Incremental Validity of a Situational Judgment Test

Michael A. McDaniel
Virginia Commonwealth University

Amy Powell Yost
Capital One Financial Corp.

Mark H. Ludwick
Capital One Financial Corp.

Richard L. Hense
Capital One Financial Corp.

Nathan S. Hartman
Virginia Commonwealth University

Paper presented at the 19th Annual Conference of the Society for Industrial and Organizational Psychology. Chicago. April, 2004.

The authors wish to thank Malcom James Ree for the RangeJ program (Johnson & Ree, 1994) and his instructions in its use.

Situational judgment tests (SJTs) present job applicants with work-related situations and possible responses to those situations. These tests have received increased research interest in recent years (e.g., Chan & Schmitt, 1997; Clevenger, Pereira, Wiechmann, Schmitt, & Schmidt Harvey, 2001; McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001; McDaniel & Nguyen, 2001; Motowidlo, Dunnette, & Carter, 1990). Research in this area has continued because these instruments have been found to predict job performance reasonably well, have less adverse impact than cognitive measures, and they have substantial face validity. The criterion-related validity of SJTs has been evaluated in several primary studies (Chan & Schmitt, 1997; Hanson & Borman, 1989; Motowidlo et al., 1990; Smith & McDaniel, 1998); in a recent meta-analysis, the typical criterion-related validity was estimated at .34 (McDaniel et al., 2001). The construct validity of SJTs also has been assessed (Beaty Jr. & Howard, 2001; Leaman & Vasilopoulos, 1997; Vasilopoulos, Reilly, & Leaman, 2000; Weekly & Jones, 1997, 1999). Meta-analytic results on the construct validity of SJTs have shown that they primarily assess conscientiousness, emotional stability, agreeableness (McDaniel, Hartman, & Grubb, 2002; McDaniel & Nguyen, 2001), and cognitive ability (McDaniel et al., 2001, McDaniel et. al., 2003). This construct heterogeneity may have practical value for applicant screening in that SJTs may account for more of the variance in predicting job performance than do unidimensional personality or cognitive ability tests.

Incremental validity

Clevenger et al. (2001) found SJTs to have incremental validity above that provided by cognitive ability, conscientiousness, job experience, and job knowledge. O'Connell, McDaniel, Grubb, Hartman, and Lawrence (2002) found that SJTs demonstrated small degrees of incremental validity in predicting both contextual and task

performance after cognitive ability and personality were controlled. Chan and Schmitt (2002) found that their SJT not only predicted three job performance dimensions (task, motivational contextual, and interpersonal contextual performance), it also provided incremental validity over the combination of cognitive ability, personality, and job experience. Chan and Schmitt (2002) stated that SJTs have been shown to measure stable individual difference attributes that do not completely overlap with measures of job experience, cognitive ability, and the Big Five personality traits, thus giving the SJTs potential for incremental prediction above both cognitive ability and personality. The current study adds to the cumulating literature assessing the incremental validity of a SJT over cognitive and personality predictors.

Method

Sample

The sample consisted of 384 professionals in a Fortune 500 financial services firm. The sample members were participating in a concurrent validation study.

Measures

Cognitive ability. The cognitive test was a measure of numerical reasoning that had been used as a selection test for the screening of these incumbents; their scores were retrieved from archival records. Because the cognitive test had been used in operational selection and the other study predictor measures had not, the validity of the cognitive test and its correlations with other variables are attenuated due to range restriction. The personality and SJT scores suffer from indirect range restriction to the extent that they are correlated with the cognitive test.

The Big 5. As part of the concurrent validation study, participants completed the ipsative version of the Occupational Personality Questionnaire (OPQ-32i; SHL, 1999).

For the purposes of this paper, the OPQ scales were configured to the Big 5 variables consistent with the OPQ manual (OPQ-32; SHL, 1999). Scores on the neuroticism scale were multiplied by -1 to yield a scale where high scores indicated emotional stability.

The SJT. The SJT was developed for the concurrent validation effort and asked respondents to rate the effectiveness of response options for each situational stem using a six-item scale. For the purposes of this paper, the SJT was empirically-keyed. To key the items, we correlated the SJT response options with the criterion in a keying sample. The keying sample size was approximately 669 individuals varying somewhat from item to item due to small amounts of missing data. (These individuals were not included in subsequent incremental validity analyses.) An item was retained for the key if its correlation with the criterion was .10.or higher (or -.10 or lower). Items negatively correlated with the criterion were inversely keyed. Retained items with criterion correlations above .15 were double weighted. The 23 items that met the scoring criteria were summed to obtain the SJT score. High scores on the resulting scale indicated favorable job performance. The validity in the keying sample was .42 which is, of course, an inflated estimate of the operational validity due to capitalization on chance.

The Criterion Measure. The criterion measure used for this paper was the mean of managerial (supervisory) performance level ratings across 15 competencies. These ratings were completed by managers as part of the concurrent validation study and were used for research purposes only. Managers were asked to rate their associate's "level of proficiency consistently demonstrated on the job for each competency." Competency levels ranged from 1 (lowest) to 4 (highest) with Low, Medium, and High ratings within each level -- resulting in a 12-point scale for the ratings (Level 1/Low to Level 4/High).

Results

Table 1 presents the correlation matrix for all variables based on the incremental validity sample. Correlations above the diagonal comprise the observed correlation matrix. Correlations below the diagonal comprise the correlation matrix corrected for multivariate range restriction.

Table 2 reports incremental validity results separately for the observed data and the correlations corrected for multivariate range restriction. Multiple regression was used to analyze the incremental contribution of predictors. Analyses were conducted both with the observed correlations and with the correlations corrected for multivariate range restriction. The RangeJ program (Johnson & Ree, 1994) was used to correct correlations for multivariate range restriction, These calculations were limited by the available data. There were no U.S. national norms available for the OPQ ipsative version and thus we did not know the unrestricted variance of the scales. The SJT measure had never been given to an applicant population and thus we did not know its unrestricted variance. However, we did know the unrestricted variance of the *g* test and used that information in the multivariate range restriction calculations. Because most of the unrestricted variances were not known, the estimated unrestricted correlation matrix may vary some from the actual unrestricted correlation matrix. Since the results of the analyses for the observed and unrestricted correlation matricies are similar, we will focus the discussion on the results from the observed distribution.

The first analysis examined the incremental validity of the SJT over cognitive ability using multiple regression. The addition of the SJT raised the R from .08 to .24. The second analysis examined the incremental validity of the SJT over and above

cognitive ability and the Big 5. The addition of the SJT raised the validity of the battery from .22 to .30

Secondary Analysis. To better understand the relatively low correlation between *g* and the performance criterion (r = .08), we examined the correlations between the cognitive measure and the individual performance level ratings that made up the composite performance rating used for this study. We also reported the correlations between the SJT and the individual performance level ratings as well as the Big 5 and the individual performance level ratings. Results are presented in Table 3.

Discussion

Table 1 presents the intercorrelations among the variables. A few observations are noteworthy. First, cognitive ability is not a strong predictor of the criterion used for this paper (*r* = .08). The correction for range restriction raised the correlation to .09. This correlation of the cognitive ability measure with job performance is inconsistent with previous research (Schmidt & Hunter, 1998). One plausible explanation for this result may be the limited breadth of the cognitive test; it is a measure of numerical reasoning and relatively narrow in its assessment of cognitive ability relative to tests of general mental ability. The correlation may also indicate that the criterion may not tap primarily the cognitive demands of the job, although the personality correlations with the criterion were not large either. The criterion was a composite of ratings on a variety of cognitive and non-cognitive competencies. When the cognitive ability measure (a numerical reasoning test) was correlated with criterion ratings pertaining specifically to cognitive performance dimensions (e.g., Analyzes Information, Identifies and Solves Problems, Applies Integrated Decision Making), the correlations tended to be higher (see Table 3). Another likely explanation of the relatively low correlation of the cognitive measure with

overall performance is that the criterion measure was computed across a variety of different jobs with large differences in cognitive ability requirements. If the same analysis were done only with those jobs having greater cognitive requirements, we would expect to see a higher correlation (Gandy & McDaniel, 1986). Another limitation that may have attenuated the correlation for the cognitive predictor in this study is that the measure used was a stanine score. Use of the stanine (as opposed to the test raw score) increased the N available for analysis due to more missing data in the raw scores, but it did reduce the variance in the predictor. The same analysis conducted with the raw score predictor might have yielded a higher correlation with the overall criterion measure.

The correlation of the SJT with the cognitive test (r = .15) is also low compared to previous research. Observed correlations between *g* and SJTs of about .30 are common (McDaniel et al., 2003). Also, the correlations between the SJT and personality are also quite low compared to typical values. For example, the observed correlation between Conscientiousness and the SJT is -.03 where values are typically about .28 (McDaniel et al., 2003). Likewise the correlations between the SJT and Agreeableness and Emotional Stability are -.04 and -.11 compared to typical values of about .25 and .31 (McDaniel et al., 2003). Finally, a few of the correlations among the Big 5 measures are moderately large and negative. This is likely due to the ipsative nature of the OPQ.

SJTs can be scored in many ways. For this study, the SJT was scored empirically. Other scoring procedures may yield keys that are not highly correlated with this one. Such keys might produce different results concerning incremental validity.

The relatively low criterion-related validity of *g* and the relatively low correlations among the predictors increased the likelihood of incremental prediction for the SJT in this study. The SJT showed substantial incremental prediction over *g* alone,

raising the multiple R from .08 to .24. However, due to the weak predictive power of *g*, the composite of *g* and the SJT (r = .24) was only slightly higher than the correlation of SJT alone with the criterion (r = .23). The SJT also showed substantial incremental prediction over *g* and the Big 5, raising the prediction from .22 to .30.

The substantial degree of incremental validity for the SJT will likely overestimate the incremental validity of such tests in other settings for two reasons. First, the *g* measure had low validity in this study relative to past research (Schmidt & Hunter, 1998). Second, the situational judgment test used in this study had smaller correlations with *g*, Conscientiousness, Emotional Stability, and Agreeableness compared to past research (McDaniel, et al., 2003).

References

Beaty Jr., J. C., & Howard, M. J. (2001). *Constructs measured in situational judgment tests designed to assess management potential and customer service potential.* Paper presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

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

Chan, D., & Schmitt, N. (2002). Situational judgment and job performance, *Human Performance*, 15 233-254.

Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N., & Harvey, V. S. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology, 86,* 410-417.

Hanson, M. A., & Borman, W. C. (1989). *Development and construct validation of a situational judgment test of supervisory effectiveness for first-line supervisors in the U.S. Army.* Paper presented at the Symposium conducted at the 4th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, Georgia.

Gandy, J. A., & McDaniel, M. A. (1986). Large samples are better, aren't they?: Pooling subjects across organizations. In H. R. Hirsh (Chair) *New Examinations of Meta-analysis.* Symposium presented at the 94th Annual Convention of the American Psychological Association, Washington, DC.

Johnson, J.T. & Ree, M.J. Ree (1994). RangeJ: A Pascal program to compute the multivariate correction for range restriction. Educational and Psychological Measurement, 54, 693-695.

Leaman, J. A., & Vasilopoulos, N. L. (1997). *Development and validation of the U.S. immigration officer applicant assessment*. Washington, D.C.: U.S. Immigration and Naturalization Service.

McDaniel, M.A., Hartman, N. S.& Grubb III, W. L. (2003, April). *Situational judgment tests, knowledge, behavioral tendency, and validity: A meta-analysis*. Paper presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychology. Orlando.

McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Predicting job performance using situational judgment tests: A clarification of the literature. *Journal of Applied Psychology, 86,* 730-740.

McDaniel, M. A., & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment, 9*, 103-113.

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

O'Connell, M. S., McDaniel, M. A., Grubb, W. L., III, Hartman, N. S., & Lawrence, A. (2002). *Incremental validity of situational judgment tests for task and contextual performance.* Paper presented at the 17th Annual Conference of the Society of Industrial Organizational Psychology, Toronto, Ontario.

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. *Journal of Applied Psychology, 124*, 262-274.

SHL (1999). *OPQ32 Manual and User's Guide*. Surrey, UK. SHL Group plc.

Smith, K. C., & McDaniel, M. A. (1998). *Criterion and construct validity evidence for a situational judgment measure.* Paper presented at the 13th Annual Convention of the Society for Industrial and Organizational Psychology, Dallas.

Vasilopoulos, N. L., Reilly, R. R., & Leaman, J. A. (2000). The influence of job familiarity and impression management on self-report measure scale scores and response latencies. *Journal of Applied Psychology, 85*, 50-64.

Weekly, J. A., & Jones, C. (1997). Video-based situational testing. *Personnel Psychology, 50*, 25-49.

Weekly, J. A., & Jones, C. (1999). Further studies of situational tests. *Personnel Psychology, 52*, 679-700

Table 1. Intercorrelation of Cognitive Ability, SJT, and the Big 5.

	g	SJT	Conscientiousness	Agreeableness	Emotional Stability	Extroversion	Openness	Job Performance
g	-	**.15**	**.13**	**-.20**	**-.10**	-.03	.03	.08
SJT	.18	-	-.03	-.04	**-.11**	.05	**.10**	**.23**
Conscientiousness	.16	-.02	-	-.05	**-.18**	**-.31**	**-.30**	.03
Agreeableness	-.23	-.05	-.06	-	**.11**	**-.18**	-.09	-.04
Emotional Stability	-.11	-.11	-.19	.11	-	**.13**	.01	-.06
Extroversion	-.03	.05	-.31	-.17	.13	-	**.24**	**.16**
Openness	.04	.10	-.30	-.09	.01	.24	-	.09
Job Performance	.09	.24	.04	-.04	-.06	.16	.09	-

Note. N = 384. Correlations above the diagonal are the observed correlations. Observed correlations in bold are statistically significant ($p \leq .05$). Correlations below the diagonal are corrected for multivariate range restriction.

Table 2. Incremental Validity Results for Observed Correlations and Corrections Corrected for Multivariate Range Restriction

	Observed Data Multiple R	Range Restriction Corrected Data Multiple R
g	.08	.09
SJT	.23	.24
Big 5	.21	.21
g + SJT	.24	.24
g + Big 5	.22	.22
g + Big 5 + SJT	.30	.31

Table 3. Correlation of SJT and *g* with performance rating dimensions.

Performance Dimension	SJT	*g*	Conscientiousness	Emotional Stability	Openness to Experience	Extroversion	Agreeableness
Analyzes Information	.08	.20**	.19**	-.09	.09	.00	-.12*
Generates and Pursues Ideas	.13*	.04	.02	-.05	.13**	.16**	-.04
Develops and Shapes Strategies	.27**	.09	.05	-.08	.13*	.09	-.10
Identifies and Solves Problems	.18**	.10	.08	-.08	.11*	.11*	-.03
Applies Integrated Decision Making	.19**	.09	.08	-.09	.04	.10*	-.02
Focuses on Strategic Priorities	.21**	.08	0	-.03	.11*	.14**	-.08
Treats Others with Respect	.11*	.02	-.02	.00	.00	.00	.13*
Collaborates with Others	.15**	.08	-.04	-.04	.02	.15**	0
Organizes and Manages Multiple Tasks	.18**	.06	.09	-.12*	0	.13*	-.04
Directs and Coordinates Work	.15**	.06	0	0	.01	.14**	-.03
Gets the Job Done	.11*	.03	.13*	-.05	.01	.20**	-.10*
Motivates and Develops	.25**	-.04	-.08	.06	.10	.19**	-.02
Builds and Leads Teams	.15**	-.01	-.03	.05	.03	.12*	.03
Influences Others	.22**	.04	-.04	-.04	.13*	.20**	0

Notes. Correlations are based on an N that ranges from 341 to 384.
*Significant at .05.
**Significant at .01.