



Comparing the validity of structured interviews for managerial-level employees: Should we look to the past or focus on the future?

Henryk T. Krajewski^{1*}, Richard D. Goffin², Julie M. McCarthy³,
Mitchell G. Rothstein⁴ and Norman Johnston⁵

¹ MICA Consulting Partners, Inc., Canada

² Department of Psychology, University of Western Ontario, Canada

³ Rotman School of Management, University of Toronto, Canada

⁴ Richard Ivey School of Business, University of Western Ontario, Canada

⁵ Private Practice, Toronto, Ontario, Canada

The current research investigated questions that persist regarding the criterion-related and construct validity of situational (SI) versus past-behaviour (PBI) structured interview formats in predicting the job performance of managers. Analyses of data collected from 157 applicants to managerial positions showed that the PBI format significantly predicted job performance ratings ($r = .32, p < .01$), whereas the SI format did not ($r = .09, ns$). Investigation of potential construct differences between the SI and PBI formats showed that the PBI was more highly related to manager-relevant cognitive ability measures, assessment centre exercises and personality traits, as compared with the SI. Such differences help to explain the predictive validity differences between the SI and PBI observed in current and previous research.

Despite two recent meta-analyses (Latham & Sue-Chan, 1999; Taylor & Small, 2002) that have investigated the criterion-related validity of the situational interview (SI) and the past-behaviour interview (PBI), there remains debate as to the criterion-related validity of each of these structured interview types in predicting job performance – particularly with respect to the domain of managerial job performance (see Huffcutt, Weekley, Weisner, DeGroot, & Jones, 2001; Pulakos & Schmitt, 1995; Taylor & Small, 2002). There are three problems with past research that have given rise to the debate surrounding the criterion-related validity of the SI and PBI. First, there are few studies that have concomitantly examined the relation of the SI and PBI with managerial performance in an actual selection context – a fact that has, of course, attenuated the ability to build consensus regarding the relative predictive validity of the SI versus the PBI. Second,

*Correspondence should be addressed to Henryk T. Krajewski, MICA Consulting Partners, 229 Yonge Street, Suite 400, Toronto, Ontario M5B 1N9, Canada (e-mail: hkrajewski@micaworld.com).

similarly designed studies that have investigated the SI, the PBI or both continue to produce conflicting results. For example, some past research has found that the SI is a valid predictor of job performance (Campion, Campion, & Hudson, 1994; Day & Carroll, 2003; Latham & Skarlicki, 1995; Latham & Sue-Chan, 1999) and is superior to the PBI in predicting job performance (Latham & Sue-Chan, 1999). Yet other past research has shown that the SI is a less valid predictor of job performance than the PBI for a range of job types (Taylor & Small, 2002) and that the SI actually shows zero validity in predicting performance in higher-level jobs (Huffcutt *et al.*, 2001; Pulakos & Schmitt, 1995). Third, close inspection of the studies that have investigated the validity of the SI and the PBI in predicting managerial job performance shows that there are several methodological deficiencies and design limitations present that may have affected the accuracy or generality of the observed results (see Latham & Sue-Chan, 1999; discussed further below).

Additionally, questions persist regarding the construct validity of both the SI and the PBI (see Conway & Peneno, 1999; Cook, Vance, & Spector, 2000; Huffcutt, Roth, & McDaniel, 1996). Indeed, few studies have comprehensively examined construct differences between the SI and the PBI and very few studies have examined *a priori*-specified construct differences between the SI and the PBI as means of explaining observed criterion-related validity differences. More generally, Posthuma, Morgeson, and Campion (2002) support the need for additional research on the constructs measured by the employment interview.

Given the issues outlined above, the overall goals of the current study were to (a) give a clearer picture as to the criterion-related and incremental validity of the SI versus the PBI in predicting managerial job performance by improving upon methodological flaws and design limitations in past research and (b) examine relations of the SI and the PBI with numerous facets of cognitive ability, personality and work-sample tests in order to illuminate construct differences between the SI and the PBI that may help to explain any observed differences in criterion-related validity. The potential to provide more accurate and generalizable findings regarding the relative criterion-related and incremental validity of the SI and the PBI is a contribution to a relatively thin literature comparing the validity of differently structured interview types in predicting managerial job performance. Moreover, within the context of a more rigorously conducted test of criterion-related validity, any potential construct differences that are observed between the SI and the PBI may allow substantive inferences to be drawn with regard to *why* criterion-related validity differences were observed between the SI and the PBI – particularly, in the present case, with respect to managerial positions. The current study went a step beyond past research not only in examining a wide range of construct relations, but also in examining the relation of specific managerially relevant facets of these constructs with the SI and the PBI. As stated by Huffcutt *et al.* (1996), research that serves to explicate the constructs that are captured by structured interview types may provide ‘the next major breakthrough in construct research’ (p. 470). In terms of applied contributions, possessing information about which constructs are being tapped by different interview types may help organizations to choose the most appropriate interview format, depending on the specific predictors the organization is interested in assessing (Jelf, 1999; Taylor & Small, 2002).

Description of the SI and the PBI

The SI and PBI are two of the most commonly used structured interview types for selecting employees (Campion, Palmer, & Campion, 1997). The SI requires applicants to

provide responses to hypothetical dilemmas that might be encountered in the job in question (Latham, Saari, Pursell, & Campion, 1980), whereas the PBI requires that applicants recall past experiences relevant to the types of tasks and situations likely to be encountered on the job (e.g. Janz, 1982). The SI has roots in goal-setting theory in that behavioural intentions (i.e. goals), as expressed in response to hypothetical scenarios, are thought to be the immediate precursors to actions (Latham & Sue-Chan, 1996), whereas the PBI is based on the premise that the best predictor of future behaviour in a given context is past behaviour exhibited in a similar context (Janz, 1989). It is recognized that Janz labelled his interview format the 'patterned behaviour description interview' (PBDD); however, as has been the norm in past research, we use the label PBI to refer, more generally, to structured interviews using the past-behaviour format. Similarly, although we try to be faithful to Latham *et al.*'s (1980) format for designing future-oriented situational interview questions, we recognize that some may argue that we are only employing situational interview 'types' of questions.

Comparing the predictive validity of the SI and PBI

In directly comparing the validity of the SI and PBI, care must be taken to ensure the respective interviews are developed, administered and scored in as similar a manner as possible while still preserving the essential and distinctive aspects of these two formats. Holding constant such elements of the research design helps increase one's confidence that any observed differences in validity are due to the presumed core differences in the SI and PBI formats (i.e. the focus on hypothetical responses [SI] versus actual past behaviours [PBI]) rather than developmental or procedural artifacts. Indeed, Latham and Sue-Chan (1999) noted that past research comparing the criterion-related validity of the SI and the PBI has been confounded by (a) a failure to rate each SI question separately as opposed to giving a global rating after the completion of the interview (e.g. Pulakos & Schmitt, 1995), (b) a failure to include a scoring guide for the PBI (e.g. Latham & Scarlicki, 1995), (c) a failure to tap the same content areas (e.g. the interviewee's willingness to learn, team orientation etc.) in SI and PBI questions (e.g. Campion *et al.*, 1994) and (d) a failure to properly operationalize the SI by neglecting to use a true situational 'dilemma' that helps to disguise the most correct answers (e.g. Campion *et al.*, 1994; Pulakos & Schmitt, 1995).

In order to address limitations of past research that resulted in inequitable comparisons of the SI and the PBI, the current study sought to create SI and PBI measures that were as similar as possible. Accordingly, the PBI and SI interviews were strictly developed as prescribed in the literature. Further, the SI and the PBI were designed to assess identical managerial skill dimensions. Measuring the same managerial skill dimensions in each interview facilitated the use of the same behaviourally anchored scoring guide for each interview. Finally, the current study required that assessors rate each managerial skill dimension in each interview immediately after the relevant question, as opposed to making ratings after all the managerial skill dimensions have been assessed.

In addition to the above-mentioned methodological problems, past research that has directly compared the SI and the PBI has used concurrent rather than predictive validity designs (Campion *et al.*, 1994; Huffcutt *et al.*, 2001; Pulakos & Schmitt, 1995), non-employee (Conway & Peneno, 1999; Day & Carroll, 2003) or non-managerial (Campion *et al.*, 1994; Huffcutt *et al.*, 2001; Latham & Scarlicki, 1995; Pulakos & Schmitt, 1995) populations, or small sample sizes (Huffcutt *et al.*, 2001; Latham & Scarlicki, 1995). As is

well known, the use of concurrent validity designs and non-employee populations can limit the ability to generalize to applicant samples, whereas extremely small sample sizes are more likely to be unrepresentative and result in reduced power. Accordingly, the current research used a predictive validity design to directly compare the validity of the SI and the PBI within a relatively large sample of applicants to high-level managerial positions.

Predictive validity hypotheses

Although past studies have shown that the PBI predicts job performance more strongly than the SI in samples consisting of applicants for higher-level positions (Huffcutt *et al.*, 2001; Pulakos & Schmitt, 1995; Taylor & Small, 2002), methodological flaws or design limitations may have contributed to the less than impressive results for the SI (see Latham & Sue-Chan, 1999). Because the current study attempts to correct for methodological flaws and design limitations of past studies, it is expected that:

Hypothesis 1a. Both the SI and PBI will show a positive and significant relation with managerial job performance.

However, with regards to examining the relative predictive validity of the SI and the PBI, it is instructive to consider the distinction between 'will do' and 'can do' determinants of job performance. 'Will do' determinants of performance refer to individuals' motivation or willingness to expend effort in their jobs, whereas 'can do' determinants of performance refer to individuals' capacity for performing the job (for reviews, see Blumberg & Pringle, 1982; Campbell, 1990). Taylor and Small (2002) suggested that, because of its focus on hypothetical, future-oriented scenarios rather than actual past performance, the SI predominantly taps the 'can do' performance determinants and may not adequately reflect the 'will do' determinants. In contrast, by requiring applicants to provide examples of *actual past behaviours* that are reflective of their day-to-day motivation to apply relevant knowledge, skills and abilities to the job, the PBI is purported to more adequately capture both the 'can do' and 'will do' determinants of job performance (Taylor & Small, 2002). Thus, it is expected that:

Hypothesis 1b. The PBI will show incremental validity over the SI in predicting managerial job performance, whereas the SI will not show incremental validity over the PBI in predicting managerial job performance.

Examination of the construct differences between the SI and PBI

Although few studies have explicitly examined whether construct differences can help explain differences in the predictive validity of the SI and PBI, past research does support the notion that there are meaningful construct differences between the SI and PBI. Indeed, past research has shown that responses to SI and PBI questions designed to measure the same job characteristic do not tend to correlate (Conway & Peneno, 1999; Huffcutt *et al.*, 2001). Further, as stated above, it has been posited that the SI predominantly measures the knowledge-based aspects of job performance, whereas the PBI is purported to measure both the knowledge-based and motivational aspects of job performance (Taylor & Small, 2002). Nevertheless, very few studies have simultaneously examined, in detail, whether meaningful construct differences exist between the SI and PBI that could explain differences in criterion-related validity. Thus, with a view to

explaining the pattern of results derived from the criterion-related and incremental validity analyses, we examined the relation of the SI and the PBI with cognitive ability, leaderless group discussion and in-basket exercise performance, and relevant facets of personality.

Hypotheses concerning the relations of the SI and PBI with cognitive ability

Cognitive ability has been widely acknowledged as a strong predictor of job performance, particularly in higher-level jobs (Hunter & Hunter, 1984; Hunter & Schmidt, 1989; Schmidt & Hunter, 1998). The specific cognitive ability facets selected for the current study were chosen based on meta-analytic results showing that tests of verbal reasoning, verbal comprehension and numerical ability from the employee aptitude survey (EAS; Ruch & Ruch, 1980) are particularly relevant to the job of manager (Ruch & Ruch, 1980).

Previous studies have posited that performance on the SI and PBI should be related to cognitive ability (e.g. Conway & Peneno, 1999; Huffcutt *et al.*, 1996, 2001; Pulakos & Schmitt, 1995). Indeed, it seems reasonable that individuals with well-developed cognitive abilities should have had past successes that could be demonstrated in the PBI, or should have amassed the requisite job knowledge that can be used to appropriately answer questions in the SI. Consistent with the above arguments, it is expected that:

Hypothesis 2a. The SI and PBI will each show a positive and significant relation with managerially relevant cognitive ability facets.

Nevertheless, as suggested above, the SI requires applicants to provide the 'right' answers to challenging hypothetical dilemmas 'on the spot' (i.e. what would you do if 'x' occurred). In contrast, the PBI requires applicants to report about actual past activities (i.e. tell me about a time when you did 'y'). Although it is possible that the PBI may draw on cognitive ability to the extent that respondents judiciously select and embellish their choices of past behaviours, the possibility that these behaviours might potentially be verified is expected to reduce respondents' efforts to expend cognitive effort in recounting overly flattering accounts of themselves. In support of this, research on faking in selection settings has suggested that applicants are less likely to provide fake answers when the items in question are verifiable, as can be the case with past work behaviours (Becker & Colquitt, 1992; Donovan, Dwight, & Hurtz, 2003). Thus, the SI would seem to be more heavily reliant on deduction and reasoning than the PBI. Although it was unable to control for the various methodological confounds present in past research (discussed earlier), a past meta-analysis supported the reasoning that the SI should show stronger relations with cognitive ability than the PBI (Huffcutt *et al.*, 1996). On the basis of the research reviewed above, it is expected that:

Hypothesis 2b. The SI will be more highly correlated with manager-relevant cognitive ability facets than the PBI.

Hypotheses concerning the relation of the SI and PBI with work sample tests

In terms of identifying tests of skill in specific managerial tasks, both the leaderless group discussion (LGD) and in-basket test have been identified as work samples that assess critical and largely non-overlapping aspects of managerial functioning (Cascio, 1998). Therefore, the relations of the SI and PBI with both the LGD and

in-basket test were examined. In terms of the constructs underlying LGD and in-basket performance, Craik *et al.* (2002) provided some evidence that the predominant construct of interest underlying the LGD may be 'interpersonal managerial style', whereas the construct of interest underlying the in-basket may be 'strategic management style'. The positive pole of interpersonal managerial style reflects adeptness at oral communication, initiative, energy and leadership, whereas the positive pole of strategic management style reflects adeptness at decision-making, fact-finding, delegating, analytic thinking and planning-organization. Given that the questions for the SI and the PBI are derived on the basis of a thorough analysis of the managerial job-content domain, both interviews should capture some of the variance relevant to interpersonal management style (as reflected in the LGD) as well as strategic management style (as reflected in the in-basket test). Thus, it is expected that:

Hypothesis 3a. The SI and PBI will each show a positive and significant relation with LGD performance and in-basket performance.

We further hypothesized that the PBI would more effectively assess interpersonal managerial style (e.g. initiative, leadership) and, by extension, be more highly related to LGD scores than would the SI. Our rationale for this expectation was that, as discussed earlier, there is reason to believe that the PBI more adequately assesses both the 'can do' and 'will do' determinants of managerial performance, whereas the SI is likely to be less effective in its assessment of 'will do' aspects of performance. Thus, there are grounds for expecting that the PBI will tap the interpersonal managerial style that an applicant would actually employ on the job whereas the SI is more likely to tap the interpersonal managerial style that the applicant thinks is correct in a hypothetical situation, *whether or not he/she would actually use it on the job*. Using similar reasoning, it follows that the PBI would also more effectively assess strategic managerial style and, by extension, be more highly related to in-basket scores than the SI. On the basis of the preceding rationale, it is expected that:

Hypothesis 3b. The PBI will be more highly correlated with LGD performance and in-basket performance than will the SI.

Hypotheses concerning the relation of the SI and PBI with personality

Researchers have consistently found, or suggested, that personality traits associated with the 'Big Five' factor of conscientiousness are related to job performance (Barrick & Mount, 1991; Tett, Jackson, & Rothstein, 1991). However, further investigations have found that there is value in distinguishing between two facets of conscientiousness, namely, dependability and achievement orientation (Hough, 1992; Jackson, Paunonen, Frabroni, & Goffin, 1996), and have shown achievement orientation to be more highly correlated than dependability with job performance criteria (Hough, 1992, 1998; Hough, Barge, & Kamp, 2001). Previous meta-analytic work supports the preceding distinction with regard to predicting performance in the managerial and leadership domain (Judge, Bono, Ilies, & Gerhardt, 2002). Accordingly, the current research considered achievement orientation as the most relevant facet of conscientiousness with respect to managerial job performance (see Table 1).

Research on the correlates of effective management has also found that traits associated with the 'Big Five' factor of extraversion are positively related to managerial effectiveness (Bass, 1990; Judge *et al.*, 2002; Stogdill, 1974). In particular, the

Table 1. Descriptions of high and low scorers on PRF-E personality dimensions of achievement orientation and dominance

PRF-E scale	Description of high scorer	Defining trait adjectives
Achievement orientation	Aspires to accomplish difficult tasks; maintains high standards; willing to put forth effort to attain excellence	Striving, self-improving, capable, enterprising, ambitious, attaining, aspiring, driving
Dominance	Attempts to control environment and to influence and direct other people; expresses opinions forcefully; enjoys the role of leader and may spontaneously assume it	Governing, controlling, commanding, domineering, influential, persuasive, forceful, leading, directing, assertive, supervising

Note. Descriptors were taken from Jackson (1984, p.6). Reproduced with permission from the publisher Sigma Assessment Systems (Port Huron, MI: D/B/A Research Psychologists Press prior to 1989) and the test author. Copyright ©1984, 1989.

extraversion facet of dominance has a venerable history of being linked to leadership potential and performance (e.g. Bass, 1990; Bradley, Nichol, Charbonneau, & Meyer, 2002; Goffin, Rothstein, & Johnston, 1996; Gough, 1990; Howard & Bray, 1988; Jackson, Peacock, & Holden, 1982; Judge *et al.*, 2002; Steinberg & Shapiro, 1982; Stogdill, 1974). Accordingly, the current research also focused on the personality trait of dominance because it would appear to be the most relevant facet of extraversion with respect to the personality substrate of managerial job performance (see Table 1).

In terms of specific expectations, performance in the SI and the PBI should relate to managers' predilection for achievement orientation and dominance. Indeed, the nature of the managerial skill dimensions typically assessed in SI and PBI questions (e.g. willingness to lead, coaching, planning and organizing) – and managerial performance in general – appear to be at least somewhat dependent on the degree to which candidates possess the personality traits of achievement orientation and dominance. Although past research has been somewhat mixed with regard to the relation of personality to structured interviews, there is evidence in favour of a link between applicant personality and interview performance; particularly when an *a priori*, construct-driven strategy (see Tett *et al.*, 1991; Goffin, Rothstein, & Johnston, 2000) is used for matching predictors and criteria (Cook *et al.*, 2000; Huffcutt *et al.*, 2001). Thus, it is expected that:

Hypothesis 4a. The SI and PBI will each relate to the managerially relevant personality traits of achievement orientation and dominance.

However, as stated above, the PBI explicitly asks for actual evidence of the predilection to perform as a manager, whereas the SI merely assumes that candidates' statements as to how they would perform in hypothetical scenarios will accurately indicate their level of motivation to perform such activities once on the job. Given that the PBI may tap into a candidate's predilection to perform to a greater degree than the SI, it is hypothesized that:

Hypothesis 4b. The PBI will be more highly correlated with the managerially relevant personality traits of achievement orientation and dominance than the SI.

Method

Participants

The current sample consisted of 157 applicants (139 males, 17 females, 1 undeclared) to high-level managerial positions in a large forestry-products organization. The positions applied for involved activities such as planning and directing the activities of front-line personnel, ensuring that company policies and procedures were followed and evaluating subordinates. The ages of the participants ranged from 23 to 60 ($M = 38.46$, $SD = 7.77$). The current sample size far exceeded that required in order to meet the most conservative participant-to-predictor ratios recommended to ensure adequate power with respect to the intended analyses (i.e. 20 to 1; Tabachnick & Fidell, 1995).

Interview development and rating scales

The development of the interviews and associated scoring guides was a meticulous process that involved a range of company personnel (i.e. subject matter experts (SMEs), job incumbents) and consisted of multiple drafts and pre-tests. Six dimensions reflective of the knowledge, skills and abilities required of effective managers were derived on the basis of a thorough job analysis involving both SMEs and job incumbents in the participating organization. The six managerial skill dimensions identified were (a) planning and organizing, (b) coaching, (c) results orientation, (d) willingness to learn, (e) team orientation and (f) oral communication. Three senior management personnel and one industrial-organizational psychologist (the final author on the current paper) served as the assessor team. Two assessors were assigned to rate candidates in each interview format. All assessors were thoroughly trained prior to interview administration.

Situational interview

Six situational questions were used to measure each of the performance dimensions identified above. Notably, each situational question involved a hypothetical 'dilemma' for which the most correct answer was not apparent.

Past behaviour interview

Six past-behaviour questions were used to assess each of the six managerial skill dimensions. In the event that the applicant did not provide sufficient information for scoring the relevant dimension, standardized probes were made available to assessors to prompt responses, which is typical of the PBI (Janz, Helervik, & Gilmore, 1986).

Interview comparability

Two issues relevant to comparability between the interviews in the current study should be reiterated. First, comparability in content across the two interview types was ensured by the fact that both SI and PBI questions were written to tap the same managerial skill dimensions that were derived on the basis of the job analysis reported above. Extensive pre-tests and judgments of content validity were conducted to ensure that the same managerial skill dimensions could be tapped by each set of respective SI and PBI questions. To illustrate each of the question types, consider the examples below that are illustrative of the type of questions designed to measure the 'team orientation' skill dimension - defined as the willingness and ability to display interpersonal awareness,

skill, conflict management and team building skills. The actual SI and PBI interview questions cannot be reproduced in their entirety due to a privacy agreement held with the participating company.

SI question: As a manager, you have scheduled a recognition dinner to reward your team for their recent performance. However, you find out that your employees really don't care about the dinner, don't want to go and may think that you are using the dinner as a form of ingratiation tactic. What, if anything, would you do?

PBI question: As a manager, what are some of the actions that you have taken to ameliorate employee/management conflicts and ensure that your subordinates acted as a team? What were the main problems you encountered? How did you address them? What were the outcomes?

As can be seen, the question formats differ yet it is clear that both questions are designed to tap into actions that concern team building and conflict resolution. Importantly, the SI question above depicts a dilemma in which the most appropriate course of action and the exact nature of the skill dimensions being assessed are not overtly apparent to participants. Questions were written by an industrial-organizational psychologist and an executive SME team from the participating organization, with absolute consensus having to be reached regarding the equivalence of the different question types for assessing each performance dimension.

The second issue that should be reiterated is that the scoring keys used to evaluate each managerial skill dimension were identical in each interview and answers to each question were scored immediately after the relevant question was asked. With respect to the 'team orientation' skill dimension presented above, behaviourally anchored scales for both the SI and the PBI ranged from: 1 (is unaware of own impact on others; fails to address, or creates, interpersonal conflict; prefers to work alone; displays little team building initiative, or interpersonal effectiveness or skill) to 5 (demonstrates self-awareness and participatory leadership skills; actively contributes to helping others problem solve; effectively identifies and resolves differences).

Measure of job performance

Eight different behavioural job performance facets and one overall rating of job performance were identified as facets of performance that were critical to effective managerial performance in the participating company. The eight different facets of performance identified were: effectively applies work practices and procedures, plans and problem solves, promotes group cooperation and displays cross-functional management, promotes safety, effectively uses oral communication, promotes customer and supplier relations, engages in personnel development and displays a supportive and personal work style. All job performance ratings were made using the relative percentile method (RPM). The RPM requires that interviewees be evaluated relative to one another and rated on the same 101-point scale (0-100) where 50 = 'average'. Thus, the total scores on the job performance measure could range from 0 to 900. The RPM has shown considerable evidence of validity in past research (Goffin, Gellatly, Paunonen, Jackson, & Meyer, 1996; Goffin & Jelley, 2003; Wagner & Goffin, 1997). In particular, Goffin *et al.* (1996) and Goffin and Jelley (2003) found that the RPM method had superior criterion-related validity compared with absolute performance rating scales. Also, laboratory research by Wagner and Goffin (1997) found that the comparative rating format used in the RPM resulted in higher levels of accuracy than more commonly used absolute rating

formats. Supervisors made job performance ratings approximately one year after the SI and PBI evaluations. The supervisors who rated job performance were not part of the interview assessor teams, nor did they have access to any of the interviewees' scores from the application process. Supervisors were informed that the performance ratings were for research purposes only and would not be retained by the company or made available to others except in aggregate form.

Cognitive ability measures

Cognitive ability was measured using the EAS (Ruch & Ruch, 1980). As stated earlier, the verbal comprehension (test 1; a maximum possible score of 30), verbal reasoning (test 7; a timed test with a maximum possible score of 30) and numerical reasoning (test 6; the maximum possible score is 20) subtests were used as these subtests have been identified as particularly important to managerial performance (Ruch & Ruch, 1980). Alternative-form reliability estimates for individual EAS subtests range from .76 to .91. Because the EAS tests are timed tests, internal consistency reliability estimates (i.e. coefficient alpha) are inappropriate (Ruch & Ruch, 1980). Thus, internal consistency reliability estimates are not reported. Evidence of the criterion-related and construct validity of the EAS is strong (Fitzpatrick, 1984). The scoring procedures used were identical to those outlined in the EAS technical manual (see Ruch & Ruch, 1980).

Work sample measures

Leaderless group discussion

All applicants participated in a leaderless group discussion exercise. Each group consisted of five people. Four managerial skill dimensions were assessed in the LGD: planning and organizing, results orientation, team orientation and oral communication. It is common and accepted that not all assessment centre (AC) exercises permit the measurement of all the managerial skill dimensions derived for a particular job (Thornton & Byham, 1982). To illustrate, the nature of the LGD precluded using it as a measure of the 'coaching' managerial skill dimension. As with the SI and the PBI, two trained, experienced assessors (including the last author) evaluated each performance dimension using 5-point, behaviourally anchored rating scales that ranged from 1 to 5. The LGD has shown strong evidence of validity and inter-rater reliability (e.g. Casio, 1998; Tziner & Dolan, 1982).

In-basket test

The in-basket test is a work simulation that requires participants to respond to various memos, reports and letters that represent problem situations likely to be encountered in a manager's day-to-day activities. Each participant was rated on five managerial skill dimensions that were derived on the basis of job analysis: planning and organizing, coaching, results orientation, team orientation and oral communication. As explained with regard to the LGD, two trained, experienced assessors evaluated the participants on each performance dimension using behaviourally anchored response scales that ranged from 1 to 5. In-basket exercises such as the one currently used have shown high inter-rater reliability and strong evidence of validity (Casio, 1998).

Personality measures

Applicants' achievement orientation and dominance were assessed using the achievement and dominance scales of the Personality Research Form-E (PRF-E; see Table 1; Jackson, 1984) - an inventory widely acknowledged for its excellent predictive and construct validity (Goffin *et al.*, 2000). The achievement and dominance scales from the PRF-E are not only paradigmatic of the achievement orientation and dominance constructs found to underlie managerial performance (see Table 1) but have also been shown to be substantially related to managerial job performance ratings in past research ($r = .45$ and $r = .33$, respectively; Goffin *et al.*, 1996). The PRF-E scales each consisted of 16 true/false items, which resulted in total scores ranging from 0 to 16.

Procedure

Candidates participated in the SI, PBI, LGD and in-basket test as part of a managerial assessment centre (AC). Teams of two assessors each took turns making evaluations in the SI, PBI, LGD and in-basket test, respectively, thus minimizing the potential for carry-over effects on the part of the assessors. For example, if the current approach had not been used, a carry-over effect could have occurred if the same rater saw the same candidate in both interview types and was consciously or subconsciously motivated to provide a similar rating for the applicant in both interviews. In addition, with the exception of the in-basket test and the LGD, candidates participated in the remaining AC exercises in a random order. Specifically, the in-basket test was administered as the first exercise on the first day of the AC and the LGD was administered as the first exercise on the second day for logistical purposes. For the remainder of the second day, candidates participated in the SI, PBI, cognitive ability assessment and personality assessment in a random order. Thus, no order effects could be reasonably expected with respect to comparing SI and PBI performance. Assessors rated each managerial skill dimension in each exercise immediately after the corresponding answers were given. To the best of our knowledge, the assessors did not have any prior contact with the candidates, nor were they aware of any of their application information (e.g. prior experience, previous performance, etc.). Although participants were required to complete the selected cognitive ability and personality measures in conjunction with the AC, only a subset of the current sample completed the personality measures ($N = 84$; see Results).

Results

Preliminary analyses

As has been the norm in past research (e.g. Huffcutt *et al.*, 2001; Pulakos & Schmitt, 1995), we aggregated the scores of the skill dimensions assessed in each interview to form total scores for the SI and the PBI, respectively. Thus, it was of interest to examine whether it was appropriate to aggregate scores on the six managerial skill dimensions measured within the SI and the PBI. A principal axis factor analysis of the six skill dimensions measured in the SI supported a single-factor solution that accounted for 59% of the variance (average loading was .76). Similarly, a principal axis factor analysis of the six skill dimensions measured in the PBI supported a single factor that accounted for 62% of the variance (average loading was .79). Confirmatory factor analyses revealed that a measurement model, wherein the respective performance dimensions of the SI and the PBI load on separate factors, adequately fitted the data (e.g. RMSEA = .06;

absolute value of average standardized residuals = .05). Thus, scores on the six managerial skill dimension scores were averaged to form a single score for each interview type.

Scores on the nine supervisory job performance ratings (i.e. eight facet ratings and one overall rating) also showed a high degree of convergence. Results of a principal axis factor analysis supported a single factor solution that accounted for 63% of the variance with an average loading of .79. Accordingly, the nine performance measures were averaged to form a single job performance composite.

Descriptive statistics

Table 2 shows the means, standard deviations, internal consistency reliabilities and zero-order correlations of the study variables. Unfortunately, inter-rater reliability estimates were not available for SI and PBI ratings in the current data set (the data made available from the participating company simply did not contain all relevant pieces of data). Nevertheless, past research has consistently found that inter-rater reliability estimates for the SI and the PBI are highly similar (e.g. Taylor & Small, 2002). Furthermore, recall that steps were taken to ensure that (a) each interview was developed and scored as similarly as possible, (b) managerial skill dimensions were rated by experienced, trained assessors and (c) the same teams of assessors rated both SI and PBI responses. Therefore, it was expected that inter-rater reliability would be reasonably high and consistent across the SI and PBI.

As shown in Table 2, the internal consistency reliabilities of the study variables were all acceptably high, ranging from .89 to .93 (Nunnally, 1978). Internal consistency estimates could not be calculated for the achievement and dominance scales, as item-level data were not available. However, past research using managers has shown internal consistency estimates of .82 and .75 for the achievement orientation and dominance, respectively (Anderson, 1999). Normative estimates for scales from the PRF-E were not based on those values given in the PRF-E handbook (Jackson, 1984) because those estimates were based on a relatively small college sample ($N = 84$). Because Anderson's data were collected from a relatively large sample of managers ($N = 227$), they were expected to provide a more accurate and relevant estimate for the current research. The reliability values presented for the individual EAS subtests represent alternative-form reliability estimates and were taken from the EAS manual (Ruch & Ruch, 1980). Notably, the correlation between the two interview types was .41 ($p < .01$), indicating that the two interviews only shared 17% of their variance.

In terms of the psychometric comparability of the two interview types, the means and standard deviations of the SI and PBI (see Table 2) do not suggest evidence of extreme scoring (e.g. ceiling/floor effects) or a restricted range. In fact, ancillary analyses showed that the preceding was true of the scores on each of the managerial skill dimensions that constituted the SI and PBI overall scores.

Hypothesis tests

Hypothesis 1 stated that the SI and PBI would each show a positive and significant relation with job performance. In order to test Hypothesis 1, the zero-order correlations of the SI and PBI with job performance were examined (see Table 2). Somewhat contrary to expectations, the PBI showed a significant and medium-sized (Cohen, 1992) relation to job performance of ($r = .32, p < .05$), whereas the SI did not ($r = .09, ns$). Hypothesis 1b stated that the PBI would predict job performance over and above the SI,

Table 2. Means, standard deviations, internal consistency reliabilities and correlations

Variable	1	2	3	4	5	6	7	8	9	10
1. SI	(.89)	—	—	—	—	—	—	—	—	—
2. PBI	.41** (.89)	—	—	—	—	—	—	—	—	—
3. JP	.09 (.93)	.32**	—	—	—	—	—	—	—	—
4. Verb-C	.04 (.85) ^a	.18**	.18**	—	—	—	—	—	—	—
5. Verb-R	.14*	.20**	.20**	.55** (.82) ^a	—	—	—	—	—	—
6. Num-R	.15*	.20**	.20**	.34**	.39** (.87) ^a	—	—	—	—	—
7. LGD	.20*	.28**	.48**	.30**	.22**	.28**	—	—	—	—
8. INB	.13*	.21**	.28**	.12	.29**	.23**	.35** (.90)	—	—	—
9. ACH	.13	.22*	.30**	.06	-.08	-.14	.02	.12 (.82) ^b	—	—
10. DOM	.19	.26*	.43**	.11	.03	.05	.20	.14	.19 (.75) ^b	—
N	157	157	157	157	157	157	157	157	84	84
M	3.16	3.42	595.50	20.03	14.98	12.38	3.16	2.82	12.88	12.49
SD	0.55	0.47	95.33	4.07	5.12	2.72	0.64	.79	2.19	2.71

Note. N = 157 for all correlations save for those involving personality variables (N = 84). SI = situational interview; PBI = past-behaviour interview; JP = job performance; Verb-C = EAS verbal comprehension; Verb-R = EAS verbal reasoning; Num-R = EAS numerical reasoning; LGD = leaderless group discussion; INB = in-basket test; ACH = achievement orientation; DOM = dominance.

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

^aBecause the EAS subtests are timed tests, internal consistency reliability estimates (i.e. coefficient alpha) are inappropriate. Thus, reliability coefficients for the EAS tests represent alternative form reliabilities and are based on the normative estimates from Ruch and Ruch (1980).

^bInternal consistency estimates were taken from a normative managerial sample in Anderson (1999; see Footnote 3).

whereas the SI would not show incremental prediction over the PBI. Hypothesis 1b was tested using hierarchical multiple regression. With regard to the interpretation of multiple regression results, it should be noted that the unadjusted R^2 values are influenced by sample size and the number of predictors used, and represent upwardly biased estimates of the true population R^2 (Pedhazur, 1982). Therefore, the *adjusted* R^2 ($adjR^2$) value was relied upon most heavily as it represents the expected proportion of variance that would be predicted in the population (Pedhazur, 1982). Supportive of Hypothesis 1b, the PBI predicted a significant and medium-sized (Cohen, 1992) proportion of variance in job performance beyond the SI ($\Delta adjR^2 = .10, p < .001$). In contrast, the SI predicted zero variance in job performance beyond the PBI (see Table 3).

Table 3. Hierarchical multiple regressions of job performance on SI and PBI total scores

Squared multiple correlations			
Regression 1: SI entered first and PBI entered second			
	SI	SI + PBI	ΔR^2_{PBI}
Job performance	.01 (.00)	.11 (.10)*	.10 (.10)*
Regression 2: PBI entered first and SI entered second			
	PBI	PBI + SI	ΔR^2_{SI}
Job performance	.11 (.10)*	.11 (.10)*	.00 (.00)

Note. $N = 157$. SI = situational interview; PBI = past-behaviour interview. Values in parentheses represent *adjusted* R^2 values. ΔR^2_{PBI} = change in R^2 due to adding the PBI to the equation. ΔR^2_{SI} = change in R^2 due to adding the SI to the equation.

* $p < .001$.

Hypotheses 2a stated that the SI and PBI would each show positive and significant relations with managerially relevant cognitive ability facets. As shown in Table 2, the PBI showed a consistent pattern of small-sized (Cohen, 1992) but significant positive relations with the cognitive ability facets. Hypothesis 2b stated that the SI would correlate more highly with managerially relevant cognitive ability facets than the PBI. In order to assess the relative degree of association of the SI and the PBI with cognitive ability, the SI and PBI were each regressed on to the five EAS subtests. The respective proportions of variance that the SI and the PBI shared with cognitive ability were then evaluated. Contrary to Hypothesis 2b, multiple regression results showed that cognitive ability shared a significant proportion of variance in the PBI ($adjR^2 = .06, p < .05$) but not in the SI ($adjR^2 = .02, ns$; see Table 4). Tests of the difference between two multiple correlations (see Tabachnick & Fidell, 1995) were deemed unnecessary due to the fact that the PBI had a significant (i.e. non-zero) relation with cognitive ability whereas the SI did not (i.e. the best estimate of the nature of this relation was zero).

Hypothesis 3a stated that the SI and PBI would each show positive and significant relations with the LGD and in-basket test. Although the SI showed a small-sized (Cohen, 1992) but significant positive relation with the LGD ($r = .20, p < .01$), it had a small and non-significant positive relation with the in-basket ($r = .13, ns$). Only the PBI showed a consistent pattern of positive, small-sized (Cohen, 1992) but significant relations with both the LGD and in-basket test ($r = .28, p < .01$ and $r = .21, p < .01$, respectively).

Table 4. Summary of multiple regression results

	SI		PBI	
	β	sr^2	β	sr^2
(a) Cognitive ability regressed on to the SI and PBI ($N = 157$)				
Verb-C	0.14	.02	0.11	.01
Verb-R	0.07	.00	0.08	.01
Num-R	0.12	.01	0.13	.02
R^2	.03(.02)		.07(.06)*	
(b) Work sample tests regressed on to the SI and PBI ($N = 157$)				
LGD	0.23**	.05**	0.43***	.17***
INB	0.12	.02	0.12	.02
R^2	.08(.07)**		.24(.23)***	
(c) Personality regressed on to the SI and PBI ($N = 84$)				
ACH	0.10	.01	0.15	.02
DOM	0.17	.04	0.22*	.05*
R^2	.04(.02)		.09(.07)*	

Note. Verb-C = EAS verbal comprehension; Verb-R = EAS verbal reasoning; Num-A = EAS numerical abilities; Num-R = EAS numerical reasoning; Sym-R = EAS symbolic reasoning; LGD = leaderless group discussion; INB = in-basket test; DOM = dominance. β = standardized beta coefficient. sr^2 = squared semi-partial correlation. Values in parentheses represent *adjusted* R^2 .

* $p < .05$, ** $p < .01$, *** $p < .001$.

Hypothesis 3b stated that the PBI would correlate more highly with the LGD and in-basket tests than the SI. In order to assess the relative degree of association of the SI and the PBI with the work sample measures, the SI and the PBI were each regressed on to the LGD and in-basket test. The respective proportions of variance that the SI and the PBI shared with the work sample measures were then evaluated. As Table 4 shows, the work sample measures shared a significant and small-sized (Cohen, 1992) proportion of variance with the SI ($adjR^2 = .07$, $p < .001$) and a significant and large (Cohen, 1992) proportion of variance with the PBI ($adjR^2 = .23$, $p < .001$). Given both sets of relations were significant, a test of the difference between two correlated multiple correlations was conducted to determine whether the PBI more highly related with the work sample measures than the SI (see Tabachnick & Fidell, 1995). As expected, results showed that the PBI was more strongly related to work sample measures than the SI ($Z = 1.70$, $p < .05$; one-tailed).

Hypothesis 4a stated the SI and PBI would each relate to the managerially relevant personality traits of achievement orientation and dominance. As Table 4 shows, the PBI showed small-sized (Cohen, 1992) but significant positive relations with achievement orientation and dominance ($r = .22$, $p < .05$ and $r = .26$, $p < .05$, respectively), whereas the SI did not. Hypothesis 4b stated that the PBI would correlate more highly with managerially relevant personality traits than the SI. In order to assess the relative degree of association of the SI and the PBI with personality, the SI and the PBI were each regressed on to the achievement orientation and dominance scales and the respective proportions of variance that the SI and the PBI shared with the personality measures

were then evaluated. As shown in Table 4, personality shared a trivial and non-significant proportion of variance with the SI ($adjR^2 = .02$, *ns*) but a significant proportion of variance with the PBI ($adjR^2 = .07$, $p < .05$). As with the results involving cognitive ability, because the PBI had a significant (i.e. non-zero) relation with personality and the SI did not (i.e. the best estimate of the nature of this relation was zero), tests of the difference between two multiple correlations (see Tabachnick & Fidell, 1995) were deemed unnecessary.

Discussion

Results showed that that PBI outperformed the SI format in predicting managerial job performance and that the PBI incrementally predicted performance over the SI, whereas the SI did not incrementally predict the job performance over the PBI. The current results are consistent with past findings showing that the PBI outperforms the SI in predicting job performance for higher-level occupations (e.g. Huffcutt *et al.*, 2001; Pulakos & Schmitt, 1995). However, as stated previously, there were potential methodological flaws in previous research that were more rigorously controlled for in the present study. Given the procedures followed in the current study, it appears that there is little evidence to suggest that the design or scoring of the interviews are plausible explanations for the observed differences in criterion-related validity. The fact that the PBI predicted managerial performance over and above the SI suggests that the PBI captures relevant variance that the SI does not capture and provides further impetus to the investigation of possible construct differences between the SI and the PBI.

With respect to examining construct differences between the SI and PBI, results showed that only the PBI evidenced a consistent pattern of significant relations with the managerially relevant facets of cognitive ability. This finding stands in contrast to past research that showed that the SI is more strongly correlated than the PBI with cognitive ability (Huffcutt *et al.*, 1996). It was suggested earlier that, because the SI involves deriving the 'right' answer to job-related questions, it may be more heavily reliant on deduction and reasoning than the PBI (as per Huffcutt *et al.*, 1996) and may thus be more likely to relate to cognitive ability than the PBI. However, upon further analysis, the 'right' answer to a typical SI dilemma within a managerial sample might be more a function of the general work style or values of the applicant (Taylor & Small, 2002) as opposed to his or her cognitive ability. For example, consider the sample SI question given in the current study or the following SI question taken from Latham and Skarlicki (1995, p. 72):

You become aware that a student is experiencing financial difficulty. The student was told by the Department Chairperson that she could expect a scholarship of \$12,000 to defray costs of living while attending graduate school. However, the student did not get the scholarship. You were in no way involved with the initial communication with the student. What, if anything, would you do?

As can be seen from the typical SI question format, it appears that the 'right' answers to SI questions might be based more on whether or not the work styles or values represented in the applicant's approach to the situation fit with those valued by the company, rather than on the applicant's cognitive ability. In showing the SI to have trivial relations with cognitive ability, the current results provide some evidence to support the preceding view, particularly within a managerial sample. Speculatively, the

fact that the PBI was related to cognitive ability to a greater degree than the SI may have been due to the fact that responses to typical PBI questions within a managerial sample are inherently more complex and verbose than responses to SI questions (Huffcutt *et al.*, 2001; Pulakos & Schmitt, 1995), which may have allowed the applicant more of an opportunity to demonstrate his or her cognitive ability.

In sum, we speculate that with managerial samples, as opposed to lower-level samples, there may be a higher likelihood that the SI questions capture elements of work style or values more than they capture elements of cognitive ability. Furthermore, within managerial samples, the complexity and verbosity of PBI responses are likely to be higher than in lower-level samples and this may tend to increase the cognitive loading of PBI scores.

The fact that the PBI, but not the SI, showed significant relations with each of the work sample measures suggests that the PBI may capture the ability to perform practical managerial behaviours to a greater degree than the SI. Thus, organizations interested in assessing the degree to which a candidate has the tacit knowledge that is required of managerial situations may be better served using the PBI as opposed to the SI.

The fact that the PBI, but not the SI, was significantly related to managerially relevant personality traits of achievement orientation and dominance supports the current hypothesis that the PBI measures applicants' managerially relevant personality traits (i.e. achievement orientation and dominance) to a greater degree than the SI. Thus, having people provide direct evidence of ways that they have behaved in the past can be expected to capture candidates' underlying motivation to a greater extent than their answers to situational dilemmas. Indeed, the typical SI dilemma provides very specific information about situational context factors that may reduce the extent to which a candidate's personality is predictive of their stated response.

One alternative explanation of the differences in criterion-related validity between the PBI and SI might be considered. Although both interview formats were developed according to recommended standards and assessed the same content areas, the use of probes may potentially have differed to some extent between the PBI and the SI. It is conceivable, for example, that interviewers in the PBI probed for more specific facts in the actual experiences of interviewees, whereas interviewers in the SI would have only been able to probe for more specific details of what interviewees thought they would do in the situations. Such a difference in the specific questions constituting the probes may have slightly altered the nature of the respective interview structures. However, structure can be defined in a variety of ways (Campion *et al.*, 1997) and there is no evidence or reason to believe that the overall level or amount of structure would be substantially different across the two formats. Thus, potential differences in the level of structure arising in the PBI and the SI were unlikely to be a contributing factor to differences obtained in their respective criterion-related validity. A more parsimonious explanation for the differences in validity is that the two interview types are assessing different constructs, as demonstrated by the pattern of results in the current study.

In sum, as suggested above, the dilemmas posed by SI questions may be best suited to soliciting examples of applicants' likely work styles or evaluating whether their work values and preferences match those of the hiring company. On the other hand, within managerial samples, the questions posed by the PBI may be better at tapping applicants' performance of job-related behaviours, the extent to which they possess managerially relevant personality traits and cognitive ability.

Limitations and future directions

One limitation of the current work is that the number of years of experience was not measured as a control variable. Indeed, it could be argued that candidates with experience performed better on the PBI than on the SI. However, there was available data on whether candidates had worked, or were currently working for, the participating organization. Ancillary analyses showed that candidates with experience inside the participating organization did not perform any better or worse than candidates without such experience. The preceding result coheres with past research on the SI and the PBI (e.g. Huffcutt *et al.*, 2001; Pulakos & Schmitt, 1995) that has shown that experience is not related to the SI or PBI interview format when job performance is used as the criterion. Nevertheless, given that we did not directly measure years of experience, future studies using a similar paradigm need to replicate these results while controlling for experience.

Some might argue that one potential limitation of the current research is that it dealt exclusively with applicants to managerial positions. Results concerning the criterion-related validity of the SI and the PBI may, therefore, only be generalizable to applicants for managerial positions or jobs of higher complexity. Nevertheless, recall that the current research was explicitly designed to examine validity differences using a sample of applicants to higher-level positions due to the current debate in the literature concerning the predictive efficacy of the SI and the PBI, and the fact that the greatest divergence in SI and PBI validity has been found in such samples (e.g. Huffcutt *et al.*, 2001; Pulakos & Schmitt, 1995).

Consistent with the above, the construct differences currently observed between the SI and the PBI may not be found in samples of lower-level job applicants. In order to make more accurate conclusions regarding the generalizability of the criterion-related and construct validity of the SI and PBI, future research should continue to conduct research using samples drawn from a diversity of occupations using rigorous methodology and stringently developed situational and past-behaviour interviews.

Although the current research advances the literature by improving upon the methodology of past studies and looking at the relation of the SI and the PBI to specific facets of cognitive ability, work samples and personality, future research needs to delve deeper to find out more about variance in job performance that is not explained by structured interviews. Interestingly, the work sample measures tended to share the greatest proportion of variance with the structured interview types in the current study – particularly with respect to the PBI. A possible avenue for future research, therefore, is to examine whether structured interviews assess a series of discrete managerial task or exercise competencies that form the building-blocks of overall managerial competence. Indeed, previous work has suggested it might be fruitful to evaluate whether a large part of the variance in the SI and the PBI may be due to applicants' degree of job-related tacit (i.e. action-oriented) knowledge (Conway & Peneno, 1999; Taylor & Small, 2002).

The finding that cognitive ability within a managerial sample relates more strongly to PBI than SI scores was unexpected. Past studies have generally found cognitive ability to relate more strongly to SI than PBI scores (e.g. Huffcutt *et al.*, 1996); however, we are not aware of research that has examined the cognitive loading of SI versus PBI questions as a function of different job levels. We have provided two rationales (above) as to why PBI questions might be more cognitively loaded than SI questions in a higher-level managerial sample but future research is needed to provide empirical support for these rationales.

Finally, it could be argued that one limitation was that we tested only SI and PBI question 'types' and not actual versions of Latham's original SI (Latham *et al.*, 1980) and Janz's (1982) PBDI. However, as was noted, we attempted to remain faithful to both Latham and Janz in creating the current interviews as they related to the particular jobs under study. In fact, with respect to the SI, we attempted to correct misspecifications in the creation of SI-type questions that arose in past research. Nevertheless, we acknowledge that the current study could be more relevant to comparisons of SI and PBI 'types' of questions as opposed to the original version SI and PBDI questions *per se*.

Practical implications

The results of the current research suggest that the SI and the PBI may not be interchangeable forms of structured interviews. In fact, together with previous work (e.g. Huffcutt *et al.*, 2001; Pulakos & Schmitt, 1995), the current research suggests the SI might not be a suitable selection tool for managerial-level occupations. Indeed, it appears that the SI and the PBI provide different types of information: whereas the PBI may be a better indicator of candidates' cognitive ability, expected managerial task performance and managerially relevant personality traits, the SI may be better suited to assessing an applicant's fit with the organization's preferred work style and values. Organizations should be aware of the nature of these differences and choose or combine the appropriate interview format(s) depending on the nature of the desired information. Future research should continue to investigate differences between the constructs tapped by the SI and PBI so that recommendations regarding the preferred interview type can continue to be fine-tuned.

Conclusions

By improving on past research designs and methodologies, the findings of the current research represent an important contribution to the knowledge base concerning the criterion-related and construct validity of situational versus past behaviour interviews in predicting managerial job performance. In particular, our findings suggest that the SI and PBI may be differentially valid in predicting the performance of managers and that this differential validity may be due to the fact that the two interview types may be assessing different constructs.

Acknowledgements

The preparation of this article was facilitated through grants from the Social Sciences and Humanities Research Council of Canada to Richard D. Goffin.

References

- Anderson, D. W. (1999). *Personality, self-efficacy, and leadership behaviour*. Unpublished doctoral dissertation, University of Western Ontario, London, Ontario, Canada.
- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*, 1-26.
- Bass, B. M. (1990). *Bass and Stogdill's handbook of leadership*. New York: Free Press.
- Becker, T. E., & Colquitt, A. L. (1992). Potential versus actual faking of a biodata form: An analysis along several dimensions of item type. *Personnel Psychology*, *45*, 389-406.

- Blumberg, M., & Pringle, C. D. (1982). The missing opportunity in organizational research: Some implications for a theory of work performance. *Academy of Management Review*, 7, 560-567.
- Bradley, J. P., Nichol, A. M., Charbonneau, D., & Meyer, J. P. (2002). Personality correlates of leadership development in Canadian forces officer candidates. *Canadian Journal of Behavioural Science*, 34, 92-103.
- Campbell, J. P. (1990). Modeling the performance prediction problem in industrial/organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology*, (Vol. 1, pp. 687-732). Palo Alto, CA: Consulting Psychologists Press.
- Campion, M. A., Campion, J. E., & Hudson, J. P. (1994). Structured interviewing: A note on incremental validity and alternative question types. *Journal of Applied Psychology*, 79, 998-1002.
- Campion, M. A., Pursell, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. *Personnel Psychology*, 50, 655-702.
- Cascio, W. F. (1998). *Applied psychology in human resource management*. Upper Saddle River, NJ: Prentice-Hall.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.
- Conway, J. M., & Peneno, G. M. (1999). Comparing structured interview question types: Construct validity and applicant reactions. *Journal of Business and Psychology*, 13, 485-506.
- Cook, K. W., Vance, C. A., & Spector, P. E. (2000). The relation of candidate personality with selection-interview outcomes. *Journal of Applied Social Psychology*, 30, 867-885.
- Craik, K. H., Ware, A. P., Kamp, J., O'Reilly, C., Staw, B., & Zedeck, S. (2002). Explorations of construct validity in a combined managerial and personality assessment programme. *Journal of Occupational and Organizational Psychology*, 75, 171-193.
- Day, A. L., & Carroll, S. A. (2003). Situational and patterned behaviour description interviews: A comparison of their validity, correlates, and perceived fairness. *Human Performance*, 16, 25-47.
- Donovan, J. J., Dwight, S. A., & Hurtz, G. M. (2003). An assessment of the prevalence, severity, and verifiability of entry-level applicant faking using the randomised response technique. *Human Performance*, 16, 81-106.
- Fitzpatrick, R. (1984). Employee aptitude survey tests. In D. J. Keyser & R. C. Sweetland (Eds.), *Test critiques* (Vol. 1, pp. 266-273). Kansas City, MO: Westport.
- Goffin, R. D., Gellatly, I. R., Paunonen, S. V., Jackson, D. N., & Meyer, J. P. (1996). Criterion validation of two approaches to performance appraisal: The behavioural observation scale and the relative percentile method. *Journal of Business and Psychology*, 11, 23-34.
- Goffin, R. D., & Jelley, R. B. (2003, April). *Facilitating social comparisons in performance appraisal*. Paper Presented at the Annual Conference of the Society for Industrial-Organizational Psychology, Orlando.
- Goffin, R. D., Rothstein, M. G., & Johnston, N. (1996). Personality testing and the assessment center: Incremental validity for managerial selection. *Journal of Applied Psychology*, 81, 746-756.
- Goffin, R. D., Rothstein, M. G., & Johnston, N. G. (2000). Predicting job performance using personality constructs: Are personality tests created equal? In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 249-264). Boston, MA: Kluwer.
- Gough, H. G. (1990). Testing for leadership with the California psychological inventory. In K. E. Clark & M. B. Clark (Eds.), *Measures of leadership* (pp. 355-379). West Orange, NJ: Leadership Library of America.
- Hough, L. M. (1992). The big five personality variables-construct confusion: Description versus prediction. *Human Performance*, 5, 139-155.
- Hough, L. M. (1998). Personality at work: Issues and evidence. In M. Hakel (Ed.), *Beyond multiple choice: Evaluating alternatives to traditional testing for selection* (pp. 131-159). Hillsdale, NJ: Erlbaum.

- Hough, L. N., Barge, B., & Kamp, J. (2001). Assessment of personality, temperament, vocational interests, and work outcome preferences. In J. P. Campbell & D. J. Knapp (Eds.), *Exploring the limits of personnel selection and classification* (pp. 111-154). Hillsdale, NJ: Erlbaum.
- Howard, A., & Bray, D. W. (1988). *Managerial lines in transition*. New York: Guilford.
- Huffcutt, A. I., Roth, P. L., & McDaniel, M. A. (1996). A meta-analytic investigation of cognitive ability in employment interview evaluations: Moderating characteristics and implications for incremental validity. *Journal of Applied Psychology, 81*, 459-473.
- Huffcutt, A. I., Weekley, J., Weisner, W. H., DeGroot, T., & Jones, C. (2001). Comparison of situational and behaviour description interview questions for higher-level positions. *Personnel Psychology, 54*, 619-644.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 92*, 72-98.
- Hunter, J. E., & Schmidt, F. L. (1989). Meta-analysis: Facts and theories. In M. Smith & I. T. Robertson (Eds.), *Advances in selection and assessment* (pp. 203-216). Oxford: Wiley.
- Jackson, D. N. (1984). *Manual for the personality research form*. Port Huron, MI: Research Psychologists Press.
- Jackson, D. N., Paunonen, S. V., Fraboni, M., & Goffin, R. D. (1996). A five-factor versus six-factor model of personality structure. *Personality and Individual Differences, 20*, 33-45.
- Jackson, D. N., Peacock, A. C., & Holden, R. R. (1982). Professional interviewers' trait inferential structures for diverse occupational groups. *Organizational Behaviour and Human Performance, 29*, 1-20.
- Janz, T. (1982). Initial comparisons of patterned behaviour description interviews versus unstructured interviews. *Journal of Applied Psychology, 67*, 577-580.
- Janz, T. (1989). The patterned behaviour description interview: The best prophet of the future is the past. In R. W. Eder & G. R. Ferris (Eds.), *The employment interview: Theory, research, and practice* (pp. 158-168). Thousand Oaks, CA: Sage.
- Janz, T., Hellervik, L., & Gilmore, D. C. (1986). *Behaviour description interviewing: New, accurate, cost-effective*. Boston: Allyn & Bacon.
- Jelf, G. S. (1999). A narrative review of post-1989 employment interview research. *Journal of Business and Psychology, 14*, 25-58.
- Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology, 87*, 765-780.
- Latham, G. P., Saari, L. M., Pursell, E. D., & Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology, 65*, 569-573.
- Latham, G. P., & Skarlicki, D. (1995). Criterion-related validity of the situational and patterned behaviour description interviews with organizational citizenship behaviour. *Human Performance, 8*, 67-80.
- Latham, G. P., & Sue-Chan, C. (1996). A legally defensible interview for selecting the best. In R. S. Barrett (Ed.), *Fair employment strategies in human resource management* (pp. 134-143). Westport, CT: Quorum Books/Greenwood Publishing.
- Latham, G. P., & Sue-Chan, C. (1999). A meta-analysis of the situational interview: An enumerative review of reasons for its validity. *Canadian Psychology, 40*, 56-67.
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw Hill.
- Pedhazur, E. J. (1982). *Multiple regression in behavioural research*. New York: Holt, Reinhart & Winston.
- Posthuma, R. A., Morgeson, F. P., & Campion, M. A. (2002). Beyond employment interview validity: A comprehensive narrative review of recent research and trends over time. *Personnel Psychology, 55*, 1-81.
- Pulakos, E. D., & Schmitt, N. (1995). Experienced-based and situational interview questions: Studies of validity. *Personnel Psychology, 48*, 289-308.
- Ruch, F. L., & Ruch, W. W. (1980). *Employee aptitude survey technical report*. Los Angeles, CA: Psychological Services.

432 Henryk T. Krajewski et al.

- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262-274.
- Steinberg, R., & Shapiro, S. (1982). Sex differences in personality traits of female and male MBA students. *Journal of Applied Psychology*, *67*, 306-310.
- Stogdill, R. M. (1974). *Handbook of leadership: A survey of theory and research*. New York: Free Press.
- Tabachnick, B. G., & Fidell, L. S. (1995). *Using multivariate statistics*. New York: HarperCollins.
- Taylor, P. J., & Small, B. (2002). Asking applicants what they would do versus what they did do: A meta-analytic comparison of situational versus past-behaviour employment interview questions. *Journal of Occupational and Organizational Psychology*, *75*, 277-294.
- Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology*, *44*, 703-742.
- Thornton, G. C., & Byham, W. C. (1982). *Assessment centers and managerial performance*. New York: Academic Press.
- Tziner, A., & Dolan, S. (1982). Validity of an assessment center for identifying future female officers in the military. *Journal of Applied Psychology*, *67*, 728-736.
- Wagner, S. H., & Goffin, R. D. (1997). Differences in accuracy of absolute and comparative performance appraisal methods. *Organizational Behaviour and Human Decision Processes*, *70*, 95-103.

Received 16 June 2004; revised version received 10 May 2005

Copyright of Journal of Occupational & Organizational Psychology is the property of British Psychological Society and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.