported correlations of .21 between response judgment and task performance, and .25 between response judgment and contextual performance (Christian et al., 2010).

Hypotheses tests. We proposed that situational judgment would predict task performance (H1) and interpersonal OCB (H2), over and above response judgment. Results support H1. Situational judgment predicted task performance (β = .26, $p < .01$; Model 3, Table 4) and explained an additional 4% variance in task performance ($p < .01$) over and above response judgment and controls. Relative weights analysis shows that response judgment accounted for 25.3% (95% CI = 8.3%, 38.6%; $p < .05$) and situational judgment accounted for 24.5% (95% CI = 5.5%, 55.1%; $p < .05$) of explained variance in task performance. Relative weights for situational and response judgments were not significantly different from each other (95% CI = −41.3%, 27.5%; *ns*), suggesting that the two judgments are not significantly different in their importance as predictors of task performance. Results without control variables replicated these findings.

Our results also support H2. Situational judgment predicted interpersonal OCB (β = .24, $p < .05$; Model 6), explaining an additional 4% variance ($p < .01$) over and above response judgment and controls. Relative weights analysis shows situational judgment accounted for 30.0% (95% CI = 9.1%, 64.6%; $p < .05$) of explained variance in interpersonal OCB. By contrast, response judgment accounted for 11.5% (95% CI = 3%, 41.8%; *ns*). This difference in relative weights between situational and response judgment approached the traditional level of statistical significance (90% CI = −51.0%, −4.4%; $p < .06$), suggesting that situational judgment may be a more important predictor of interpersonal OCB than response judgment. Results without control variables replicated these findings.

Taken together, these analyses support a key assumption motivating our research—namely, that adding an explicit assessment of situational judgment provides incremental validity for predicting two key types of performance (task performance and interpersonal OCB). The pattern of findings also suggests that beyond the joint variance shared between situational and response judgments: (a) both situational and response judgments significantly predict unique variance in task performance; and (b) situational judgment significantly explains unique variance in interpersonal OCB but response judgment does not. These results provide initial support for our argument that response judgment in SJTs might predict interpersonal OCB because it inherently assesses situational judgment.

We realize that typical SJTs ask only about response judgment. It is therefore possible that adding situational judgment may have affected the predictive validity of response judgment in this study. For example, this could be due to cognitive fatigue, learning, context, or accessibility effects (e.g., Feldman & Lynch, 1988). Thus, we wanted to explore whether asking respondents to make situational judgments might have eroded the predictive validity of response judgment in Study 2. To this end, we conducted Study 3 with a similar sample and had respondents make response judgments but not make situational judgments.

Study 3

Method

Participants. Eighty-nine (n = 89) university seniors from a different cohort of the same international organizational behavior course as Study 2 provided data. Participants' mean age was 22 years (*SD* = 1.5 years), and 55% were female. Participants were diverse, representing 17 countries across five continents. On average, each participant had 4.2 years (*SD* = 2.06 years) of previous work experience and had traveled to 13 countries (*SD* = 8.30); 91% spoke at least two languages.

Procedures. We used the same task, procedures, and measures as Study 2. Culturally diverse teams (n = 13; with 6–8 members per team; average index of nationality heterogeneity =

Table 4
Hierarchical Linear Regression Results and Relative Weights Analysis (Study 2)

	Task performance				Interpersonal OCB			
Variable	Model 1	Model 2	Model 3	RW	Model 4	Model 5	Model 6	RW
Sex	−.04	−.05	−.06	0.6%	.10	.09	.08	5.8%
Number of languages spoken	.12	.11	.10	4.5%	−.00	−.01	−.02	0.1%
International experience	.21*	.15	.07	2.8%	.28**	.25**	.17	10.7%
Work experience	−.35**	−.33**	−.29**	10.0%	−.13	−.12	−.08	2.4%
Extraversion	.19*	.18*	.23*	8.4%	−.08	−.08	−.04	1.0%
Agreeableness	−.04	.01	−.02	1.0%	.22*	.25**	.22*	23.5%
Conscientiousness	.29**	.23**	.27**	14.7%	.12	.09	.12	5.7%
Emotional stability	−.13	−.10	−.11	1.5%	−.10	−.08	−.08	1.8%
Openness to experience	.08	−.00	−.04	1.6%	.12	.08	.04	3.3%
Cognitive ability	.15	.02	−.01	5.2%	.11	.03	.00	4.3%
Response judgment		.33**	.23*	25.3%		.19*	.09	11.5%
Situational judgment			.26**	24.5%			.24*	30.0%
F	3.94** (10,121)	5.15** (11,120)	5.66** (12,119)		3.09** (10,121)	3.22** (11,120)	3.54** (12,119)	
R^2	.25	.32	.36		.20	.22	.26	
ΔR^2		.07**	.04**			.02*	.04*	
adjusted R^2	.18	.26	.30		.14	.16	.19	

Note. N = 132. Table reports standardized beta coefficients. RW = relative weight (%) of R^2.
* $p < .05$. ** $p < .01$.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

.86, *SD* = 0.10) produced multimedia dramatizations of challenging intercultural interactions.

Time 1 data included personality, cognitive ability, and demographics. At Time 2, participants completed intercultural SJT response judgment but did not complete situational judgment. At Time 3, peers provided round-robin ratings of task performance and interpersonal OCB.

Measures. We used the same measures as Study 2 for substantive constructs. Table 5 reports descriptive statistics, internal consistency reliability (alpha) coefficients, and correlations.

Criterion measures: Task performance and interpersonal OCB. Peers rated task performance ($r_{WG(J)}$ = .87; α = .91) and interpersonal OCB ($r_{WG(J)}$ = .81; α = .85). Group variances were not significant for either task performance (group variance = .09, *Z* = 1.05, *ns*) or interpersonal OCB (group variance = .10, *Z* = 1.21, *ns*), supporting the removal of group differences for the estimation of target scores. In addition, target scores showed agreement between peer-ratings and statistically significant amounts of target variance for task performance (target variance = .19, *Z* = 3.12, $p < .01$ and interpersonal OCB (target variance = .14, *Z* = 2.89, $p < .01$).

Predictor measure: Response judgment. Two research assistants assessed the quality of response judgment. Interrater agreement exceeded .70 (ICC2.1 = .81) so we averaged ratings of response judgment (α = .75) across raters.

Control measures. We controlled for cognitive ability (Wonderlic, 1999), Big Five (α = .72 – .90), previous work experience, international experience, number of languages spoken, and gender.

Analytic strategy. Following recommendations of Byrne, Shavelson, and Muthén (1989), we conducted a series of multigroup confirmatory factor analyses to assess the equivalence of response judgments across the samples in Studies 2 and 3. First, we tested the hypothesized model for each sample separately. Second, we tested for configural invariance between the samples. Third, we tested for metric invariance of response judgment between samples by fixing factor loadings to be equal across the samples. Finally, we tested for predictive invariance by fixing relationships of response judgment with task performance and interpersonal OCB to be equal across both samples.

Results and Discussion

Descriptive statistics. Table 5 reports means, standard deviations, and correlations. Response judgment was positively related to task performance ($r = .38$, $p < .01$) and interpersonal OCB ($r = .29$, $p < .01$). The strength of these correlations is not significantly different from those observed in Study 2 (task performance: $r = .44$, $Z = -.52$, *ns*; interpersonal OCB: $r = .28$, $Z = .08$, *ns*).

Multivariate regression analyses. Table 6 summarizes results of hierarchical linear regression. Response judgment predicted task performance (β = .33, $p < .01$; Model 2, Table 6) and interpersonal OCB (β = .26, $p < .05$; Model 4, Table 6), after accounting for the controls. Relationships were similar to those in Study 2 (task performance: β = .33, $p < .01$; interpersonal OCB: β = .19, $p < .01$).

Multigroup confirmatory factor analyses. The hypothesized three-factor model (task performance, interpersonal OCB, and response judgment) had good fit to the data in both samples (Study 2: $\chi^2[62df] = 87.54$, $p < .05$, $\chi^2/df = 1.41$, *IFI* = .97, *RMSEA* = .06; Study 3: χ^2[62df] = 97.08, $p < .01$, $\chi^2/df = 1.57$, *IFI* = .95, *RMSEA* = .08). Results further show configural invariance between both samples: $\chi^2[124df] = 144.18$, ns, $\chi^2/df = 1.16$, *IFI* = .98, *RMSEA* = .04 as well as metric invariance ($\Delta\chi^2[6df] = 3.98$, ns). Fixing relationships of response judgment with task performance and interpersonal OCB to be equal across the two samples did not worsen fit ($\Delta\chi^2[2df] = 4.91$, *ns*). Taken together, these results suggest that asking respondents to make situational judgments first (as in Study 2) did not significantly alter the predictive validity of response judgments.

In our final study, we sought to replicate and extend Study 2 using a sample of working adults to strengthen confidence in the

Table 5
Means, Standard Deviations, and Correlations (Study 3)

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13
Criterion measures															
1. Task Performance	−.02	.54	(.91)												
2. Interpersonal OCB	.01	.52	.56**	(.85)											
Predictor Measure															
3. Response Judgment	2.35	.41	.38**	.29**	(.75)										
Control measures															
4. Extraversion	3.60	.65	.05	.00	.03	(.90)									
5. Agreeableness	4.08	.45	.07	.16	−.04	.28**	(.81)								
6. Conscientiousness	3.37	.59	.22*	.12	.05	.11	.28**	(.80)							
7. Emotional Stability	3.34	.65	−.01	.04	.05	.19	.14	.09	(.85)						
8. Openness to Experience	3.55	.44	.01	−.07	.08	.28**	.11	.04	.13	(.72)					
9. Cognitive Ability	24.91	7.72	.29**	.23*	.23*	−.13	−.16	−.07	−.20	.06	—				
10. Work Experience (in years)	4.18	2.06	.08	−.02	−.02	.21*	.19	.07	.26*	.07	−.27**	—			
11. International Experience	13.45	8.30	.04	.01	−.08	.39**	.32**	.02	.12	.20	−.24*	.39**	—		
12. Number of Languages Spoken	2.71	.96	.01	−.03	.08	−.13	−.11	−.04	.17	−.03	.09	.15	−.08	—	
13. Sex (0 = male, 1 = female)	.55	.50	.10	.08	−.08	.11	.14	−.01	−.27**	−.07	.04	−.16	−.18	−.06	—

Note. *N* = 89. Alpha reliabilities are shown in parentheses along the diagonal.
* $p < .05$. ** $p < .01$.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 6
Hierarchical Linear Regression Results (Study 3)

	Task performance			Interpersonal OCB		
Variable	Model 1	Model 2	RW	Model 3	Model 4	RW
Sex	.13	.16	5.0%	.08	.10	3.5%
Number of languages spoken	−.03	−.04	0.3%	−.06	−.07	1.4%
International experience	.09	.13	2.2%	.06	.08	1.4%
Work experience	.15	.15	4.5%	.02	.01	0.6%
Extraversion	−.00	−.03	0.3%	−.05	−.07	0.8%
Agreeableness	−.02	−.02	0.8%	.16	.16	13.0%
Conscientiousness	.24*	.22*	15.7%	.10	.08	5.6%
Emotional stability	.04	.03	0.5%	.11	.10	2.3%
Openness to experience	−.04	−.06	0.3%	−.12	−.13	5.1%
Cognitive ability	.37**	.29**	27.4%	.31**	.25*	28.2%
Response judgment		.33**	43.1%		.26*	37.9%
F	1.75 (10,78)	2.77** (11,77)		1.17, (10,78)	1.67, (11,77)	
R^2	.18	.28		.13	.19	
ΔR^2		.10**			.06*	
adjusted R^2	.08	.18		.02	.08	

Note. $N = 89$. Table reports standardized beta coefficients. RW = Relative weights (%) of R^2.
* $p < .05$. ** $p < .01$.

generalizability of the findings. We also expanded the controls to rule out alternative explanations.

Study 4

Method

Participants. One hundred eighty-eight ($n = 188$) working adults from a master of business administration (MBA) course on international business with an overseas consulting assignment, offered by a large business school in Singapore, participated in the study. Mean age was 33 years ($SD = 5.36$ years), and 38% were female. Participants were diverse, representing 26 countries across five continents. On average, each had 9.2 years ($SD = 5.27$) of previous work experience and had traveled to 11 countries ($SD = 7.86$); 95% spoke at least two languages.

Procedures. Participants formed their own culturally diverse consulting teams (average team size of four; average index of nationality heterogeneity = .81, $SD = 0.21$). Teams worked on an intensive 3-month consulting project where they negotiated access to an organization outside of Singapore and completed a project on specific intercultural management challenges in the target organization. Teams consulted with a wide range of industries (e.g., manufacturing, retail, information, finance, insurance, service, professional service, etc.). They also researched their organization and conducted interviews through on-site visits with executives, middle managers, and employees to understand the intercultural challenges faced by members of the organization. At the end of the project, teams prepared written reports (average number of pages = 116, $SD = 23.64$) and made a formal presentation to the organization. This included a comprehensive analysis of the intercultural and institutional challenges faced by the organization and recommendations for the organization to consider.

We obtained archival data on cognitive ability (Graduate Management Admissions Test [GMAT] from participants' MBA application records) and collected data from participants at three points in time. At Time 1 (beginning of the course), participants completed online surveys of personality and demographic characteristics. At Time 2 (2 weeks into the course, and at the beginning of their project), participants completed the intercultural SJT. At Time 3 (at the end of their project), peers rated task performance and interpersonal OCB of team members using the same round-robin design as Studies 2 and 3.

Measures. Unless otherwise noted, we measured all variables using a 7-point Likert-type scale (1 = *strongly disagree*, 7 = *strongly agree*).

Criterion measures: Task performance and interpersonal OCB. Team members assessed task performance ($r_{\mathrm{WG(J)}} = .84$; $\alpha = .91$) and interpersonal OCB ($r_{\mathrm{WG(J)}} = .76$; $\alpha = .90$) with the same items as Studies 2 and 3. Group variances were again not significant for either task performance (group variance = .04, $Z = .76$, *ns*) or interpersonal OCB (group variance = .05, $Z = .86$, *ns*), supporting the removal of group differences for the estimation of target scores. Target scores indicated agreement between peer ratings and showed statistically significant amounts of target variance for task performance (target variance = .38, $Z = 5.83$, $p < .01$) and interpersonal OCB (target variance = .26, $Z = 5.23$, $p < .01$). Confirmatory factor analyses of the hypothesized two-factor model (task performance and interpersonal OCB) showed excellent fit: $\chi^2[8df] = 10.16$, *ns*; $\chi^2/df = 1.27$, $IFI = .99$, $RMSEA = .04$. All factor loadings were statistically significant (.88 – .99, $p < .01$), and the two-factor model showed significantly better fit than a single-factor model ($\Delta\chi^2[1df] = 111.00$, $p < .001$).

Predictor measures: Situational judgment and response judgment. We employed the same four research assistants as in Study 2 for scoring situational and response judgments. Interrater agreement exceeded .70 (situational judgment: ICC2.1 = .92; response judgment: ICC2.1 = .80). We averaged ratings of situational judgment ($\alpha = .83$) and response judgment ($\alpha = .75$) across both raters.

Control measures. We controlled for cognitive ability (GMAT), Big Five ($\alpha = .74$–.89), previous work experience, international experience, number of languages spoken, and gender.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

We also included additional controls in Study 4 to rule out alternative explanations based on individual characteristics. Given that our operationalization of situational judgment parallels the broader trait of empathy, defined as the tendency to perceive how the world appears to others and to feel compassion for them, we controlled for two types of empathy with four items each adapted from Davis (1983). An example item for intercultural cognitive empathy is "I try to understand people from other cultures better by imagining how things look from their perspective" ($\alpha = .80$), and an item for intercultural affective empathy is "I'm often quite touched by things that I see happen in intercultural interactions" ($\alpha = .72$). Self-efficacy, defined as perceived capability to enact a specific behavior (Bandura, 1997), is an important predictor of performance (Stajkovic & Luthans, 1998). Accordingly, we also controlled for intercultural self-efficacy (e.g., "I know how to put people from different cultures at ease in intercultural situations"; $\alpha = .90$) with six items adapted from Van Dyne et al. (2012).

Table 7 summarizes results of confirmatory factor analyses and shows that the hypothesized five-factor measurement model (situational judgment and response judgment with correlated uniqueness factors between indicators from the same SJT items; cognitive and affective empathy; and intercultural self-efficacy) had good fit to the data: $\chi^2[333df] = 553.41$, $p < .01$; $\chi^2/df = 1.66$, $IFI = .94$, $RMSEA = .06$. All factor loadings were statistically significant (.45 – .80, $p < .01$).

Comparison with alternative models demonstrated that the hypothesized five-factor model was a significantly better fit than (a) a five-factor model that did not allow for correlated uniqueness factors (Model 2: $\Delta\chi^2[7df] = 60.83$, $p < .001$); (b) a four-factor model with correlated uniqueness factors that combined both situational judgment and response judgment (Model 3: $\Delta\chi^2[1df] = 23.74$, $p < .001$); (c) a four-factor model without correlated uniqueness factors that combined both situational judgment and response judgment (Model 4: $\Delta\chi^2[8\text{df}] = 85.62$, $p < .001$); (d) a two-factor model combining situational judgments and response judgments versus all self-reported measures (Model 5: $\Delta\chi^2[4\text{df}] = 54.22$, $p < .001$); and (e) a single-factor model (Model 6: $\Delta\chi^2[11df] = 221.65$, $p < .001$). Together, these results support the discriminant validity of all five constructs.

Results and Discussion

Descriptive statistics. We report descriptive statistics, internal consistency reliability (alpha) coefficients, and correlations in Table 8. Situational judgment was positively related to task performance ($r = .41$, $p < .01$) and interpersonal OCB ($r = .37$, $p < .01$). Response judgment was positively related to task performance ($r = .37$, $p < .01$) and interpersonal OCB ($r = .22$, $p < .01$).

Hypotheses tests. We proposed that situational judgment would predict task performance (Hypothesis 1) and interpersonal OCB (Hypothesis 2), over and above response judgment. Results support Hypothesis 1. Situational judgment predicted task performance ($\beta = .25$, $p < .01$; Model 3, Table 9) and explained an additional 4% of variance in task performance ($p < .01$) over and above response judgment and the controls. Relative weights analysis shows situational judgment accounted for 21.0% (95% CI = 6.8%, 31.0%; $p < .05$) and response judgment accounted for 15.3% (95% CI = 6.3%, 26.8%; $p < .05$) of explained variance in task performance. Relative weights for situational and response judgments were not significantly different from each other (95% CI = −17.8%, 10.7%; *ns*). This replicates results of Study 2 and suggests that both situational judgment and response judgment are not significantly different in their importance as predictors of task performance. As in Study 2, results without control variables replicated these findings.

Our results also support Hypothesis 2. Situational judgment predicted interpersonal OCB ($\beta = .32$, $p < .01$; Model 6). Situational judgment explained an additional 6% of variance in interpersonal OCB ($p < .01$) over and above response judgment and controls. Relative weights analysis shows situational judgment accounted for 32.6% (95% CI = 15.4%, 55.0%; $p < .05$) of explained variance in interpersonal OCB. By contrast, response judgment accounted for 7.0% (95% CI = 1.8%, 22.1%; *ns*). This difference in relative weights between situational and response judgment is statistically significant (95% CI = −50.0%, −6.8%; $p < .05$). Like Study 2, this suggests that situational judgment is a more important predictor of interpersonal OCB than response

Table 7
CFA Comparisons of Alternative Nested Model Fit (Study 4)

Model	Description	χ^2	*df*	χ^2/df	IFI	RMSEA	$\Delta\chi^2$	Δdf
1	Hypothesized five-factor model with correlated uniqueness factors for the same SJT items	553.41^{**}	333	1.66	.94	.059		
	Alternate nested models compared to the hypothesized five-factor model							
2	Five-factor model without correlated uniqueness factors for the same SJT items	614.24^{**}	340	1.81	.92	.066	60.83^{**}	7
3	Single-SJT-factor model with correlated uniqueness factors for the same SJT items	577.15^{**}	334	1.73	.92	.062	23.74^{**}	1
4	Single-SJT-factor model without correlated uniqueness factors for the same SJT items	639.03^{**}	341	1.87	.91	.068	85.62^{**}	8
5	Two-factor model combining both SJT measures and combining all self-reported measures	607.63^{**}	337	1.80	.90	.066	54.22^{**}	4
6	Single-factor model	775.06^{**}	343	2.26	.85	.082	221.65^{**}	11

Note. $N = 188$. IFI = incremental fit index; RMSEA = root-mean-square error of approximation. The hypothesized model includes situational judgment, response judgment, intercultural self-efficacy, intercultural cognitive empathy, and intercultural affective empathy.
** $p < .01$.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 8
Means, Standard Deviations, and Correlations (Study 4)

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Criterion measures																			
1. Task Performance	.01	.60	(.91)																
2. Interpersonal OCB	.02	.62	.59**	(.90)															
Predictor measures																			
3. Situational Judgment	2.91	.50	.41**	.37**	(.83)														
4. Response Judgment	2.61	.50	.37**	.22**	.49**	(.75)													
Control measures																			
5. Intercultural Self-Efficacy	5.40	.82	.23**	.20**	.12	.26**	(.90)												
6. Intercultural Cognitive Empathy	5.49	.75	.19*	.21**	.25**	.14	.38**	(.80)											
7. Intercultural Affective Empathy	5.63	.75	.12	.20**	−.10	−.07	.32**	.34**	(.72)										
8. Extraversion	3.20	.73	−.16	−.04	.03	.07	.20**	−.02	.10	(.89)									
9. Agreeableness	3.91	.40	.19**	.24**	.10	−.01	.06	.07	.38**	.31**	(.74)								
10. Conscientiousness	3.51	.45	.28**	.20**	.02	.16*	.12	.05	.14	.08	.22**	(.74)							
11. Emotional Stability	3.29	.66	.05	.08	.04	.00	.12	.16*	.01	.10	.15*	.08	(.84)						
12. Openness to Experience	3.62	.44	.11	.12	.24**	.17*	.12	.10	.07	.32**	.23**	.10	.18*	(.77)					
13. Cognitive Ability	643.69	39.47	.29**	.04	.28**	.31**	.08	.06	−.11	−.12	−.01	.02	−.14	.16*	—				
14. International Experience	10.56	7.86	.12	.20**	.20*	.10	.11	.06	−.04	.14	.04	.06	.15*	.15*	.01	—			
15. Work Experience (in years)	9.16	5.27	.02	.08	−.05	−.12	.07	.13	.12	−.08	.02	.17*	.21**	.02	−.09	.29**	—		
16. Number of Languages Spoken	2.60	.93	.19*	.14	.12	.10	.09	.08	.08	.10	.09	.02	.02	.21**	.12	.06	−.09	—	
17. Sex (0 = male, 1 = female)	.38	.49	.14	.06	.10	.04	−.10	.04	−.01	−.11	.15*	.02	−.13	−.10	−.11	−.15*	−.12	.02	—

Note. $N = 188$. Alpha reliabilities are shown in parentheses along the diagonal.
* $p < .05$ ** $p < .01$.

Table 9
Hierarchical Linear Regression Results and Relative Weights Analysis (Study 4)

	Task performance				Interpersonal OCB			
Variable	Model 1	Model 2	Model 3	RW	Model 4	Model 5	Model 6	RW
Sex	.14*	.12	.10	3.6%	.04	.02	−.00	0.7%
Number of Languages Spoken	.13	.12	.12	5.0%	.09	.09	.08	3.7%
Work Experience	−.06	−.03	−.02	0.3%	−.04	−.01	−.00	0.8%
International Experience	.13	.11	.07	2.0%	.20**	.18*	.13	8.6%
Extraversion	−.27**	−.28**	−.27**	11.0%	−.20*	−.20*	−.19*	4.8%
Agreeableness	.16*	.18*	.14	5.8%	.21*	.22**	.17*	12.3%
Conscientiousness	.23**	.19**	.21**	13.1%	.13	.10	.13	8.1%
Emotional Stability	.04	.04	.04	0.5%	−.00	−.01	.00	0.5%
Openness to Experience	.04	.01	−.02	1.0%	.05	.04	−.00	1.4%
Cognitive Ability	.23**	.17*	.14*	10.3%	−.02	−.06	−.09	1.0%
Intercultural Cognitive Empathy	.05	.03	−.03	1.8%	.10	.09	.01	4.3%
Intercultural Affective Empathy	.00	.04	.08	2.4%	.05	.07	.12	8.1%
Intercultural Self-Efficacy	.19**	.14	.15*	6.6%	.13	.09	.10	5.9%
Response Judgment		.25**	.15*	15.3%		.17*	.05	7.0%
Situational Judgment			.25**	21.0%			.32**	32.6%
F	6.38** (13,174)	7.28** (14,173)	7.92** (15,172)		3.28** (13,174)	3.48** (14,173)	4.55** (15,172)	
R^2	.32	.37	.41		.20	.22	.28	
ΔR^2		.05**	.04**			.02*	.06**	
adjusted R^2	.27	.32	.36		.14	.16	.22	

Note. $N = 188$. Table reports standardized beta coefficients. RW = Relative weights (%) of R^2.
* $p < .05$. ** $p < .01$.

judgment. As in Study 2, results without control variables replicated these findings.

In sum, results from Study 4 replicated the key findings from Study 2 with working adults and a broader range of control variables. Thus, Study 4 strengthens the generalizability of our results and supports further our assumption that adding situational judgment to SJTs provides valuable information beyond that provided by response judgment.

General Discussion

In this research, we expanded the SJT paradigm and made three key contributions. First, following the suggestions of Ployhart (2006), we used verbal protocol analysis to open the black box of SJTs and examined the types of situational judgments made by SJT respondents. Our findings show that understanding the intentions, emotions, and thoughts of the parties in the situation were the dominant types of situational judgments made by our participants. This is important because it is the first study to illuminate how people perceive and interpret situations presented in SJTs. Knowing how people perceive and interpret the situation is important because it deepens our understanding of why some people make better or worse response judgments than others.

Second, results demonstrated that situational judgment incrementally predicted both task performance and interpersonal OCB—over and above response judgment. This result holds even after ruling out a possible alternative explanation in Study 3—that is, that assessing situational judgment might affect the predictive validity of response judgment due to cognitive fatigue, learning, context, or accessibility effects. Taken together, these findings attest to the usefulness of complementing response judgment with situational judgment in SJTs.

Third, both situational judgment and response judgment significantly predicted task performance. However, in the case of interpersonal OCB, situational judgment emerged as the only significant predictor when controlling for both situational and response judgment.

Theoretical Implications and Future Research Directions

In a recent review of the SJT literature, Ployhart and MacKenzie (2011) noted that the role of judgment in SJT research and practice is a neglected issue. To address this gap, we took a novel approach to SJTs by expanding on the type of judgments measured by SJTs from a focus on response judgments, to include a focus on situational judgments. Our primary theoretical contribution lies in putting judging situations back into SJT theorizing. Reconnecting SJT theorizing with situational judgment grounds SJTs more firmly in an interactionist paradigm (Campion & Ployhart, 2013), which emphasizes the importance of both situational judgments and response judgments. In so doing, we show how we can expand the SJT paradigm and provide a foundation for future SJT research to advance our understanding of judgment processes in SJTs.

To deepen our understanding of situational and response judgments in SJTs, future research could examine cognitive mechanisms that help explain the processes respondents use to arrive at situational and response judgments. For instance, Ployhart (2006) suggested that memory retrieval is one psychological mechanism underlying judgments in SJTs. Consistent with this notion, our verbal protocol results provide anecdotal evidence that people draw on their previous experiences with similar situations when making situational and response judgments. Thus, future studies could compare situational and response judgments of people with more or less experience relative to the situations presented by a SJT.

Given the promising role of situational judgments in predicting performance, future research could also examine factors that affect

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

situational judgments. For instance, attribution theory suggests that situational cues about internal versus external causes of behavior, controllability of behavior, and the stability of behavior influence situational judgments (Weiner, 1995). Future SJT research could systematically vary the amount and presence of such situational cues and examine their influence on situational judgments, as well as response judgments. Further, research could also examine interactions between situational features and personality traits in affecting situational judgment. For example, individuals' need for closure, which refers to the epistemic desire to seize immediately on a firm answer to an ambiguous situation and to subsequently neglect consideration of alternative answers (Kruglanski & Webster, 1996), may interact with situational cues to affect situational judgment.

By identifying situational and response judgment as two key constructs measured by our SJT, this study shed greater light on why SJTs predict task performance and interpersonal OCB. This is important, as Christian et al. (2010) noted that "identifying the constructs measured by selection tests such as SJTs is important for theory testing and understanding why a given test is or is not related to the criterion of interest" (p. 85). Our findings suggest that SJT scores might relate to task performance and interpersonal OCB based on different types of judgments assessed in the SJTs. For task performance, we found that both situational judgment and response judgment contributed unique variance. For interpersonal OCB, however, only situational judgment contributed unique variance while response judgment did not. By explicitly measuring situational judgment in addition to response judgment, we found empirical evidence that supports Christian et al.'s speculation that SJTs relate to interpersonal OCB because SJTs assess to some extent the "ability to perceive and interpret social dynamics in such a way that facilitates judgments regarding the timing and appropriateness of contextual behaviors" (p. 92). Future research could further open the black box of the effects of situational judgment on task performance and interpersonal OCB. For task performance, future research could draw on our arguments that situational judgment facilitates understanding others' role expectations and examine role expectations as a possible mediator of the relationship between situational judgment and task performance. For interpersonal OCB, future research could draw on Christian et al.'s arguments and examine appropriateness of timing and nature of help offered as potential mediators of the relationship between situational judgment and interpersonal OCB.

Finally, our research should also have implications for performance outcomes across a wide range of jobs because situational judgments about intentions, emotions, and thoughts are crucial to performance in many jobs that involve interpersonal relationships. For example, physicians need to relate to their patients (Silvester, Patterson, Koczwara, & Ferguson, 2007); or service providers need to understand the concerns of their clients (Parker & Axtell, 2001). Thus, future research could examine situational judgment about others' intentions, emotions, and thoughts as predictors of task performance and interpersonal OCB in a wide variety of jobs and domains.

Implications for Selected-Response SJTs

We have discussed the implications of our findings on situational judgments and response judgments in expanding future SJT research. At the same time, our results should be interpreted in the context of our constructed-response (i.e., open-ended) methodology, which differs from the "selected-response" methodology (i.e., close-ended questions) commonly adopted in the extant SJT literature (Motowidlo, Dunnette, & Carter, 1990). Our choice was predicated on the notion that a constructed-response format might provide greater response fidelity (i.e., the extent to which the response format corresponds to similar real-life ways of responding; Sackett, 1987; Weekley, Ployhart, & Holtz, 2006) than the selected-response format. As Ryan and Greguras (1998) noted, "life is not multiple choice" (p. 183). Further, a constructed-response format avoids cuing respondents about the correct solution (Thornton & Rupp, 2006), which we felt was particularly important for an intercultural SJT because intercultural interactions are prone to misjudgments (Earley & Ang, 2003).

Thus, an important caveat of this study is that our findings on situational judgments and response judgments are based on "constructed-response" format, which may not necessarily generalize to SJT studies using the selected-response format. As such, we recommend three key areas for future research to better understand selected- versus constructed-response SJTs.

First, future research could test the generalizability of our findings to selected-response SJTs. This requires future studies to assess whether situational judgments predict task performance and OCB over and above response judgments, using a selected-response format to measure both the situational and response judgments.

Second, future research could explore the differential effects of construct (situation judgment vs. response judgment) versus method (selected response vs. constructed response) by employing a 2 (construct) × 2 (method) research design, with SJT item stems held constant. Such a design allows us to determine, for instance, whether the options provided in the selected-response SJTs might be incomplete compared to judgments obtained in the constructed-response SJTs. These findings will have implications on the predictive validity of the two types of SJTs. Future research could also use verbal protocols to compare the cognitive processes underlying the generation of situational and response judgments across the four cells in the 2 × 2 research design. This will deepen our understanding of the effects of different response formats on the cognitive processes in SJTs.

Third, future research could examine the utility of constructed- versus selected-response SJTs, to better inform SJT practitioners on the pros and cons of the different types of SJTs. We note that while the constructed-response format offers greater response fidelity than selected-response formats, it is more time and resource consuming. For example, our respondents took an average of 3.7 min ($SD = 42$ s) to complete each constructed-response question, and our raters took on average 1 min to score each response. Selected-response SJTs on the other hand, are "easier to score and implement in large-scale testing programs, making them attractive options for early stages of recruitment and selection" (Weekley & Ployhart, 2006, p. 5). They also offer the possibility of immediate feedback. Therefore, we recommend that future studies assess the incremental validity of constructed situational judgment over and above (a) selected response judgment, (b) selected situational judgment, and (c) selected response and situational judgment. Such research may further our understanding of the benefits provided by greater response fidelity of a constructed-response format

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

relative to the benefit of easier scoring associated with a selected-response format.

Practical Implications

To date, SJTs only assess response judgment. Our finding that situational judgment predicts task performance and interpersonal OCB over and above response judgment reinforces the value of assessing both forms of judgments in SJTs. This finding also has practical implications for other assessment and selection techniques where candidates are required to respond to situations, such as situational interviews (Latham, Saari, Pursell, & Campion, 1980). For instance, managers could ask candidates to provide situational judgments in addition to response judgments during situational interviews to better assess the qualities of the candidates.

In light of the different advantages that selected-and constructed-responses offer, organizations may consider developing SJTs that contain items with selected-response formats and items with constructed-response formats. This may allow organizations to strike a balance between concerns associated with the greater testing and scoring time required for constructed-response SJTs and their benefits in terms of response fidelity. Notably, constructed-response formats need not be limited to the written responses used in our research. For example, a growing body of SJT research demonstrates the feasibility of using webcam based constructed-response formats to enhance the response fidelity of SJTs (e.g., Lievens, De Corte, & Westerveld, in press).

Our study also highlights the potential usefulness of verbal protocols in uncovering situational judgments made by respondents. Thus, verbal protocols could be deployed in other assessment contexts that require situational responses, such as assessment centers. Similarly, verbal protocol analysis could be useful as a prescreening tool in the development of SJT items, especially for text-based SJT items. McDaniel, Psotka, Legree, Yost, and Weekley (2011) noted that text-based SJT items are often ambiguous, requiring respondents to make additional assumptions about the presented situation. As item ambiguity reduces the predictive validity of SJT items (McDaniel et al., 2011), we recommend that test developers use verbal protocol analysis to discover the kinds of assumptions that respondents make, which can be used to identify and improve ambiguous SJT items.

Finally, having an intercultural SJT that predicts performance outcomes in culturally diverse contexts makes a significant practical contribution. As noted by Deardorff (2009), "Intercultural competence is a very complex concept with a variety of components and aspects. One tool or method does not provide a comprehensive measurement of the complexity of this concept" (p. 486). Given that current measures of intercultural competence are dominantly based on self-reported instruments (Leung, Ang, & Tan, 2014), the intercultural SJT provides an alternative performance-based assessment tool that has good predictive validity. Organizations may complement validated report-based measures of intercultural competence (e.g., Ang et al., 2007) with our intercultural SJT to enhance their selection for international assignments (Leung et al., 2014) or global leadership positions (Rockstuhl, Seiler, Ang, Van Dyne, & Annen, 2011).

Strengths and Limitations

A methodological strength of our study is the use of verbal protocols to clarify the nature of situational judgments made by respondents. To date, researchers have often attempted to understand judgments in SJTs based on relationships of SJT performance with established predictors such as cognitive ability (Weekley & Jones, 1997), personality (McDaniel et al., 2007), job experience (Motowidlo & Beier, 2010), or job knowledge (Clevenger, Pereira, Wiechmann, Schmitt, & Schmidt-Harvey, 2001). Verbal protocols complement such efforts because they offer a unique opportunity for directly examining the judgments made by SJT respondents. As our results show, such an approach holds great potential to deepen our understanding of the constructs assessed by SJTs. Examining respondents' situational judgments should allow SJT researchers to uncover alternative question prompts beyond asking respondents what they would do.

Our lagged, multisource design is another methodological strength that responds to calls in the SJT literature for greater use of predictive validity designs instead of concurrent validity designs (Whetzel & McDaniel, 2009). Additionally, replicating our results across two lagged, multiple source studies strengthens the confidence in the generalizability of our findings.

We note a few limitations in our study. First, our focus here is on interpersonal intercultural SJTs. Hence, our findings that situational judgments mainly involve judgments of someone's intentions, emotions, and thoughts may not generalize to other types of SJTs. We encourage future studies to extend this research to other types of SJTs, such as SJTs that assess knowledge and skills and basic personality tendencies (Christian et al., 2010). We expect that SJTs that are less interpersonal and more focused on a task may require different situational judgments.

Second, our measure of situational judgment in Study 2 and Study 4 combined judgments of thoughts, emotions, and intentions, in order to reduce the administration time for participants. However, combining these different judgments precludes us from examining which judgment (i.e., thoughts, emotions, intentions) might be most important when judging a situation. To deepen our understanding of the relative importance of these judgments, we recommend future studies to assess judgments of thoughts, emotions, and intentions separately and to compare their predictive validity.

Third, our studies relied on undergraduate and MBA students, which may evoke questions regarding the external validity of the findings. However, we note that despite being students, our participants worked in teams similar to teams in real-world contexts. For instance, participants had to work interdependently within their team on a high-stakes task under time pressure. In Study 4, teams also had to present their project to an external client. Nonetheless, future research could replicate our findings with managerial samples to strengthen the external validity of our results.

Finally, results from Study 3 strengthen the internal validity of our findings by addressing questions about inadvertent effects of priming respondents to make situational judgments on the predictive validity of response judgments. At the same time, respondents in Study 2 and Study 4 consistently provided situational judgments

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

before making response judgments. The order in which respondents provide situational- and response judgments might therefore affect the incremental validity of situational judgment over and above response judgment. Thus, future research could further strengthen the internal validity of our results by replicating our findings while asking respondents to generate situational judgments after providing response judgments.

Conclusion

SJT scholars have repeatedly called for research to open the black box of situational judgment in SJTs (Ployhart, 2006; Schmitt & Chan, 2006; Whetzel & McDaniel, 2009). Our study responds to these calls and reinvigorates SJT research by highlighting the importance of putting situational judgments back into SJTs. Specifically, results of our verbal protocol analysis of SJT responses identified the dominant types of situational judgments made. More important, results of two time-lagged, multiple-source studies demonstrate the value of asking respondents to make both situational judgments and response judgments. Results consistently show that situational judgment predicts task performance and interpersonal OCB over and above response judgment and other established predictors. Overall, our results provide timely insights to situational judgment in SJTs and suggest promising benefits—both theoretical and practical—for future research on SJTs.

References

Ang, S., Rockstuhl, T., & Ng, K. Y. (2014). *Performance-based cultural intelligence (CQ): Development and validation of an intercultural situational judgment test (iSJT)*. Singapore: Nanyang Technological University, Center for Leadership and Cultural Intelligence.

Ang, S., & Van Dyne, L. (2008). Conceptualization of cultural intelligence: Definition, distinctiveness, and nomological network. In S. Ang & L. Van Dyne (Eds.), *Handbook of cultural intelligence: Theory, measurement, and applications* (pp. 3–15). Armonk, NY: Sharpe.

Ang, S., Van Dyne, L., Koh, C., Ng, K. Y., Templer, K. J., Tay, C., & Chandrasekar, N. A. (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. doi:10.1111/j.1740-8784.2007.00082.x

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.

Blau, P. M. (1977). *Inequality and heterogeneity*. New York, NY: Free Press.

Byrne, 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. doi:10.1037/0033-2909.105.3.456

Byham, W. C. (1977). Assessor selection and training. In J. L. Moses & W. C. Byham (Eds.), *Applying the assessment center method* (pp. 89–126). New York, NY: Pergamon Press. doi:10.1016/B978-0-08-019581-0.50011-6

Campion, M. C., & Ployhart, R. E. (2013). Assessing personality with situational judgment measures: Interactionist psychology operationalized. In N. D. Christiansen & R. P. Tett (Eds.), *Handbook of personality at work* (pp. 439–456). New York, NY: Routledge.

Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology, 82,* 143–159. doi:10.1037/0021-9010.82.1.143

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

Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N., & Schmidt-Harvey, V. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology, 86,* 410–417. doi:10.1037/0021-9010.86.3.410

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44,* 113–126. doi:10.1037/0022-3514.44.1.113

Deardorff, D. K. (2009). Implementing intercultural competence assessment. In D. K. Deardorff (Ed.), *The Sage handbook of intercultural competence* (pp. 477–491). Thousand Oaks, CA: Sage.

Earley, P. C., & Ang, S. (2003). *Cultural intelligence: Individual interactions across cultures*. Palo Alto, CA: Stanford University Press.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin, 128,* 203–235. doi:10.1037/0033-2909.128.2.203

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Farh, J. L., Zhong, C. B., & Organ, D. W. (2004). Organizational citizenship behavior in the People's Republic of China. *Organization Science, 15,* 241–253. doi:10.1287/orsc.1030.0051

Feldman, J. M., & Lynch, J. G. Jr. (1988). Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior. *Journal of Applied Psychology, 73,* 421–435. doi:10.1037/0021-9010.73.3.421

Gelfand, M. J., Erez, M., & Aycan, Z. (2007). Cross-cultural organizational behavior. *Annual Review of Psychology, 58,* 479–514. doi:10.1146/annurev.psych.58.110405.085559

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. doi:10.2307/2667088

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, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 7–28). Tilburg, the Netherlands: Tilburg University Press.

Heider, F. (1958). *The psychology of interpersonal relations*. New York, NY: Wiley. doi:10.1037/10628-000

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.

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

Johnson, J. W., & LeBreton, J. M. (2004). History and use of relative importance indices in organizational research. *Organizational Research Methods, 7,* 238–257. doi:10.1177/1094428104266510

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: User's reference guide*. Chicago, IL: Scientific Software International.

Kenny, D. A. (1995). *SOREMO Version 2: A FORTRAN program for the analysis of round-robin data structures*. Unpublished manuscript, University of Connecticut.

Kenny, D. A., & La Voie, L. (1984). The social relations model. *Advances in Experimental Social Psychology, 18,* 141–182. doi:10.1016/S0065-2601(08)60144-6

Kruglanski, A. W., & Webster, D. M. (1996). Motivated closing of the mind: "Seizing" and "freezing". *Psychological Review, 103,* 263–283. doi:10.1037/0033-295X.103.2.263

Lance, C. E., Woehr, D. J., & Meade, A. W. (2007). A Monte Carlo investigation of assessment center construct validity models. *Organiza-*

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

tional Research Methods, 10, 430–448. doi:10.1177/1094428106289395

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159–174. doi:10.2307/2529310

Latham, G. P., Saari, L. M., Pursell, E. D., & Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology, 65,* 422–427. doi:10.1037/0021-9010.65.4.422

LeBreton, J. M., & Senter, J. L. (2007). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods, 11,* 815–852. doi:10.1177/1094428106296642

Leung, K., Ang, S., & Tan, M. L. (2014). Intercultural competence. *Annual Review of Organizational Psychology and Organizational Behavior, 1,* 489–519. doi:10.1146/annurev-orgpsych-031413-091229

Lievens, F. (2006). International situational judgment tests. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement and application* (pp. 279–300). Mahwah, NJ: Erlbaum.

Lievens, F., De Corte, W., & Westerveld, L. (2012). Understanding the building blocks of selection procedures: Effects of response fidelity on performance and validity. *Journal of Management*. Advance online publication. doi:10.1177/0149206312463941

Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: A review of recent research. *Personnel Review, 37,* 426–441. doi:10.1108/00483480810877598

Magnusson, D., & Ekehammar, B. (1975). Perceptions of and reactions to stressful situations. *Journal of Personality and Social Psychology, 31,* 1147–1154. doi:10.1037/h0077032

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. doi:10.1111/j.1744-6570.2007.00065.x

McDaniel, M. A., Psotka, J., Legree, P. J., Yost, A. P., & Weekley, J. A. (2011). Toward an understanding of situational judgment item validity and group differences. *Journal of Applied Psychology, 96,* 327–336. doi:10.1037/a0021983

Molinsky, A. L. (2013). The psychological processes of cultural retooling. *Academy of Management Journal, 56,* 683–710. doi:10.5465/amj.2010.0492

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

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

Motowidlo, S. J., Hooper, A. C., & Jackson, H. L. (2006). A theoretical basis for situational judgment tests. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and practice* (pp. 57–81). Mahwah, NJ: Erlbaum.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Parker, S. M., & Axtell, C. M. (2001). Seeing another viewpoint: Antecedents and outcomes of employee perspective taking. *Academy of Management Journal, 44,* 1085–1100. doi:10.2307/3069390

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

Ployhart, R. E., & MacKenzie, W. I. Jr. (2011). Situational judgment tests: A critical review and agenda for future research. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology* (Vol. 2, pp. 237–252.). Washington, DC: American Psychological Association.

Pulakos, E. D. (1984). A comparison of rater training programs: Error training and accuracy training. *Journal of Applied Psychology, 69,* 581–588. doi:10.1037/0021-9010.69.4.581

Rockstuhl, T., Seiler, S., Ang, S., Van Dyne, L., & Annen, H. (2011). Beyond EQ and IQ: The role of cultural intelligence in cross-border leadership effectiveness in a globalized world. *Journal of Social Issues, 67,* 825–840. doi:10.1111/j.1540-4560.2011.01730.x

Ryan, A. M., & Greguras, G. J. (1998). Life is not multiple choice: Reactions to the alternatives. In M. D. Hakel (Ed.), *Beyond multiple choice: Evaluating alternatives to traditional testing for selection* (pp. 183–202). Mahwah, NJ: Erlbaum.

Sackett, P. R. (1987). Assessment centers and content validity: Some neglected issues. *Personnel Psychology, 40,* 13–25. doi:10.1111/j.1744-6570.1987.tb02374.x

Schmitt, N., & Chan, D. (2006). Situational judgment tests: Method or construct? In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement and application* (pp. 135–155). Mahwah, NJ: Erlbaum.

Shore, L. M., Randel, A. E., Chung, B. G., Dean, M. A., Ehrhart, K. H., & Singh, G. (2011). Inclusion and diversity in work groups: A review and model for future research. *Journal of Management, 37,* 1262–1289. doi:10.1177/0149206310385943

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420–428. doi:10.1037/0033-2909.86.2.420

Silvester, J., Patterson, F., Koczwara, A., & Ferguson, E. (2007). "Trust me . . .": Psychological and behavioral predictors of perceived physician empathy. *Journal of Applied Psychology, 92,* 519–527. doi:10.1037/0021-9010.92.2.519

Smith, P. C., & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors to rating scales. *Journal of Applied Psychology, 47,* 149–155. doi:10.1037/h0047060

Spencer-Oatey, H. (2008). Face, (im)politeness, and rapport. In H. Spencer-Oatey (Ed.), *Culturally speaking: Culture, communication, and politeness theory* (2nd ed., pp. 11–47). London, England: Continuum.

Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin, 124,* 240–261. doi:10.1037/0033-2909.124.2.240

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. doi:10.1111/1464-0597.00139

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.

Thomas, K. W. (1976). Conflict and conflict management. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 889–935). Chicago, IL: Rand McNally.

Thornton, G. C., III, & Rupp, D. E. (2006). *Assessment centers in human resources management*. Mahwah, NJ: Erlbaum.

Triandis, H. C. (2006). Cultural intelligence in organizations. *Group & Organization Management, 31,* 20–26. doi:10.1177/1059601105275253

Van Dyne, L., Ang, S., Ng, K. Y., Rockstuhl, T., Tan, M. L., & Koh, C. (2012). Sub-dimensions of the four factor model of cultural intelligence: Expanding the conceptualization and measurement of cultural intelligence. *Social and Personality Psychology Compass, 6,* 295–313. doi:10.1111/j.1751-9004.2012.00429.x

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. doi:10.2307/256902

Weber, E. U., & Johnson, E. J. (2009). Mindful judgment and decision making. *Annual Review of Psychology, 60,* 53–85. doi:10.1146/annurev.psych.60.110707.163633

Weekley, J. A., & Jones, C. (1997). Video-based situational testing. *Personnel Psychology, 50,* 25–49. doi:10.1111/j.1744-6570.1997.tb00899.x

Weekley, J. A., & Ployhart, R. E. (2006). An introduction to situational judgment testing. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and practice* (pp. 1–10). Mahwah, NJ: Erlbaum.

Weekley, J. A., Ployhart, R. E., & Holtz, B. C. (2006). On the development of situational judgment tests: Issues in item development, scaling, and scoring. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and practice* (pp. 157–182). Mahwah, NJ: Erlbaum.

Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct*. New York, NY: Guilford Press.

Whetzel, D. L., & McDaniel, M. A. (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review, 19,* 188–202. doi:10.1016/j.hrmr.2009.03.007

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. doi:10.1177/014920639101700305

Wonderlic, E. F. (1999). *Wonderlic Personnel Test user's manual*. Libertyville, IL: Wonderlic.

Appendix

Development of the Intercultural SJT

Ang, Rockstuhl, and Ng (2014) presented a detailed report of the development and validation of the intercultural SJT used in this research. The intercultural SJT depicts intercultural interpersonal interactions at work using multimedia vignettes. We chose to develop a multimedia rather than a text-based SJT because multimedia SJTs are of higher fidelity and greater validity (Chan & Schmitt, 1997; Christian et al., 2010).

Script Development

We followed Weekley et al.'s (2006) recommendations for scripting multimedia SJT items. First, we constructed a taxonomy of the situational domain of an intercultural SJT. We then collected critical incidents from interviews with subject matter experts (SMEs; executives in international assignments, experienced cross-cultural researchers and trainers) and from extensive reviews of the literature. We identified prototypical incidents for scripting. Specifically, we focused on interactions between individuals from two different cultural backgrounds including North America, South America, Europe, Asia, and the Middle East.

A professional scriptwriter drafted and revised scripts using input from SMEs (executives from the countries depicted in the scripts and experienced cross-cultural researchers). We produced vignettes in authentic, work-related settings using professional actors from countries and ethnicities depicted in the scripts. We deployed intercultural experts during the film production to assure cultural fidelity of the multimedia vignettes. As a final manipulation check, subject matter experts not involved in the SJT development independently mapped each multimedia vignette to the underlying situational taxonomy. The Cohen's kappa agreement between expert ratings and the intended situational domain averaged .92 (range from .83 to 1).

Received October 5, 2013
Revision received July 22, 2014
Accepted August 12, 2014 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.