

## Personality Testing and the Assessment Center: Incremental Validity for Managerial Selection

Richard D. Goffin and Mitchell G. Rothstein  
University of Western Ontario

Norman G. Johnston  
Toronto, Ontario, Canada

The authors investigated the relative validity of personality testing and the assessment center (AC) in order to evaluate the incremental validity of both assessment methods in the prediction of managerial performance. Managers ( $N = 68$ ) from a forestry products organization were evaluated with an AC and selected traits on the Personality Research Form (D. N. Jackson, 1984). Performance and promotability served as criteria. Overall, personality was found to be at least equivalent to the AC in predicting performance. Personality testing resulted in significant incremental validity over that of the AC in the prediction of performance, but the converse was also true. This suggested that personality and the AC assess different domains, with each uniquely and significantly predicting performance. Promotability was not significantly predicted. Practical implications and future directions were discussed.

Recent research (e.g., Christiansen, Goffin, Johnston, & Rothstein, 1994; Gellatly, Paunonen, Meyer, Jackson, & Goffin, 1991; Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; Tett, Jackson, & Rothstein, 1991) has generally supported the existence of a significant relationship between personality measures and job performance. Additionally, several studies have dealt with the extent to which personality may improve the prediction of performance beyond that of other selection methods. Using a sample of accountants, Day and Silverman (1989) found that job-relevant personality traits significantly predicted job performance, even when cognitive ability was partialled from the relationship. Schippman and Prien (1989) achieved similar results by studying a sample of managers. Furthermore, McHenry, Hough, Toquam, Hanson, and Ashworth (1990) found personality traits to have incremental validity over cognitive ability in predicting three out of the five job performance criteria for U.S. Army personnel. Similar results were re-

ported by Rosse, Miller, and Barnes (1991), who studied a sample of health care workers, Arneson, Millikin-Davies, and J. Hogan (1993), who studied a sample of insurance claims adjusters, and Gellatly et al. (1991), who studied a sample of first-line managers. Conversely, Cortina, Doherty, Schmitt, Kaufman, and Smith (1992) failed to find evidence of the incremental predictive validity of personality over cognitive ability in a sample of state police recruits. However, the cognitive ability test used by Cortina et al. (1992) was unusual in that it included a panel interview component that may have overlapped with the domain of personality.

Whereas the literature suggests that personality tests have incremental validity over cognitive ability, there is little research on the incremental validity of personality over other common selection methods. The current study addressed this deficiency by using personality testing and the assessment center (AC) method in a sample of managers.

### Personality and the AC Method

Meta-analytic results have suggested that ACs possess respectable levels of validity (Gaugler, Rosenthal, Thornton, & Bentson, 1987). Nonetheless, given the relatively high cost of ACs, it is not surprising that numerous researchers have investigated the possibility that other selection devices may have comparable levels of validity (e.g., Borman, 1982; Bray & Grant, 1966; Drakeley, Herriot, & Jones, 1988; Dulewicz & Fletcher, 1982; Hinrichs, 1969, 1978; J. Hogan & Zenke, 1986; Pynes & Bernardin, 1989; Thornton & Byham, 1982; Turnage & Muchinsky, 1984; Tziner & Dolan, 1982; Wollowick & McNamara, 1969). In particular, both Borman (1982)

---

Richard D. Goffin and Mitchell G. Rothstein, Department of Psychology, University of Western Ontario, London, Ontario, Canada; Norman G. Johnston, Independent Practice, Toronto, Ontario, Canada.

This research was supported by Grant 410-92-1857 from the Social Sciences and Humanities Research Council of Canada. Preliminary results from this study were presented at the Annual Meeting of the Society for Industrial and Organizational Psychology, San Francisco, April 1993.

Correspondence concerning this article should be addressed to Richard D. Goffin, Department of Psychology, Social Science Centre, University of Western Ontario, London, Ontario, Canada N6A 5C2. Electronic mail may be sent via Internet to goffin@sscl.uwo.ca.

and Tziner and Dolan (1982) compared personality testing and AC in terms of their ability to predict training performance. J. Hogan and Zenke (1986) made a similar comparison using actual job performance, whereas Bray and Grant (1966), as well as Hinrichs (1969), compared the two with salary progress as the criterion. Additionally, Hinrichs (1978) compared the validity of personality with that of AC scores on the basis of attained position level. In Bray and Grant, Borman, Hinrichs (1978), J. Hogan and Zenke, and Tziner and Dolan, the validity of the AC appeared to exceed that of personality testing. However, in Bray and Grant, Hinrichs (1978), and J. Hogan and Zenke, the validity coefficients for the personality tests were nonetheless significant. Unfortunately, none of the above studies dealt with whether personality tests contributed significant incremental predictive variance over and above that of the AC. Despite the fact that personality tests and ACs have been shown to predict job performance (e.g., Gaugler et al., 1987; Tett et al., 1991), on average, relations between personality scores and AC scores tend to be small (Borman, 1982; Bray & Grant, 1966; Crawley, Pinder, & Herriot, 1990; Hinrichs, 1978; Tziner & Dolan, 1982). Therefore, it is quite conceivable that if used in combination, personality and the AC may both uniquely and significantly predict job performance. Furthermore, from Russell and Domm's (1995) finding that the criterion-related validity of ACs is largely attributable to task- or role-based inferences rather than trait-based inferences, it seems likely that relevant trait variance, as measured with a personality test, may predict criterion variance beyond what is attributable to the AC.

An additional limitation of some existing research concerns the manner in which personality was measured. A modern construct-oriented test development strategy may be advantageous because of resistance to faking and higher validity (Goffin, Rothstein, & Johnston, 1995; Goffin & Woods, 1995; Holden & Jackson, 1985; Wiggins, 1973). The Personality Research Form (PRF; Jackson, 1984) used in this study is a promising exemplar of the construct-oriented strategy. Because much of the existing research on the use of personality testing versus the AC was conducted without construct-oriented measures, our study was able to assess whether such measures have more potential for adding unique predictive variance.

A further problem with some of the existing research concerns the use of exploratory research designs in which personality measures were used with little or no a priori rationale. Tett et al. (1991) found that studies undertaken with a priori hypotheses concerning expected relations between specific personality traits and job performance yielded a much higher mean validity than studies with exploratory designs. Thus, to the extent that an a priori rationale in the choice of personality traits was lacking in some of the earlier research comparing the predictive

validity of personality measures and ACs, the contribution of personality to predicting job performance may have been underestimated.

Finally, an underlying consideration pertaining to much of the research involving ACs has been construct validity (e.g., Klimoski & Brickner, 1987; Russell & Domm, 1995). Specifically, two questions have figured prominently in this regard: (a) To what extent is the criterion-related and convergent validity of AC exercises greater than that of AC dimensions (i.e., assuming that exercise scores are aggregated across dimensions, whereas dimension scores are aggregated across exercises (e.g., Bycio, Alvares, & Hahn, 1987) and (b) To what extent is the criterion-related validity of ACs attributable to trait-based inferences, as opposed to task-based, role-congruency inferences (Russell & Domm, 1995). Investigation of the second concern has suggested that trait-based inferences in AC ratings may be inferior to task- or role-based inferences in terms of criterion-related validity. Unfortunately, direct investigation of the task-based versus role-based distinction in AC scores requires a modification of the traditional AC strategy (see Russell & Domm, 1995) that was precluded in the current work. However with respect to the first concern, we acknowledged that scores might reasonably be derived on the basis of AC dimensions or AC exercises; therefore, we considered both types of scores in our analysis.

This study expanded the current knowledge base through the use of a priori hypotheses and hierarchical regression (Pedhazur, 1982) to assess the incremental validity of personality over the AC. Moreover, we capitalized on developments in the measurement of personality through the use of a construct-oriented test (Jackson, 1984). Finally, extrapolating from Bycio et al. (1987) among others, we considered composite AC exercise scores in addition to AC dimensional scores.

The specific research questions of interest in this study were as follows:

Do relevant personality trait measures provide significant amounts of incremental predictive variance over and above that of the AC?

What is the relative importance of the AC and relevant personality trait measures in predicting performance criteria?

We chose the PRF scales of dominance, exhibition, and achievement as possible predictors of managerial performance on the basis of past research (for a review, see R. T. Hogan, 1991). These three traits have been strongly and reliably linked to criteria of general managerial performance and success (e.g., Brenner, 1982; Hanson & T. W. Harrell, 1985; M. S. Harrell, T. W. Harrell, McIntyre, & Weinberg, 1977; R. T. Hogan, 1991; Howard & Bray, 1988; Wertheim, Widom, & Wortzel, 1978). These traits have also been found to constitute the

stereotype of an effective manager (Steinberg & Shapiro, 1982) and are related to academic performance in graduate business education (Rothstein, Paunonen, Rush, & King, 1994). In addition, Barrick and Mount (1991) classified personality traits into the Big Five dimensions and found in a meta-analysis of personality-job performance relations that two dimensions, namely, extraversion and conscientiousness, showed relations with managerial performance. Barrick and Mount support the research cited above because dominance and exhibition are two of the primary traits comprising extraversion, and achievement is an important component of conscientiousness. We chose to use primary personality traits in the present study, as there is some indication that such measures have increased predictive ability over broad personality dimensions (Tett, Jackson, Rothstein, & Reddon, 1994).

## Method

### *Participants and Procedure*

Participants in this study were 68 AC candidates in a large forestry products organization. Their mean age was 36 ( $SD = 7.5$ ) with a range from 23 to 57 years. Candidates were identified for participation in the AC through their applications for management positions in the organization. For each vacancy, applicant resumés were screened to identify 5 candidates for further assessment: One of them was subsequently selected for the position. A total population of 315 candidates for management positions were evaluated over a 3-year period, from which our sample of 68 selected managers was derived. Thus, the selection ratio was .215. The data were part of a larger project evaluating the predictive validity of an AC. All 68 participants were selected for promotion to supervisory positions. Performance ratings by supervisors collected 1-year after the AC evaluation were available. Personality testing was used extensively in this organization, primarily as part of the selection process. Almost everyone in this study was administered the personality test concurrently with the AC, such that results from both assessment methods were used in the selection decisions. However, for administrative reasons, a small proportion of candidates (approximately 10 of those in the current study) were not given the personality test until shortly after they were selected. We tracked participants by identification numbers and their names were not known to us.

### *Specification of the Content Domain*

The content domains for the AC predictors and the performance criteria were determined by total quality management (TQM) principles implemented throughout the organization. Nine job performance criteria (provided below) for first-level supervisory positions were identified through a job analysis and defined in terms of TQM principles. Similarly, AC dimensions were determined on the basis of these TQM principles. The predictor constructs that comprised the AC dimensions are described in the Appendix. Personality predictors were selected

on the basis of the research cited above that indicated traits found to be related consistently to managerial performance.

### *Assessment Center*

These data came from a 2-day managerial selection program used by a large forestry products organization. The AC was developed by an industrial-organizational psychologist and was operated according to conventional standards (i.e., Thornton & Byham, 1982). With the AC dimensions in mind (see Appendix), exercises were developed to allow candidates multiple opportunities to demonstrate the desired behavioral tendencies. In particular, an effort was made to ensure that each dimension was sampled in at least two (and typically more than two) exercises. Each exercise was tailor-made to accurately assess as many of the AC dimensions as possible with a job-relevant format. Development of each exercise and the associated scoring manuals was a painstaking process involving consideration of input from company personnel, multiple drafts and revisions, and multiple pretesting trials.

The AC consisted of (a) a leaderless group discussion (LGD) involving 5 candidates at a time engaged in a problem-solving discussion; (b) a situational interview (e.g., Latham, Saari, Pursell, & Campion, 1980) requiring candidates to describe their preferred course of action in response to several hypothetical scenarios; (c) a past-behavior interview (e.g., Janz, 1982) in which participants were to provide specific examples of past situations wherein they demonstrated certain job-relevant capabilities; (d) an in-basket; and (e) a simulation requiring the candidate to play the role of a manager providing performance feedback. Trained assessors administered the above exercises, recorded and scored observations, and participated in consensus meetings during which overall AC ratings were obtained. Observations were made and evaluated against the six AC dimensions described in the Appendix. The breakdown of which dimensions were measured by which exercises is provided in Table 1. Not all dimensions were amenable to measurement through each exercise; hence, the LGD, in-basket, and simulation did not afford measurement of willingness to learn. Additionally, the LGD did not allow measurement of coaching, and the simulation did not measure team orientation.

In all exercises except the past-behavior interview, observed behavior scores and judgment scores were available for each of the measured dimensions. Observed behavior scores were obtained by summarizing assessors' ratings of the presence of key effective behaviors for a given dimension, whereas judgment scores required the assessor to rate the candidate on a 5-point scale, where 3 = *minimally acceptable performance*. There were no observed behavior scores obtained for any of the dimensions measured with the past-behavior interview, whereas the in-basket lacked observed behavior scores for the communication dimension. This occurred because the past-behavior interview and the communication dimension in the in-basket did not readily afford evaluation by rating the presence of key effective behaviors. The in-basket differed from other exercises by also providing a second set of rating scores that were based on the assessors' evaluations of the degree to which more general types of behaviors relevant to the respective dimensions (e.g., requests additional information, delegates appropriately, etc.) were descriptive of the candidates. These general behavior scores were

Table 1  
Breakdown of Assessment Center (AC) Scores by Dimension and Exercise

Dimension	Exercise score												Consensus rating	Dimension total
	LGD		Situational interview		Past-behavior interview		In-basket		Simulation		Consensus rating			
	Observed	Judgment	Observed	Judgment	Observed	Judgment	Observed	General	Observed	Judgment				
Planning and organizing	1	1	1	1	1	1	1	1	1	1	1	1	1	11
Coaching	0	0	1	1	1	1	1	1	1	1	1	1	1	9
Results orientation	1	1	1	1	1	1	1	1	1	1	1	1	1	11
Willingness to learn	0	0	1	1	1	1	0	0	0	0	0	0	0	4
Team orientation	1	1	1	1	1	1	1	1	1	1	1	1	1	9
Communication	1	1	1	1	1	1	0	0	1	1	1	1	1	10
Exercise total	8	8	12	12	6	6	14	14	8	8	6	6	6	

Note. Dimension total = number of AC scores available for each dimension, and exercise total scores include all available for each exercise or total number of consensus ratings. Because unit weighting was used throughout, all totals equal the maximum possible score for the respective composite scores. LGD = leaderless group discussion.

used along with the other scores to compute the composite scores described below. In addition to the exercise-based scores, assessors provided consensus judgments for each candidate on each of the six dimensions, as well as an overall AC consensus rating. The row marginals in Table 1 provide the total number of scores available for each of the six AC dimensions, whereas the column marginals show the number of scores available for each of the five exercises and the consensus ratings.

Four different sets of AC scores were considered for use in the current analyses. First, following a traditional strategy, the overall AC consensus rating could be used as the sole predictor. Second, the seven consensus ratings (one for each of the six dimensions plus the overall consensus rating) could all be used as predictors. Third, a composite score for each of the six dimensions could be derived by unit weighting and summing the appropriate scores as shown in Table 1 (e.g., the planning and organizing dimension composite would aggregate 11 scores). Fourth, from Bycio et al. (1987), six composite scores could also be derived by unit weighting and summing the respective scores for each of the five exercises (e.g., as shown in Table 1, the LGD composite would be comprised of eight scores), plus the consensus scores could be summed to yield another exercise score. Given that unit weighting was used throughout, the respective column and row marginals in Table 1 indicate the maximum possible scores for each of the composite exercise scores and dimensional scores. From the principle of aggregation (Paunonen, 1984), it was expected that the latter two sets of scores would hold the most promise in the prediction of criteria.

*Personality Measures*

Personality was measured with the PRF-Form E (Jackson, 1984). The PRF is a self-report personality test consisting of 22 scales with 16 items (see Appendix). As discussed above, the scales of dominance, achievement, and exhibition were chosen a priori as being relevant to the target job. The PRF is one of the most highly cited personality inventories (Mitchell, 1983); its excellent psychometric properties have been repeatedly acknowledged (e.g., for reviews, see Anastasi, 1972; Kelly, 1972), as has its validity in personnel selection applications (e.g., Gellatly et al., 1991; Goffin et al., 1995; Rothstein et al., 1994). Following Jackson's (1984) instructions, PRF raw scores were converted to T scores ( $M = 50, SD = 10$  in the population).

*Performance Measures*

Eleven supervisory ratings, including nine different performance dimensions determined through job analysis, an overall performance rating, and a promotability rating were used as criterion variables. The nine performance dimensions were work practices and procedures, planning and problem solving, monitoring and controlling, group cooperation and cross-functional management, promoting safety, communication, personnel development, customer and supplier relations, and personal work style. Supervisors were clearly informed that these ratings were designated for research purposes only, and, as such, they would not influence the ratees' careers in any manner or be made available to others except in a summarized form, as in this article. All ratings were made with the Relative Percentile Method (RPM), a comparative performance appraisal ap-

proach that has been validated in previous research (Goffin, Gellatly, Paunonen, Jackson, & Meyer, 1996; Wagner & Goffin, 1996). As explained by Goffin et al. (1996), the RPM approach requires that ratings be made on 101-point scales, where 50 = *average*. For each performance dimension, all of the rateses of a given rater are considered relative to one another and rated on the 0-to-100 scale. Goffin et al. suggested that the RPM approach has superior criterion-related validity to that of the behavioral observation scale, whereas Wagner and Goffin's (1996) laboratory research found that the comparative rating format used in the RPM tends to result in higher levels of accuracy than does the more common absolute or individual rating format used in many performance ratings.

The supervisors provided the performance ratings 1 year after the ratees' AC evaluations. At no time did the supervisors see the specific scores from the AC or PRF.

## Results

### *Preliminary Analyses*

A visual examination of the relations among the performance dimensions and the rating of overall performance suggested a high degree of similarity. Analogously, a principal components analysis revealed that a single component explained 62% of the variance. Therefore, a total performance measure was formed on the basis of an equally weighted composite of the nine performance dimensions and the rating of overall performance. The rating of promotability was kept separate because, rationally, promotability was seen as distinct from performance, and, empirically, research has shown promotability and performance to be differentially predictable by the AC (Turnage & Muchinsky, 1984). The maximum possible score was 1,000 on the total performance composite and 100 on the promotability composite.

Prior to addressing our main research questions, we had to decide which set of AC scores to use as predictors in all the analyses. As discussed in the Method section, it was expected that the composite dimensional scores and the composite exercise scores were the most promising. Preliminary multiple regression analyses (not reported in this article) showed that the composite dimensional scores predicted more variance in both the total performance and the promotability criteria than did the composite exercise scores, although the differences in variance predicted were small. Moreover, the predictive advantage of the composite dimensional scores over the exercise scores held true for the unadjusted and shrunken  $R^2$  values; thus, they were chosen for use in all the current analyses.<sup>1</sup>

Table 2 contains descriptive statistics, reliability estimates (coefficient alpha), and uncorrected intercorrelations among the predictors (personality and AC dimensions) and performance criteria. From PRF normative data (Jackson, 1984), the dominance, achievement, and exhibition means corresponded respectively to

the 74th, 74th, and 55th percentiles in the population at large but were similar to means reported by Rothstein et al. (1994) for a large sample ( $N = 450$ ) of graduate business administration students who took the PRF for research purposes. Standard deviations of PRF scores were not markedly different from the population value of 10. All computed reliabilities were acceptable, but reliability could not be computed for the promotability rating because it was a single-item measure.

Relations between the AC predictors and the criteria ranged from .02 to .40, and most reached statistical significance. Criterion relations involving the personality traits ranged from .02 to .45, and all but one reached statistical significance.

Relations among the six AC composite dimension scores were significant and ranged from .52 to .87. Although significant, the relations among the three personality traits were lower, ranging from .21 to .46, whereas intercorrelations between personality traits and AC scores were very low and typically nonsignificant, ranging from  $-.02$  to .24. The distinctiveness of the AC and personality domains was further investigated through a principal components analysis of the six AC predictors together with the three PRF scales. Two components with eigenvalues greater than one were extracted and subjected to a varimax rotation. A clear two-component solution explaining 71% of the variance with near-perfect simple structure arose such that all AC scores loaded highly, positively, and solely on the first component; the three PRF traits loaded highly, positively, and solely on the second component.

### *Hierarchical Regression Analyses*

Hierarchical regression analyses were carried out to study the degree to which the AC dimensions and the per-

<sup>1</sup> It is unlikely that any other choice of scores would have improved the prediction of performance over the composite dimensional scores. For example, in additional analyses not reported here, it was determined that the overall AC consensus rating was a less effective predictor than were the composite dimensional scores. Similarly, we carried out multiple regressions not reported here, with the seven AC consensus ratings (one consensus rating for each of the six dimensions plus the overall AC consensus rating) as predictors. The consensus ratings were less effective in the prediction of performance and promotability than were the currently used composite dimensional scores. Additionally, from results of the principal components analysis in the *Preliminary Analyses* section, we formed a single unit-weighted composite comprised of all the AC composite dimensional scores and one further unit-weighted composite comprised of the PRF dominance, achievement, and exhibition scores. Use of these macro unit-weighted composite scores failed to improve the prediction of criteria beyond that shown by the shrunken  $R^2$  values in Table 3, nor did it change the relative importance of the AC and personality in the prediction of criteria.

Table 2  
Zero-Order Correlations and Descriptive Statistics for Assessment Center (AC) Dimensions and Personality Research Form (PRF) Scales

Variable	1	2	3	4	5	6	7	8	9	10	11
AC predictors											
1. Planning	.81										
2. Coaching	.76**	.80									
3. Results orientation	.87**	.81**									
4. Willingness to learn	.73**	.52**	.75								
5. Team orientation	.77**	.72**	.61**	.85							
6. Communication	.82**	.70**	.77**	.66**	.75						
				.59**	.71**	.83					
PRF predictors											
7. Dominance	.06	.17	.16	-.02	.13	.05	.82				
8. Achievement	.08	.05	.07	.24*	.10	.10	.38**	.67			
9. Exhibition	-.02	.11	.04	-.00	-.01	.11	.46**	.21*	.73		
Criteria											
10. Total performance <sup>a</sup>	.37**	.36**	.36**	.18	.40**	.33**	.45**	.33**	.30**	.93	
11. Promotability	.24*	.26*	.23*	.02	.24*	.16	.27*	.02	.21*	.60**	
<i>M</i>	6.29	5.12	6.31	2.45	5.15	6.40	55.82	56.26	51.32	650.34	62.21
<i>SD</i>	0.98	0.91	0.94	0.49	0.84	0.76	7.69	7.05	8.59	108.98	18.99

Note. *N* = 68. Alpha reliabilities appear on the diagonal. Unit-weighted composite of nine performance dimensions and overall performance rating.  
\* *p* < .05 (one-tailed). \*\* *p* < .01 (one-tailed).

sonality traits could predict the two criteria. In the first series of hierarchical analyses, reported in Table 3, the six AC dimensional composites were entered into the equation first and followed by the three personality traits. Total performance and promotability in turn served as the criterion. However, the order of entry of variables in a multiple regression analysis has a potent effect (Pedhazur, 1982). Therefore, to gain a fuller appreciation of the relative importance of AC scores and personality traits in predicting criteria, in a second series of regressions, we entered the three personality traits into the equation first, followed by the six AC dimensional scores (see Table 3).

A forced entry approach was used in these regressions to ensure that predictors remained in the equation regardless of the significance level of their beta weights. Although it would have been possible to use only the significant predictors, this was not done for two reasons. First, significance tests of *R*<sup>2</sup> values from regression analyses in which the selection of predictors is determined by their predictive ability in the sample at hand are biased in favor of giving significant results (Pedhazur, 1982). Second, our goal was to assess how well the AC scores and personality scores performed as intact sets of predictors, that is, sets that were chosen on the basis of a priori considerations regarding the target job.

With respect to the first series of regressions, wherein the AC scores were entered first, the *R*<sup>2</sup> value achieved in Step 1 was significant in the case of the total performance criterion but nonsignificant in the case of promotability (see Table 3). The addition of the personality traits in Step 2 resulted in a significant *R*<sup>2</sup> increment in the prediction of total performance but not in the case of promotability. In the second series of hierarchical regressions, wherein personality was entered into the equation first, the obtained *R*<sup>2</sup> value in Step 1 was significant with respect to total performance but not with respect to promotability. In Step 2, the AC dimensional scores resulted in a significant increment in *R*<sup>2</sup> in the case of the total performance criterion but not the promotability criterion. The combined set of personality and AC predictors resulted in a significant *R*<sup>2</sup> when predicting total performance but not promotability.

It is noteworthy that three personality traits predicted a slightly higher proportion of performance variance in the sample than did the six AC dimensions (see unadjusted *R*<sup>2</sup> values in Table 3). It is typically expected that more predictors will result in more predicted sample variance. The fact that three personality traits predicted at least as much sample variance as six AC dimensions suggests that the personality traits are on average at least as effective in predicting performance as are the AC dimensions.

The unadjusted sample estimate of *R*<sup>2</sup> tends to be an upwardly biased estimate of the true population *R*<sup>2</sup>

Table 3  
*Hierarchical Regressions*

Squared multiple correlations			
Regression 1: Assessment center (AC) dimensions first and Personality Research Form (PRF) traits second			
Performance criterion	AC	AC + PRF	$\Delta R^2_{PRF}^a$
Total performance	.21* (.13)	.43** (.34)	.22** (.21)
Promotability	.15 (.07)	.22 (.10)	.07 (.03)
Mean $R^2$	.18 (.10)	.33 (.22)	.15 (.12)
Regression 2: PRF traits entered first and AC dimensions second			
Performance criterion	PRF	PRF + AC	$\Delta R^2_{AC}^b$
Total performance	.25** (.21)	.43** (.34)	.18* (.13)
Promotability	.09 (.05)	.22 (.10)	.13 (.05)
Mean $R^2$	.17 (.13)	.33 (.22)	.16 (.09)

*Note.* AC dimensions (planning and organizing; coaching; results orientation; willingness to learn; team orientation; and communications skills) and PRF (see Jackson, 1984, and the Appendix of this article) traits (dominance, achievement, and exhibition) were each entered as predictors. Values in parentheses signify shrunken  $R^2$ 's (see Pedhazur, 1982, p. 148).  $N = 68$ .

<sup>a</sup> Change in  $R^2$  caused by adding the PRF traits to the equation. <sup>b</sup> Change in  $R^2$  caused by adding the AC dimensions to the equation.

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

(Pedhazur, 1982). Thus it is expected that the more predictors one uses, the more the sample  $R^2$  will shrink when prediction takes place in the population at large. As suggested by the shrunken  $R^2$  values (see Table 3) in the population, the personality traits alone would be expected to explain an average of 21% of the variance in total performance, whereas the corresponding value for the AC dimensions was 13%. Furthermore, the shrunken estimates of incremental variance (see Table 3) indicate that in the population the personality traits would be expected to improve prediction of total performance over that of the AC dimensions by 21%. Our complementary series of analyses suggest that the improvement in prediction of the AC scores over the personality traits is 13% in the population (see Table 3).

### *Correction for Range Restriction*

It was reasoned that because the overall AC consensus rating was always considered in making selection decisions, range restriction might cause all relationships involving AC predictors to be attenuated, thereby giving an artifactual predictive advantage to personality testing.<sup>2</sup> To empirically assess this possibility, we investigated the effect of correcting for range restriction in the regression analyses. As recommended by Bobko (1995) and used by McHenry et al. (1990), we applied the 1943 Lawley (see Lord & Novick, 1968) multivariate range restriction correction to the entire matrix of correlations (see Table 2). This is the recommended procedure for range restriction corrections when dealing with multivariate analyses such as multiple regression which simultaneously involve an

entire matrix of correlations (see Bobko, p. 111). Then we repeated all of the regression analyses reported above using the corrected matrix. However, the range-corrected  $R^2$ 's were virtually identical to the corresponding uncorrected values, and the few observed differences in  $R^2$ 's were trivial, the largest being .02, and most being .00.<sup>3</sup> Given that no consequential differences between uncorrected and corrected values were observed, it was reasoned that range restriction must not have had an appreciable effect on these analyses and that the uncorrected regression results, as reported above, were veridical. Perhaps the most compelling explanation as to why range restriction did not appreciably affect these results is simply that selection decisions involving the AC were based on the overall AC consensus rating, not the AC dimensional scores we used here, even though the latter are better predictors. Thus, the effect of range restriction on the current predictors was indirect (see Bobko, 1995) and in this case trivial. Moreover, as described in the Method section, AC candidates were assessed in groups of 5 and from each group of 5, 1 candidate was typically selected. By selecting the best candidate out of each group of 5 rather than applying straight top-down selection from the entire pool of applicants, the overall effect of range restriction was drastically reduced. Additionally, all AC

<sup>2</sup> We thank an anonymous reviewer for pointing out this potential problem.

<sup>3</sup> The detailed results of the regression analyses that were conducted with the range-corrected matrix are available upon request from the corresponding author.

candidates had been prescreened from their resumé, meaning that the unselected sample ( $n = 315$ ) was to some extent already limited in range. Finally, it is well-known that top-scoring candidates are also the most likely to turn down a job offer because of other more enticing career options (Cascio, 1991). In such cases, the position was offered to a lower scoring candidate thereby reducing range restriction. As a result, the mean and standard deviation of the overall AC consensus rating, which could range from unity to 6, did not differ drastically across the selected ( $n = 68$ ,  $M = 4.04$ ,  $SD = .87$ ) and unselected ( $n = 315$ ,  $M = 3.97$ ,  $SD = .97$ ) samples.

### Discussion

In this research, we investigated the extent to which relevant personality measures provide significant amounts of incremental predictive variance beyond that of the AC. Also of interest was the relative importance of the AC and personality traits in predicting performance. With respect to the first issue, it was found that personality traits significantly improve the prediction of a performance composite beyond the prediction afforded by the AC. Nevertheless, the AC also afforded significant incremental predictive variance over personality testing in the case of the composite performance measure. Neither the AC nor the personality traits significantly predicted a promotability rating. One limitation of the promotability rating was its reliance on a single-item scale of unknown reliability, whereas the total performance composite was comprised of 10 items ( $\alpha = .93$ ).

With regard to the relative importance of the two types of predictors, these data suggest that if used exclusively, relevant personality traits could be expected to explain proportions of variance in performance, which would be at least as high in the population as those attributable to the AC dimensions (see shrunken  $R^2$  values for total performance in Table 3). Nonetheless, it is arguable that the AC would be disadvantaged in comparisons of shrunken  $R^2$ s because the shrinkage formula results in greater downward adjustment when more predictors are used (Pedhazur, 1982) and the AC used more predictors than did the personality domain. However, this argument fails when comparing unadjusted  $R^2$  values (see Table 3). There were twice as many AC predictors as personality predictors. This means that the degree of upward bias present in the unadjusted AC  $R^2$  values is likely to be greater than that of the respective personality  $R^2$ s. Even though the AC had six predictors and the personality domain had three unadjusted  $R^2$  values attained using total performance as the criterion were larger for personality than for the AC.

As delineated by Cascio (1991) among others, criterion-related validity is only one of the factors affecting the utility of a selection method. However, assuming that

number of hires, selection ratio,  $SD_y$ , and average tenure of new hires are held constant, whereas the criterion-related validity of personality predictors is at least as high as that of the AC and testing costs are far less (i.e., estimated per candidate cost of the current AC, \$1,350, was 270 times that of the personality test), higher utility is expected for personality testing.

In addition to the issue of utility, personnel selection specialists must also be cognizant of participant reactions to testing procedures. Extrapolating from Steiner and Gilliland (1996), it appears likely that the typical AC may garner more positive reactions than the typical personality test. Nonetheless, it is unlikely that participant reactions to well-developed, appropriately applied personality tests will be negative (e.g., Goffin & Woods, 1995).

It is noteworthy that the pattern of relations in Table 2 suggests a marked lack of correlation between personality and AC scores, a pattern that was strongly reinforced through principal components analysis. The lack of relatedness of the personality and AC domains has also been supported by the pattern of relations seen in Borman (1982), Bray and Grant (1966), Crawley, Pinder, and Herriot (1990), Hinrichs (1978), and Tziner and Dolan (1982). This consistent pattern, combined with our finding of significant incremental validity for both personality and AC scores over one another when predicting performance, suggests that personality and AC scores may sample different domains, which in turn may predict relatively independent aspects of the domain of performance. Further corroboration of this perspective and some evidence as to the substantive nature of the predictive domain assessed by the AC was provided by Russell and Domm (1995), who found that the criterion-related validity of AC scores is largely independent of trait-based (e.g., personality-related) inferences. Rather, the predictiveness of the AC appeared to be attributable to task- or role-based inferences. Thus, trait-based inferences, as captured by relevant PRF scales, are likely to predict criterion variance beyond that predicted by the AC. Accordingly, in our analyses, the best prediction was achieved through a combination of personality and the AC. As shown in Table 3, the proportion of variance in total performance predicted by this combination was quite high (shrunken  $R^2 = .34$ ), substantially exceeding that of cognitive ability tests when the appropriate uncorrected values from J. E. Hunter and R. F. Hunter (1984) are considered. Consequently, our findings bode well for the use of personality testing as an adjunct to the AC and vice versa.

Our finding that personality traits performed at least as proficiently as the AC in the prediction of performance contrasts with previous research in this area wherein the validity of the AC appeared to exceed that of personality testing (Borman, 1982; Bray & Grant, 1966; Hinrichs,



1978; J. Hogan & Zenke, 1986; Tziner & Dolan, 1982). We believe that this is due to the limitations of most previous research with respect to the manner in which personality was measured or the research strategy used. In particular, sophisticated construct-oriented personality measures such as the PRF (Jackson, 1984) were not always used in the past. The construct-oriented test development strategy may be advantageous in the measurement of personality (Goffin et al., 1995; Goffin & Woods, 1995; Holden & Jackson, 1985; Wiggins, 1973). The current study adds further weight to this contention. In addition, we followed an a priori hypothesis testing research strategy in which personality measures were identified as high potential predictors on the basis of published conceptual and empirical linkages to the target job (e.g., Brenner, 1982; Hanson & T. W. Harrell, 1985; M. S. Harrell et al., 1977; R. T. Hogan, 1991; Howard & Bray, 1988; Rothstein et al., 1994; Steinberg & Shapiro, 1982; Wertheim et al., 1978). Some of the previous research comparing the validity of personality testing and the AC had adopted a more exploratory strategy in the choice of personality traits, which, on the basis of Tett et al. (1991), is likely to have led to an underestimate of the true predictive potential of personality.

In summary, the present study dealt with the incremental validity of personality testing over the AC in predicting managerial performance. We found that the predictive validity of personality tended to be at least as high as that of the AC. Perhaps more importantly, the domains encompassed by our personality and AC predictors were largely orthogonal, and each contributed uniquely and significantly to the prediction of performance. Therefore, personality testing was extremely useful as an adjunct to the AC and vice versa. However, we must caution that the generalizability of these results remains to be investigated. In particular, it is important to determine if the current results replicate with different job types or similar jobs in different organizations and industries, different personality measures, and different ACs.

## References

- Anastasi, A. (Reviewer). (1972). Personality Research Form [Test No. 123]. In O. K. Buros (Ed.), *The Seventh Mental Measurements Yearbook* (pp. 297-298). Highland Park, NJ: Gryphon Press.
- Arneson, S., Millikin-Davies, M., & Hogan, J. (1993). Validation of personality and cognitive measures for insurance claims examiners. *Journal of Business and Psychology, 7*, 459-473.
- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1-26.
- Bobko, P. (1995). *Correlation and regression: Principles and applications for industrial/organizational psychology*. New York: McGraw-Hill.
- Borman, W. C. (1982). Validity of behavioral assessment for predicting military recruiter performance. *Journal of Applied Psychology, 67*, 3-9.
- Bray, D. W., & Grant, D. L. (1966). The assessment center in the measurement of potential for business management. *Psychological Monographs, 80* (17, Whole No. 625).
- Brenner, D. C. (1982). Relationship of education to sex, managerial status, and the managerial stereotype. *Journal of Applied Psychology, 67*, 380-383.
- Bycio, P., Alvares, K. M., & Hahn, J. (1987). Situational specificity in assessment center ratings: A confirmatory factor analysis. *Journal of Applied Psychology, 72*, 463-474.
- Cascio, W. F. (1991). *Costing human resources: The financial impact of behavior in organizations*. Boston: PWS-Kent.
- Christiansen, N. D., Goffin, R. D., Johnston, N. G., & Rothstein, M. G. (1994). Correcting the 16PF for faking: Effects on criterion-related validity and individual hiring decisions. *Personnel Psychology, 47*, 847-860.
- Cortina, J. M., Doherty, M. L., Schmitt, N., Kaufman, G., & Smith, R. G. (1992). The "Big Five" personality factors in the IPI and MMPI: Predictors of police performance. *Personnel Psychology, 45*, 119-140.
- Crawley, B., Pinder, R., & Herriot, P. (1990). Assessment center dimensions, personality and aptitudes. *Journal of Occupational Psychology, 63*, 211-216.
- Day, D. A., & Silverman, S. B. (1989). Personality and job performance: Evidence of incremental validity. *Personnel Psychology, 42*, 25-36.
- Drakeley, R. J., Herriot, P., & Jones, A. (1988). Biographical data, training success and turnover. *Journal of Occupational Psychology, 61*, 145-152.
- Dulewicz, V., & Fletcher, C. (1982). The relationship between previous experience, intelligence and background characteristics of participants and their performance in an assessment centre. *Journal of Occupational Psychology, 55*, 197-207.
- Gaugler, B. B., Rosenthal, D., Thornton, G. C., III, & Bentson, C. (1987). Meta-analysis of assessment center validity [Monograph]. *Journal of Applied Psychology, 72*, 493-511.
- Gellatly, I. R., Paunonen, S. V., Meyer, J. P., Jackson, D. N., & Goffin, R. D. (1991). Personality, vocational interests, and cognitive predictors of managerial job performance and satisfaction. *Personality and Individual Differences, 12*, 221-231.
- 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 Behavioral Observation Scale and the Relative Percentile Method. *Journal of Business and Psychology, 11*, 37-47.
- Goffin, R. D., Rothstein, M. G., & Johnston, N. (1995, May). *A comparative evaluation of a construct-driven and an empirically derived personality test in personnel selection*. Paper presented at the Annual Conference of the Society for Industrial and Organizational Psychology [Division 14 of the American Psychological Association], Orlando, FL.
- Goffin, R. D., & Woods, D. M. (1995). Using personality testing for personnel selection: Faking and test-taking inductions. *International Journal of Selection and Assessment, 3*, 227-235.
- Hanson, B. A., & Harrell, T. W. (1985). Predictors of business success over two decades: An MBA longitudinal study

- (Research Rept. No. 788). Stanford, CA: Stanford University Graduate School of Business.
- Harrell, M. S., Harrell, T. W., McIntyre, S. H., & Weinberg, C. B. (1977). Predicting compensation among MBA graduates 5 and 10 years after graduation. *Journal of Applied Psychology*, 62, 636-640.
- Hinrichs, J. R. (1969). Comparison of real-life assessments of management potential with situational exercises, paper-and-pencil ability tests, and personality inventories. *Journal of Applied Psychology*, 53, 425-432.
- Hinrichs, J. R. (1978). An 8-year follow-up of a management assessment center. *Journal of Applied Psychology*, 63, 596-601.
- Hogan, J., & Zenke, L. L. (1986). Dollar-value utility of alternative procedures for selecting school principals. *Educational and Psychological Measurement*, 46, 935-945.
- Hogan, R. T. (1991). Personality and personality measurement. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 2, pp. 873-919). Palo Alto, CA: Consulting Psychologists Press.
- Holden, R. R., & Jackson, D. N. (1985). Disguise and the structured self-report assessment of psychopathology: I. An analogue investigation. *Journal of Consulting and Clinical Psychology*, 53, 211-222.
- Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities [Monograph]. *Journal of Applied Psychology*, 75, 581-595.
- Howard, A., & Bray, D. W. (1988). *Managerial lines in transition*. New York: Guilford Press.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- Jackson, D. N. (1984). *Personality Research Form manual*. Port Huron, MI: Research Psychologists Press.
- Janz, T. (1982). Initial comparisons of patterned behavior description interviews versus unstructured interviews. *Journal of Applied Psychology*, 67, 577-580.
- Kelly, E. L. (Reviewer). (1972). Personality Research Form [Test No. 123]. In O. K. Buros (Ed.), *The Seventh Mental Measurements Yearbook* (pp. 298-301). Highland Park, NJ: Gryphon Press.
- Klimoski, R. J., & Brickner, M. (1987). Why do assessment centers work? The puzzle of assessment center validity. *Personnel Psychology*, 40, 243-260.
- Latham, G. P., Saari, L. M., Pursell, E. D., & Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology*, 65, 422-427.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology*, 43, 335-354.
- Mitchell, J. V. (Ed.) (1983). *Tests in print III*. Lincoln, NE: Buros Institute.
- Paunonen, S. V. (1984). Optimizing the validity of personality assessments: The importance of aggregation and item content. *Journal of Research in Personality*, 18, 411-431.
- Pedhazur, E. J. (1982). *Multiple regression in behavioral research*. New York: Holt, Rinehart & Winston.
- Pynes, J. E., & Bernardin, H. J. (1989). Predictive validity of an entry-level police officer assessment center. *Journal of Applied Psychology*, 74, 831-833.
- Rosse, J. G., Miller, H. E., & Barnes, L. K. (1991). Combining personality and cognitive ability predictors for hiring service-oriented employees. *Journal of Business and Psychology*, 5, 431-445.
- Rothstein, M. G., Paunonen, S. V., Rush, J. C., & King, G. A. (1994). Personality and cognitive ability predictors of performance in graduate business school. *Journal of Educational Psychology*, 86, 516-530.
- Russell, C. J., & Domm, D. R. (1995). Two field tests of assessment centre validity. *Journal of Occupational and Organizational Psychology*, 68, 25-47.
- Schippman, J. S., & Prien, E. P. (1989). An assessment of the contributions of general mental ability and personality characteristics to management success. *Journal of Business and Psychology*, 3, 423-437.
- Steinberg, R., & Shapiro, S. (1982). Sex differences in personality traits of female and male MBA students. *Journal of Applied Psychology*, 67, 306-310.
- Steiner, D. D., & Gilliland, S. W. (1996). Fairness reactions to personnel selection techniques in France and the United States. *Journal of Applied Psychology*, 81, 134-141.
- 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.
- Tett, R. P., Jackson, D. N., Rothstein, M., & Reddon, J. R. (1994). Meta-analysis of personality-job performance relations: A reply to Ones, Mount, Barrick, and Hunter (1994). *Personnel Psychology*, 47, 157-172.
- Thornton, G. C., & Byham, W. C. (1982). *Assessment centers and managerial performance*. New York: Academic Press.
- Turnage, J. J., & Muchinsky, P. M. (1984). A comparison of the predictive validity of assessment center evaluations versus traditional measures in forecasting supervisory job performance: Interpretive implications of criterion distortion for the assessment paradigm. *Journal of Applied Psychology*, 69, 595-602.
- 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. (1996, April). *Differences in accuracy of individual and comparative performance appraisal methods*. Paper presentation at the Annual Conference of the Society for Industrial and Organizational Psychology [Division 14 of the American Psychological Association], San Diego, CA.
- Wertheim, E. G., Widom, C. S., & Wortzel, L. H. (1978). Multivariate analysis of male and female professional career choice correlates. *Journal of Applied Psychology*, 63, 234-242.
- Wiggins, J. S. (1973). *Personality and prediction*. Reading, MA: Addison-Wesley.
- Wolowick, H. B., & McNamara W. J. (1969). Relationship of the components of an assessment center to management success. *Journal of Applied Psychology*, 53, 348-352.

## Appendix

## Measures Used in This Study

## Description of Assessment Center Dimensions

Dimension	Description
Planning and organizing	Establishing and implementing a course of action to achieve some business objective. Being able to set goals and priorities, identify and initiate goal-relevant actions, allocate and coordinate resources, and monitor.
Coaching	Guiding, directing, and helping subordinates to develop, and facilitating their professional growth.
Results orientation	Showing initiative, perseverance, and commitment to achieving quality in work performance.
Willingness to learn	Showing commitment to further self and to professional development and an openness to and active involvement in learning.
Team orientation	Being willing and able to work effectively in a work group and exhibiting interpersonal awareness, participatory decision making, a desire to help, conflict management skills, and team-building skills.
Communication	Exchanging information effectively in oral and/or written form, including the ability to persuade and listen.

## Personality Research Form Scale Names and Descriptions (Form E)

Scale name	Description of high scorer
Abasement	Accommodates others by willingness to be helpful and assume subordinate position.
Achievement	Aspires to accomplish difficult tasks; maintains high standards.
Affiliation	Enjoys being with friends and people in general; is sociable.
Aggression	Enjoys combat and argument; easily annoyed; likes to "get even."
Autonomy	Abhors restraints, confinement, or restrictions; is independent.
Change	Likes new and different experiences; dislikes routine.
Cognitive structure	Prefers certain, definite information; dislikes ambiguity.
Defence	Takes offense easily; does not accept criticism; is defensive.
Desirability	Describes self in terms judged as desirable, presents favorable picture of self.
Dominance	Attempts to influence or direct other people; enjoys the role of leader.
Endurance	Is willing to work long hours; does not give up easily.
Exhibition	Wants to be the center of attention; enjoys having an audience.
Harm avoidance	Dislikes exciting or dangerous activities; seeks to maximize personal safety.
Impulsivity	Tends to act on the spur of the moment; is emotionally volatile.
Infrequency	Responds to test items in implausible or pseudo-random fashion.
Nurturance	Gives sympathy and comfort to those in need; assists others whenever possible.
Order	Keeps personal effects and surroundings neat and organized.
Play	Enjoys games, sports, social activities, and other amusements.
Sentience	Notices smells, sounds, sights, tastes, and textures; has a hedonistic view of life.
Social recognition	Is concerned about reputation and what other people think.
Succorance	Frequently seeks sympathy, protection, love, and advice.
Understanding	Values logical thought; is intellectually curious and reflective.

*Note.* This excerpt from the Personality Research Form is from Jackson (1984, pp. 6-7). Reproduced with permission from the publisher Sigma Assessment Systems (D/B/A Research Psychologists Press prior to 1989; Port Huron, MI) and the author. Copyright © 1984, 1989.

Received October 26, 1995  
Revision received March 20, 1996  
Accepted July 10, 1996 ■