The Criterion-Related Validities and Perceived Fairness of the Situational Interview and the Situational Judgment Test in an Iranian Organisation

Sara Banki and Gary P. Latham*
Rotman School of Management, University of Toronto, Canada

The criterion-related validity coefficients of a situational interview (SI) and a situational judgment test (SJT) were investigated in a sales department of a government-owned automobile company in Iran. Both the SI and SJT had concurrent validity with job performance \((n = 101, r = .28, r = .23, p < .05\), respectively). Only the SI, however, had incremental validity over and above the SJT in predicting job performance. Furthermore, the SI fully mediated the relationship between SJT and job performance. The two methods did not differ significantly on employee motivation to perform well, perceived fairness, or test anxiety. Nevertheless, employees recommended the use of the SI over the SJT.

Les coefficients de validité critérielle d’un entretien situationnel (SI) et d’un test de jugement situationnel (SJT) ont été calculés dans le service commercial d’une entreprise publique iranienne du secteur automobile. SI et SJT présentent tous les deux une validité concurrente avec la performance professionnelle (pour \(N = 101, r = .28\) et .23 respectivement; sign. à .05). Toutefois, seul le SI bénéficie d’une validité incrémentielle par rapport au SJT dans la prédiction de la performance professionnelle. En outre, le SI contrôle totalement la relation entre le SJT et le critère. Les deux méthodes ne diffèrent pas significativement dans l’évaluation de la motivation au travail, de l’équité perçue ou de l’anxiété. Néanmoins les salariés préféraient le SI au SJT.

INTRODUCTION

Frese and colleagues (e.g. Bledow & Frese, 2009; Fay & Frese, 2001) have used both situational interview (SI) and situational judgment tests (SJT) as measures of a person’s personal initiative. Frese argued that both methods can be used as proximal predictors of actual behavior. Both methods, he said,
present descriptions of work-related situations to respondents, and both ask them to mentally simulate that they are faced with these situations. The situations refer to critical incidents (Flanagan, 1954). Bledow and Frese (2009) argued that “‘in-situ’, situated behavioral preferences shape the development of goals and intentions (Latham & Skarlicki, 1995), their transformation into concrete ideas of how to act in a given situation (Gollwitzer, 1999) and ultimately actual behavior (Latham & Saari, 1984, p. 8)”.

Despite the fact that the criterion-related validity of the SI and SJT with job performance has been studied extensively (e.g. Latham & Sue-Chan, 1999; McDaniel, Hartman, Whetzel, & Grubb, 2007), no previous study has examined the criterion-related validity of the SI relative to the SJT in the West, let alone in a country with a strong Islamic culture. Hence, this was the purpose of the present study.

Dreher (2000) observed that throughout the 20th century, findings from the behavioral sciences have been derived for the most part from Judeo-Christian, Euro-Western countries. Zewail (2004) reported that “70–80 percent of the world’s population largely in developing countries [has] contributed less than 7 percent of . . . [recent] scientific articles” (p. 1). In their review of cross-cultural studies on organisational behavior, Tsui, Nifadkar, and Ou (2007) found that even here, 86 per cent of studies’ first authors were from Western countries. Only one study was conducted by a first author from the Middle East. Hence, Gelfand, Leslie, and Fehr (2008) concluded that to prosper, organisational psychology must adopt a global perspective. Specifically, they argued the necessity for organisational psychology to investigate global voices and culturally adapt practices rather than merely implement Western methodologies in non-Western countries. Global voices are needed, they said, because national culture can influence organisational behavior. Answers to questions that reflect the values and realities of the non-Western world are needed to determine the generalisability of behavioral science findings, as well as the acceptability of the methods to non-Western populations.

The practical significance of the present study, which focused on the selection of employees, is fourfold. First, it adds a non-Western voice to the literature. The first author is an Islamic female. Second, the study on employee selection took place in a country where the values and realities are far different from those of the West, namely, Iran. Hence, the results are of importance not only to Iranian employers and employees, but they are of particular importance to global organisations that wish to establish a business in Iran. To the authors’ knowledge, this is the first employee validation study conducted in this country. This is because unlike the West, highly distinct boundaries among such spheres of life as family, friendship, and religion with organisational behavior do not exist in Iran. In the Middle East, family and friends exert a more powerful influence on employers/employees...
than is the case in the West (Budhwar & Mellahi, 2006). Thus, employee selection in Iran is typically done with little or no thought to objectivity or criterion-related validity. Instead, emphasis is placed on the legitimisation of subjectivity and nepotism. As Blunt and Popoola (1985) noted, “In settings where paid work is scarce, and where there are strong pressures to allocate jobs in a particular fashion, the selection process constitutes a prime means for fulfilling one’s obligations to kin and other personal contacts” (p. 51).

Third, since 1979 there has been ongoing conflict between Iran and the West in general, and the United States in particular. This has led to both a religious and a government ban on anything that is distinctly Western. Men, for example, are frowned upon for having a cleanly shaven face or wearing a necktie.

The extent to which selection methods developed in the West, and unknown in Iran, are acceptable to Iranians, was not known prior to this study. Hence the answer should be useful to both Iranian and global organisations. The two selection methods that were examined were the situational interview (SI) and the situational judgment test (SJT). These two methods were examined because, unlike many selection tests developed in the West, both can be contextualised, which increases face validity. An empirical comparison of the criterion-related validities of these two selection methods, and employee reactions to them, has not taken place in the West, let alone in Iran. This comparison is important because these two methods are closely related to one another. Hence, the fourth reason for conducting this study was to determine whether an SJT is as valid, or significantly more valid than an SI, and to see whether employees’ reactions to it are as positive, if not significantly more positive than reactions to an SI. If the answers are positive, global organisations can hire people in Iran through administration of the SJT on the organisation’s website.

SI VERSUS SJT

Unlike Iran, the most frequently used method for selecting employees in Euro-American countries is the interview (Dipboye, Wooten, & Halverson, 2004). Moreover, it is the selection technique preferred by applicants (Hausknecht, Day, & Thomas, 2004). An interview can be either structured or unstructured. To the extent that the interview is structured, the criterion-related validity coefficient increases (Huffcutt & Arthur, 1994). The corrected validities of structured interviews are comparable to those obtained with tests that assess cognitive ability. This is particularly true when the questions asked in a structured interview are derived from a job analysis (Wiesner & Cronshaw, 1988).

The SI (Latham, Saari, Pursell, & Campion, 1980) is a structured interview whereby all individuals are asked the same questions. The questions are
derived from a job analysis, namely, the critical incident technique, or CIT (Flanagan, 1954). The job analysis makes explicit to the interviewers the job requirements that are critical for defining an ideal applicant. In addition, the CIT facilitates the development of a behavioral scoring guide for assessing whether an interviewee’s response to a question is excellent (5), acceptable (3), or unacceptable (1). Having a scoring guide minimises the influence of an interviewee’s non-job-related behavior on an interviewer’s decision. A scoring guide rather than a scoring key is used because real-world job demands or practical problems often have multiple solutions that go well beyond technical task knowledge (Schmitt & Chan, 2006). As Chan and Schmitt (2005) have argued, high levels of contextual knowledge of what to do and how to do it are required with regard to interpersonal, organisational, and the resource environment in order to become a high performer. This is likely to be true in all societies.

Inherent in each SI question is a dilemma that increases the likelihood that interviewees will state what they believe they would do in a given situation. Through uniform questions and a behavioral scoring guide, an interviewee has less freedom to manage impressions in ways to mislead an interviewer than is the case in an unstructured interview. The interviewers record in writing the interviewee’s response to each question. Note-taking reduces the biasing effects of pre-interview expectations (Biesanz, Neuberg, Judice, & Smith, 1999). The use of two or more interviewers minimises idiosyncratic biases of a single interviewer, and in addition reduces random error through the aggregation of ratings (Dipboye, 1992). The questions are scored at the end of the interview by two or more interviewers, independently. This delay in judgment along with the knowledge that one’s assessment will be compared with another interviewer’s assessment minimises snap judgments that typically occur in an unstructured interview (Dipboye et al., 2004).

A very similar, low-cost alternative to an SI is the SJT, typically administered as a pencil and paper test (Weekley & Ployhart, 2006). As is the case with the SI, people who take the SJT are presented with a variety of situations that they are likely to encounter on the job (McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001). These questions too are typically derived from the CIT. Unlike the SI, which asks for an interviewee to generate a course of action (i.e. “What would you do in this situation?”), the SJT, which is free of interviewer bias, requires two judgments from a respondent, one regarding the best, and the other regarding the worst course of action from those provided by the test developer. Frary (1982) found that a multiple-response format often yields a higher validity coefficient than a single response.

Further similarities between the SI and SJT include the fact that both methods have incremental validity beyond an assessment of cognitive ability (Clevenger, Pereiera, Wiechmann, Schmitt, & Schmidt-Harvey, 2001; Klehe &
Both methods minimise adverse impact on racial subgroups (Lin, Dobbins, & Farh, 1992; Hough, Oswald, & Ployhart, 2001). Klehe and Latham (2006) showed that the SI can serve as either an assessment of typical or maximum performance. McDaniel et al. (2007) found that SJTs too can assess typical or maximum performance.

The SJT has been shown to be a valid predictor of performance in simple as well as complex tasks. For example, Hedge, Bruskiewicz, Borman, Hanson, Logan, and Siem (2000) found that it had validity for selecting airline pilots. The SI has been shown to be valid for predicting the organizational citizenship behavior of university professors (Latham & Skarlicki, 1995) as well as the job performance of social workers (Gibb & Taylor, 2003) and correction officers (Stohr-Gillmore, Stohr-Gillmore, & Kistler, 1990).

Both the SI and the SJT represent a method rather than a construct that has been used to assess a range of criteria. Typically they are used to assess judgment (Weekley & Ployhart, 2006). As noted earlier, Frese and colleagues also found that both methods are valid measures of personal initiative (Frese, Fay, Hilburger, & Leng, 1997; Bledow & Frese, 2009). Because of the similarities between the SI and SJT, they have been referred to as “close cousins” (Weekley & Gier, 1987; Weekley & Ployhart, 2006). Schmitt and Chan (2006) argued that “SJTs are unique in some way, perhaps most obviously, as measures of the practical use of a variety of information in sometimes ambiguous situations to make good decisions” (p. 147). The same can be said of the SI.

**CULTURE AND CRITERION-RELATED VALIDITY**

To the extent that a selection method clashes with the values and norms of a country, one can expect problems in implementation (Dipboye, Smith, & Howell, 1994). Spence and Petrick (2000, p. 63) concluded from their research that “national culture values can supersede and undermine . . . cosmetic attempts to standardize the formal interview questions”. Similarly, Hofstede (2001) argued that caution must be exercised in transporting behavioral science procedures that have been found to be effective in one culture to another culture that has a different value system.

Iran is high on uncertainty avoidance relative to Canada and the US (Pearce & Osmand, 1999; Ronen & Shenkar, 1985). To minimise uncertainty, the Iranian government passes strict rules, laws, policies, and regulations that are to be uniformly enforced throughout the country. The ultimate goal of the government, Hofstede (1983) stated, is to control everything in order to avoid, if not eliminate, the unexpected. Thus it would appear that the Iranian value system suggests a preference for written tests, where everything is predetermined, as opposed to an interview where unanticipated spontaneous answers may cause one or more parties’ embarrassment. Further, written
choice-tests such as the SJT are likely to be seen as more objective than an interview because there is no allowance for elaboration. This reduces the uncertainty associated with the evaluation/grading of an applicant’s answer to a selection test by another individual.

Based on the extant literature, five hypotheses were tested. First, both responses to the SI and the SJT correlate significantly with job performance. Second, as a low fidelity simulation, the SJT is a more distal predictor of job performance than the SI. This is because the SI requires employees to make the same situational judgments as the SJT, but in addition the SI requires them to develop situational responses rather than make selections from standardised answers. Furthermore, an SI is closer to the situation of a performance appraisal because in both settings an employee is evaluated by another person as opposed to a computerised assessment of an employee’s responses to an SJT. Hence, the SI should have incremental validity with regard to an SJT. Third, for these same reasons it was hypothesised that: SJT→SI→job performance. That is, the SI fully mediates the relationship between the SJT and job performance.

**REACTIONS TO SELECTION METHODS**

Test-taking anxiety has been shown to correlate negatively with both test (Schmitt & Ryan, 1992) and interview performance (McCarthy & Goffin, 2004). Oral presentations are rarely required of students in the Iranian education system, even at the university level. Hence, a job applicant in Iran has rarely been exposed to face-to-face interactions where the outcome will be a formal assessment of that person’s performance, let alone a decision as to whether to make that person a job offer. Typical of a collectivist society, job offers are often made, as noted previously, on the basis of nepotism and friendship. Hence, employees in Iran may experience more anxiety during an SI than they do when they take an SJT. Thus, a fourth hypothesis is that anxiety is higher during an SI than it is when taking an SJT. If the first three hypotheses are rejected, this fourth hypothesis, if supported, provides an explanation.

Another cultural factor that may affect employees’ reactions to different selection methods in Iran is the dress code. Although Iranian society is highly homogeneous (e.g. 90% are Shiite Muslim), Iranians vary in their adherence to the dress code advocated by the Iranian government for its citizens. All women are required by law to wear a scarf covering their hair in public, and a garment that allows only their hands and ankles to be visible. Nevertheless, women differ in their conformity to these norms (e.g. amount of facial make-up worn in public, visible hair, length of garment). Men too vary in adherence to shaved vs. unshaven faces, open/closed shirt, and use of a necktie. The latter, as noted earlier, is frowned upon by the government as it
is seen as a symbol of allegiance to Western norms, as is a cleanly shaven face. The company in which this study took place is owned by the Iranian government; hence, employees may be particularly prone to anxiety during a selection interview. This is a further reason to hypothesise that employees’ anxiety is higher in the SI than it is in an SJT: One’s appearance is salient in the former, but not in the latter.

Once an organisation has made a selection decision, the person’s willingness to accept a job offer becomes an important issue (Chapman & Zweig, 2005). Reactions to and perceptions of the fairness of the selection method affect this decision (Hausknecht et al., 2004). The SJT may appear more objective and concrete than the SI to Iranians. Hence, the fifth hypothesis is that the SJT is perceived to be significantly more fair than the SI.

**METHOD**

**Sample**

Seventy-five male and 26 female employees (n = 101) in the sales department of an Iranian automobile company participated in a concurrent validity study. Their mean work experience for the company was four years (SD = 1.95). The job requirements include interacting with automobile dealerships as well as customers. The employees are responsible for (a) planning in terms of calculating monthly quotas for the production of different types of cars as well as (b) the colors of those cars to be shipped to different dealerships, (c) notifying the customer when the car will be ready, (d) processing the final payments, (e) facilitating license plate and insurance coverage, and (f) writing the legal ownership document. If the car is not ready for the customer on the agreed upon date, these employees (g) must deal with customer complaints and (h) find/suggest alternatives (e.g. a price discount; another car in the same price range). Finally, these employees are responsible for scheduling the (i) repair and maintenance of a purchased car during the warranty period. In short, the job requirements involve judgment.

**Procedure**

The employees were informed that the purpose of this study was to assess the usefulness (i.e. validity) of two selection methods for hiring future job applicants. They were reminded of the difficulties that occur for all concerned when a bad hiring decision is made.

*Development of SI.* Consistent with the procedures outlined by Latham (1989) and Latham and Sue-Chan (1996), the CIT was conducted with managers of the employees. An example of a critical incident is (a) Circumstances:
The owner’s cousin was hired. (b) Behavior: The cousin spent most days on the phone talking to friends/family rather than waiting on customers. (c) Result: The company lost business. Fellow employees had to work extra hard to wait on the customers the cousin ignored while talking on the phone.

The resulting critical incidents were transformed into 11 SI questions. In addition, the subject matter experts developed behavioral illustrations of an excellent (5), acceptable (3), and an unacceptable (1) answer to each question. Interviewee responses were scored on a 5-point Likert type scale. An example of an SI question is in the Appendix.

Development of SJT. The 11 SJT questions were identical to those in the SI. Hence, they too were derived from the same job analysis. The scoring key was developed by subject matter experts because (a) this is the procedure that is used most frequently with SJTs (Bergman, Drasgow, Donovan, Henning, & Juraska, 2006), and (b) this method is consistent with the development of a scoring guide for the SI. Specifically, the same group of managers who developed the scoring guide for the SI determined the best and the worst of those item responses for each SJT question. Thus the SI and SJT questions were identical, the same SMEs were used, and the respective scoring guide for the SI and scoring key for the SJT were similar. Consequently, unnecessary variation between the two selection techniques was minimised. The SJT questions were followed by three responses from which employees had to choose the best and the worst response. The responses were scored on a 5-point Likert type scale. The final SJT score was the mean of employees’ score on the best and worst response.

Administration of the SI and SJT. The order in which the two selection methods were administered to each interviewee was counterbalanced. The first author, an Iranian, conducted the SI in Farsi and recorded the interviewee answers in writing. Interviewees were given the same questions in the same order. The author, as well as a manager in the company, scored the answers using the behavioral scoring guide, independently. To minimise knowledge of an employee’s job performance biasing the manager’s assessment of an employee’s responses to the SI questions, the manager saw only the written answers the first author had recorded. Thus the first author was blind to the interviewee’s job performance and the manager was blind to the identity of the interviewee. The scoring of each answer was finalised through consensus. Consensus was reached after discussing a performance dimension, the question assessing that dimension, and re-reading the recorded answer to an item given by an interviewee.

In the SJT, the questions were presented to respondents in the same order as the SI. The answers were evaluated against the predetermined scoring key.
To minimise the probability of responses to one test influencing responses to the second, there was a one-week time lag between the administration of the two tests. Performance on each test was calculated by summing the scores given to the respective items.

**Anxiety.** Test and interview anxiety was assessed using the scale developed by McCarthy and Goffin (2004). The scale assesses five different aspects of anxiety (communication, behavior, performance, social, and appearance). The internal consistency of the overall scale was 0.83. A sample item for the situational interview anxiety is: “In job interviews I get very nervous about whether my performance is good enough.” Similarly, a sample item for situational judgment test anxiety is: “In a situational judgment test, I get very nervous about whether my performance is good enough.”

**Fairness.** The perceived fairness of the SI and SJT was assessed using 11 items developed by Bauer, Truxillo, Sanchez, Craig, Ferrara, and Campion (2001). This scale was used because Bauer et al. (2001) found that the internal consistency of the scale was satisfactory ($\alpha = 0.71$). A sample item is: The test was administered to all applicants in the same way.

**Motivation.** Because this was a concurrent validity study, the present employees may not have been motivated to do well on the tests. Hence the test scores might be uniformly low. If this were true, it would explain why the validity coefficient of the SI, the SJT, or both were not significant (Hausknecht et al., 2004; Schmidt & Ryan, 1992). Consequently, motivation to do well on the two selection methods was assessed using the 10-item Motivation Scale developed by Arvey, Strickland, Drauden, and Martin (1990). This scale was used because those authors had shown that the scale has high internal consistency ($\alpha = 0.85$). A sample item for the SI is: “Doing well in this interview is important to me.” A sample item for the SJT is: “Doing well on this test is important to me.”

The SI and SJT questions were presented in Farsi. Because the measures of fairness, anxiety, and motivation were originally written and published in English, those items were translated by the first author from English into Farsi. The accuracy of the translation was verified through back-translation by five people who speak both English and Farsi fluently. In addition, the wording of the scales in Farsi was subsequently examined by the company’s managers to ensure understanding of each item by their employees. These steps were taken because people from different cultures, such as Iran vs. the US where the items were originally written, may perceive the test items differently (Whetzel, McDaniel, & Nguyen, 2008). Responses to these three scales were made immediately following the administration of the SI and SJT. All these scales were measured on a 7-point Likert type scale ranging from “strongly disagree” (1) to “strongly agree” (7).

© 2009 The Authors. Journal compilation © 2009 International Association of Applied Psychology.
Performance. The job performance of each employee was assessed by their respective managers \((n = 5)\) using the company’s performance appraisal instrument. Specifically, employees were assessed on 14 items (e.g. problem-solving ability, team work capabilities, work precision) using a 7-point Likert type scale (weak–excellent). All the performance items assess task performance. No contextual performance measures are used by the company.

Managers were informed that the assessments of their employees would be shown only to the researchers. None of the managers was aware of an employee’s score on the SI or SJT when they appraised their employees’ job performance.

RESULTS

The Cronbach alpha coefficients for the SI and SJT were 0.84 and 0.83, respectively. The correlation between the scores on these two tests was .31 \((p < .01)\). The respective criterion-related validity coefficients of the SI and SJT with job performance were 0.28 and 0.23 \((p < .01\) and \(p < .01\), respectively). These two validity coefficients were not significantly different. Table 1 presents the intercorrelations among the variables of this study.

The scores on the SJT for males \((n = 72)\) and females \((n = 26)\) were 37.65 \((SD = 3.89)\) and 37.73 \((SD = 4.12)\), respectively. This difference was not significant. The SI mean scores for males \((37.55, SD = 3.78)\) and females \((37.35, SD = 3.85)\) did not differ significantly.

The incremental validity of the two methods was assessed using two regression analyses. In the first regression, the SI was entered first, followed by the SJT. In the second regression, the order of the two tests was reversed. The SI accounted for variance in the criterion over and above the SJT.

TABLE 1

<table>
<thead>
<tr>
<th>Correlation</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Performance</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. SI score</td>
<td>0.28*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. SJT score</td>
<td>0.23*</td>
<td>0.31*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. SI anxiety</td>
<td>0.14</td>
<td>-0.01</td>
<td>-0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. SI fairness</td>
<td>0.08</td>
<td>-0.16</td>
<td>-0.12</td>
<td>-0.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. SI motivation</td>
<td>0.27*</td>
<td>-0.03</td>
<td>-0.02</td>
<td>0.19</td>
<td>0.42*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. SJT anxiety</td>
<td>0.04</td>
<td>-0.14</td>
<td>-0.02</td>
<td>0.79*</td>
<td>-0.06</td>
<td>0.21*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. SJT fairness</td>
<td>0.08</td>
<td>-0.04</td>
<td>-0.16</td>
<td>0.03</td>
<td>0.63*</td>
<td>0.33*</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>9. SJT motivation</td>
<td>0.17</td>
<td>-0.01</td>
<td>-0.15</td>
<td>0.17</td>
<td>0.41*</td>
<td>0.74*</td>
<td>0.24*</td>
<td>0.41*</td>
</tr>
</tbody>
</table>

Note: * \(p < .05\).

© 2009 The Authors. Journal compilation © 2009 International Association of Applied Psychology.
Specifically, the SI accounted for an additional 5 per cent of unexplained variance ($\Delta F_{(1, \ 90)} = 5.08, \ p = .02, \ \Delta R^2 = 0.05$). The SJT did not demonstrate incremental validity beyond the SI ($\Delta F_{(1, \ 90)} = 2.16, \ p = .15, \ \Delta R^2 = 0.02$).

A mediation model was tested in which the SI was considered as a more proximal predictor of performance, and the SJT as a more distal one. Following Baron and Kenny’s (1986) model, the three steps of mediation analysis were conducted. In the first step, the mediator (SI) was regressed on the predictor variable (SJT). The coefficient was 0.30 ($p = .002$). In the second step, the criterion variable (performance) was regressed on the predictor (SJT). The corresponding beta coefficient was 1.64 ($p = .03$). In the last step the criterion variable was regressed on both the mediator (SI) and the predictor (SJT). At this step the significant relationship between the SJT and performance became non-significant ($b = 1.12, \ p = .15$), but the relationship between the SI and performance remained statistically significant ($b = 1.81, \ p = .03$). Thus, the SI fully mediated the relationship between the SJT and job performance. To test the indirect effect of the SJT on performance, the Sobel (1982) test was also conducted. The indirect effect was significant ($z = 2.17, \ p = .03$).

There was no significant difference in the employees’ motivation to perform well on the SI ($M = 5.13, SD = 0.85$) versus the SJT ($M = 5.20, SD = 0.87$). Motivation to do well on both tests was moderately high. There was no significant difference between the SI ($M = 2.44, SD = 0.73$) and the SJT ($M = 2.52, SD = 0.78$) on test anxiety. Finally, both the SI ($M = 5.31, SD = 0.64$) and the SJT were perceived as fair ($M = 5.34, SD = 0.60$).

Prior to this study, none of the employees had been exposed to formal selection procedures, let alone an SI or SJT. Hence, employees were asked to recommend which method, if any, to use for selection purposes in the company. Of the 101 employees, 42 recommended the SI, eight recommended the SJT, 33 found both acceptable, but five recommended that neither be used. A chi-square test indicated that these differences were significant ($X^2_{(3)} = 45.24, \ p < .001$).

**DISCUSSION**

The first three hypotheses of this study were supported. Both the SI and the SJT correlated significantly with job performance. Only the SI accounted for incremental variance. Further, the SI fully mediated the relationship between the SJT and job performance. No support was found for the fourth and fifth hypotheses regarding differences in reactions to the SI and SJT in terms of anxiety, motivation, and fairness.

Consistent with the findings obtained in a meta-analysis by Whetzel et al. (2008), SJT scores did not have an adverse impact by sex. The same was shown to be true with regard to SI scores. Consistent with the results.
obtained in the West (Hausknecht et al., 2004), Iranian employees indicated their preference for the interview as a method of hiring employees.

The practical significance of these findings is at least fourfold. First, the SJT appears to be useful as a first step in a selection process followed by the SI.

Second, the SI and the SJT were shown to have criterion-related validity in a non-Western country that has societal values different from those in the West. Previous studies of the SI have been limited to Canada (e.g. Klehe & Latham, 2005), Germany (e.g. Moser & Schuler, 2001), New Zealand (e.g. Gibb & Taylor, 2003), and the United States (e.g. Weekley & Gier, 1987). All previous studies on the SJT have been conducted in Australia, Mexico, the United States or Western Europe (Lievens, 2006). The present finding suggests the global applicability (Triandis & Suh, 2002) of these two selection methods.

Third, of practical significance for multinational organisations is the finding that these two selection methods can be used in an organisation regardless of whether it is located in the West or the East. Their use only requires tailoring the questions and the scoring guide/scoring key to the societal culture where they will be applied. Such and Schmidt (2004) found that when the same SJT questions and scoring guide were implemented in the same company that operated in multiple countries, the SJT was found to have criterion-related validity in Australia and the United Kingdom, but not in Mexico. This is because situations likely to be encountered in one country, or in one organisational setting, regardless of the country where it is located, may not be relevant in another. Moreover, an acceptable response in the Middle East may be less appropriate in the West (e.g. emphasise/de-emphasise group harmony; emphasise/de-emphasise religious values in the workplace). Finally, as Whetzel et al. (2008) have noted, differential item functioning theory states that people from different cultures may perceive the same test items differently and, hence, obtain different test scores. A practical advantage of the SI and SJT is that, unlike pencil and paper cognitive ability and personality tests, the questions/items can be contextualised with regard to the societal culture where they will be applied (Lievens, 2006). This explains why differences in cultural values between Euro-Western/Judeo-Christian versus Islamic countries are unlikely to be a factor limiting the validity or acceptability of these two selection methods. The contextualising of the questions/items to fit the societal culture where the SI and SJT are used enables them to be seen by a local population as congruent with the norms and values where they are to be employed. Valid tests that are copyrighted in the West are likely to be rejected in the present political climate in Iran. By using an SJT and an SI, an organisation, regardless of its geographical setting, is able to make selection decisions that are not only valid, but are seen by employees as fair. In short, the use of an SI and SJT permits a uniform format to the selection process that allows factoring in differences in societal values which in turn enhances the judgment capacity of those who make hiring decisions.
Fourth, Bauer and Truxillo (2006) argued that the SJT should be perceived positively by people because the job relatedness is “obvious”. Their contention was supported in the present study. The present finding with regard to the SI is consistent with Steiner and Gilliland’s (2001) conclusion that, with the exception of work samples, interviews receive favorable reactions relative to other selection methods in Belgium, France, Spain, South Africa, and the United States. Contrary to our hypothesis, the same appears to be true for the situational interview in Iran.

Fifth, only five of the 101 employees who participated in this study recommended that these two selection methods not be used for hiring employees. This finding is especially noteworthy given that these employees had no previous experience with either the SI or the SJT. It is also noteworthy that these employees had no knowledge of the concept of criterion-related validity. Moreover, there are no laws prohibiting employment discrimination in Iran. Employees are accustomed to hiring decisions being made on the basis of “who you know” rather than on how you perform in the selection process.

Possible Limitations and Future Research

The drawbacks of this study are arguably sixfold. First, Arthur and Villado (2008) have warned against confounding constructs and methods when comparing the validity coefficients of different predictors because doing so can lead to conclusions that are theoretically/conceptually uninterpretable. They cited approvingly a study by Lievens and Sackett (2006) where the criterion-related validity of a video-based and a written situational judgment test were compared. In doing so, the predictor constructs of the SJTs were held constant. Similarly, Bauer, Truxillo, Paronto, Weekley, and Campion (2004) examined reactions of students to a face-to-face, a telephone, and an interactive voice response interview. Consistent with those two studies, the predictor construct in the present study was identical. The SI and the SJT presented the same job-related situations. The only difference was giving an answer (SI) versus choosing an answer (SJT). Thus the hypothesis that confounding construct with method is a limitation of this study was rejected.

That the correlation between responses to the SJT and SI was only moderately high may have been due to the respective instructions. The SJT used a knowledge instructions format (i.e. choose the best/worst response); the SI used a behavioral tendency instruction format (What would you do?). Hence, the SJT may have been more cognitively loaded while the SI may have been more personality loaded. Future research is needed to test this hypothesis.

Third, this was a concurrent rather than a predictive validity study. Nevertheless, the motivation of the present employees to perform well on the SI
and SJT was relatively high. Moreover, Schmit and Motowidlo (1995) found that the concurrent validity of the SI ($r = .32$) was not affected when tenure, a proxy they used for job knowledge, was partialled out.

Fourth, the criterion for validating the two tests was the organisation’s existing performance appraisal instrument. Not only was it essentially a trait scale, it was generic for the organisation as a whole. The SI/SJT questions were developed specifically to assess the employees in the sales department. Had the criterion too been developed to assess the critical behaviors exhibited by these employees on the job, the two validity coefficients might have been significantly higher.

A fifth limitation, common to most single studies, is the relatively small sample size of the employees. However, such studies are necessary in order to enable meta-analyses of studies that examine socio-culture (e.g. East vs. West) and method (e.g. interview vs. questionnaire) as moderating variables that affect validity coefficients.

Sixth, the SJT in this study had relatively high internal consistency. To capture the breadth of job performance, a lower Cronbach alpha is typically preferable (McDaniel et al., 2001). The SJT might have had higher validity had there been additional items. The present SJT items were the same as those asked in the SI. It is possible that more SJT items are needed than SI questions to increase validity. Again, future research is required to examine this possibility.

### CONCLUSIONS

Triandis and Suh (2002) stressed the importance of investigating the external validity of methods and procedures in distinctly different societies in order to identify universal principles in the behavioral sciences. The present study is the first to show the comparability of the criterion-related validity of the SI and the SJT. The validity coefficients of both tests were significant. The magnitude of the criterion-related validity of the SI in the present study was similar to the uncorrected mean validity coefficient of .29 obtained in a meta-analysis of SI studies conducted in Western countries (Latham & Sue-Chan, 1999). Similarly, the magnitude of the criterion-related validity coefficient of the SJT obtained in the present study compares favorably with the results of a meta-analysis on SJT studies conducted in the West where the estimated population correlation coefficient was found to be .26 (McDaniel et al., 2007). Thus it would appear that the applicability and validity of these two methods for selecting employees is not restricted to Western cultures. This is likely due to the fact that there are no pre-set questions in the SJT or the SI. Hence, unlike the standardised wording of items in most cognitive ability or personality tests, the situations presented to the employees were Iranian rather than Western based.
REFERENCES


© 2009 The Authors. Journal compilation © 2009 International Association of Applied Psychology.


© 2009 The Authors. Journal compilation © 2009 International Association of Applied Psychology.


© 2009 The Authors. Journal compilation © 2009 International Association of Applied Psychology.
APPENDIX

Sample SI/SJT Question

You are in the unit meeting. Your manager blames you for not doing well on a task, in front of all your peers and managers from other divisions. You believe that your manager is wrong in his critique, and that he might have come to this conclusion hastily without knowing all the information. You feel you are being treated unfairly in front of your peers. You feel that your reputation may be affected by this critique. What would you do in this situation?