The Power to Predict Performance.™

Incremental Validity of Numerical Reasoning over Critical Thinking

Kingsley C. Ejiogu

Mark Rose

Zhiming Yang

and

John Trent

August 2007

PEARSON

70556

Incremental Validity of Numerical Reasoning over Critical Thinking

Abstract

This study was conducted to evaluate the incremental validity of numerical reasoning over critical thinking in predicting job performance and overall potential as measured by supervisors' ratings. Results indicated significant incremental validity of numerical reasoning over and above critical thinking in the prediction of overall performance and overall potential.

This paper presents the results of an examination of the incremental validity of numerical reasoning over critical thinking in the prediction of job performance and overall potential as measured by supervisors' ratings.

Professionals with responsibility for employee selection processes are faced with choosing from a broad array of assessment tools to help screen and select candidates. Making a correct choice has the potential to maximize the effectiveness of selection decisions while minimizing time and costs involved. Conversely, choosing the wrong set of assessment tools can lead to poor hiring decisions and wasted resources.

Due to the relevance and importance of critical thinking across a wide range of jobs (O*Net OnLine, 2005), the use of tools and methods for evaluating critical thinking of job applicants has been a popular practice among HR practitioners, I/O Psychologists, and other professionals working in the area of employment selection.

Ejiogu, Kingsley C., Rose, Mark, Yang, Zhiming, and Trent, John (2007). Incremental Validity of Numerical Reasoning over Critical Thinking. Poster session presented at the 115th annual convention of the American Psychological Association, San Francisco, CA,
August 17-20.

The authors gratefully acknowledge the assistance of Paul E. Williams, PsyD, on the presentation of this paper.

In general, the use of critical thinking assessments has been shown to be of potential value in making hiring decisions. For example, critical thinking scores have been found to correlate with job performance (Ejiogu, Yang, Trent, & Rose, 2005; Gaston, 1993), work sample performance (Kudisch & Hoffman, 2002; Spector, Schneider, Vance, & Hezlett, 2000), and organizational level attained (Watson & Glaser, 2006).

Despite the ubiquitous presence of critical thinking evaluations, there appears to be limited research showing the types of instruments that are most likely to complement and increase the predictive power of such evaluations as part of the hiring process. Two likely exceptions were studies on assessment centers in which critical thinking was shown to relate more strongly to cognitively loaded measures than measures of interpersonal skills (Kudisch & Hoffman, 2002; Spector, Schneider, Vance, & Hezlett, 2000). For example, in a study of managerial and executive-level participants, Watson-Glaser scores were found to relate more strongly to in-basket exercise scores ($r = .26$, $p < .05$) than a coaching exercise ($r = .16$, $p < .05$; Spector, et al., 2000). These results imply, albeit indirectly, that measures of interpersonal skills are likely to supplement critical thinking measures in predicting overall job performance. In a similar vein, the extensive research on personality, job interviews, and work samples as complements to cognitive ability tests (Schmidt & Hunter, 1998), and the burgeoning literature on emotional intelligence (e.g., Goleman, 1998), indicate that these types of soft-skill assessments are appropriate complements to critical thinking assessments.

What is more challenging to anticipate, however, is whether other cognitive ability evaluations are likely to complement critical thinking assessments. Furthermore, it is open to question whether including other cognitive ability assessments in a selection system is worthwhile with regard to the anticipated benefits versus the costs. Given the high correlation among many cognitive ability tests, even a carefully done job analysis, while certainly helpful, may not provide the information needed to estimate with precision the benefits that could be gained from adding other cognitive ability tests to an assessment of critical thinking.

In this study, we evaluated the incremental validity of a specific type of cognitive ability assessment, numerical reasoning, over and above critical thinking in the prediction of job performance.

Using O*Net online, we chose to include individuals in our sample if they worked in a job that required both strong critical thinking skills and numerical reasoning ability. Based on this sampling requirement, we hypothesized that the numerical reasoning measure would provide incremental validity to the critical thinking measure in predicting job performance.

Method

Participants

Participants were 87 job incumbents, about two-thirds of whom occupied finance-related positions (e.g., Actuary, Banking Supervisor, Business Unit Leader, Finance Analyst, Purchasing Manager) in a variety of industries. Close to two-thirds (63%) of the participants were employed in the financial services/banking industry.

The job levels of the participants included executives and directors (22%), managers (25%), and professionals/individual contributors (33%), while the rest were in administrative or other first-line supervisory levels.

The gender composition of the 87 participants was 45% males and 55% females. With regard to the highest educational level reported by the participants, 19% had a Masters degree or higher qualification, 6% had done some post-graduate work, 33% had a Bachelor's degree, 25% had an Associate Degree or 1-4 years of college education, and 17% had a high school diploma or GED. The participants were about 90% White (non-Hispanic), about 2% Black/African American, about 3% Hispanic/Latino (a), about 3% Asian/Pacific Islander, and about 1% Other.

Measures

The two predictors examined in this study were critical thinking and numerical reasoning. Critical thinking was measured with the 40-item Watson-Glaser Critical Thinking Appraisal—Short Form (Watson & Glaser, 2006). Numerical reasoning was measured with the 32-item Advanced Numerical Reasoning Appraisal (ANRA; Rust, 2006).

The Watson-Glaser is composed of the following five tests:

(1) **Inference**—discriminating among degrees of truth or falsity of inferences drawn from given data

(2) **Recognition of Assumptions**—recognizing unstated assumptions or presuppositions in given statements or assertions

(3) **Deduction**—determining whether certain conclusions necessarily follow from information in given statements or premises

(4) **Interpretation**—weighing evidence and deciding if generalizations or conclusions based on the given data are warranted

(5) **Evaluation of Arguments**—distinguishing between arguments that are strong and relevant and those that are weak or irrelevant to a particular issue

Each Watson-Glaser test is composed of scenarios similar to those typically found in a variety of settings, including the workplace, the school, and other organizational settings. Each scenario is followed by a number of items for the participant to respond to, with response options ranging from 2 for some items to 5 for other items. The Watson-Glaser score used as the measure of critical thinking ability was the total score (ranging from 0 to 40) derived from the summation of the scores on the five tests.

Internal consistency reliability (coefficient alpha) of the Watson-Glaser total score for the sample of 87 in this study was .85.

The ANRA is a set of two tests:
Test 1—Comparison of Quantities and
Test 2—Sufficiency of Information.

Test 1 requires the candidate to make a comparison between two quantities and determine which quantity, if any, is greater, or whether the two quantities are equal, or if there is insufficient information to make the comparison. Test 2 requires the candidate to consider two statements, either separately or in combination, and then decide whether the information provided in the two statements is sufficient to answer an accompanying question. ANRA requires the candidate to apply his or her numerical reasoning skills to decisions that reflect the variety of numerical estimation and analytic tasks frequently encountered in many everyday situations. There are four response options per question, and the maximum total raw score on ANRA is 32. The internal consistency reliability (coefficient alpha) of the total score on ANRA for the sample used in this study was .87.

The three criteria we examined in this study were average performance, overall performance, and overall potential. We measured the criteria with a 26-item questionnaire independently completed on each participant by the participant's work supervisor. We derived the first criterion—Average Performance—by averaging the ratings on 24 performance behaviors in the questionnaire. For individuals who did not have complete data on all 24 performance behaviors, we predicted missing values using non-missing performance ratings. Each of the other two criteria—Overall Performance and Overall Potential— was measured with a single item. All the job behaviors were derived from O*Net Online descriptions of jobs similar to the target jobs (that is, occupations related to banking, finance, or insurance).

Procedure

Participants completed the computer-administered Watson-Glaser and ANRA as part of a larger validation and normative study. The data were collected over a five-month period in 2006. The participants signed consent forms with the understanding that their data would be used purely for research purposes. The job supervisor of each participant provided ratings using the performance rating form supplied by the researchers. The performance rating form contained 26 items. Twenty-four of these items were behaviors regarding the following three composite areas that were relevant to most professional, managerial, and executive jobs in finance-related occupations: Analysis and Problem Solving, Judgment and Decision Making, and Quantitative/Professional Knowledge and Expertise. The ratings of behaviors in the above three areas ranged from 1 = "needs improvement" through 4 = "acceptable" to 7 = "outstanding." A "Not Applicable" rating was also available for behaviors that supervisors considered irrelevant to the job. Additionally, the supervisors rated their respective subordinates on single-item measures of overall performance and overall potential. The single-item measure of overall performance had the same seven-point rating scale as the twenty-four behaviors, described previously. The single-item measure of Overall Potential used a 5-point scale ranging from 1 = "no higher than current job" to 5 = "more than two levels above current job."

Data Analyses

Participant data were analyzed for subordinates whose supervisors reported (a) having supervised them for at least three months and, (b) that the supervisors were at least "knowledgeable" of the job performance of the subordinate.

After an examination of the zero-order correlations among the predictors and criteria, we used hierarchical linear regression analyses to examine the incremental validity of the ANRA score over and above the Watson-Glaser score. In step 1 of the two-step hiearchical linear regression, we regressed the critical thinking scores onto each criterion. In step 2, we added the numerical reasoning scores to the critical thinking scores and regressed both predictors onto each of the three criteria—average performance, overall performance, and overall potential.

In addition to the zero-order correlations and hierarchical linear regression analyses, we also examined the partial correlation of each predictor with the criteria while controlling for the other predictor.

Results

The means, standard deviations, internal consistency reliabilities, and zero-order correlations between the variables are presented in Table 1. All the correlation coefficients shown in Table 1 are uncorrected for unreliability in the criterion.

The results of the hierarchical regression analyses are presented in Tables 2, 3, and 4, for each of the three criteria—average performance, overall performance, and overall potential. The partial correlations are presented in Table 5.

The results in Table 1 show that, with the exception of the correlation between the Watson-Glaser total score and overall job potential, each of the two predictors is significantly correlated with each of the three criteria. Table 1 also shows that the two predictors are significantly correlated with each other ($r = .75$, $p < .01$).

Table 1
Means, Standard Deviations, Internal Consistency Reliabilities, and Zero-Order Correlations Between Critical Thinking, Numerical Reasoning, Average Job Performance, Overall Job Performance, and Overall Job Potential.

Variable	*M*	*SD*	1	2	3	4	5
1. Critical Thinking (Watson-Glaser Total Score)	30.60	6.3	(.85)				
2. Numerical Reasoning (ANRA Total Score)	20.63	6.47	.75	(.87)			
3. Average Job Performance	5.35	0.91	.34	.37	(.97)		
4. Overall Job Performance (single-item rating)	5.56	1.16	.29	.37	.84	(–)	
5. Overall Job Potential (single-item rating)	3.41	1.14	.18	.33	.63	.60	(–)

Note. Listwise. N = 87. Correlations greater than .18 are statistically significant (p < .01, two-tailed). The significance level for the correlation coefficient .18 is .09. Reliability coefficients indicated along the diagonal are alpha coefficients.

Results

Table 2 shows that numerical reasoning accounted for additional 2% of unique variance in average performance over and above the 12% of the variance explained by critical thinking. This 2% incremental variance was not statistically significant ($p > .05$).

Table 2
Hierarchical Linear Regressions: Critical thinking and Numerical Reasoning as Predictors of Average performance.

Variable		Standardized Beta Coefficient (β)	R Square	Adjusted R Square	R Square Change
Model 1	Critical Thinking	.34**	.12	.11	
Model 2	Critical Thinking	.16			
	Numerical Reasoning	.25	.14	.12	.02

Note. Average performance rating was based on the average of supervisory ratings on 24 behaviors related to Analysis and Problem Solving, Judgment and Decision Making, and Quantitative/Professional Knowledge and Expertise. ** p < .01 (two-tailed).

In Table 3, we can see that numerical reasoning accounted for a statistically significant incremental variance of 6% ($p < .05$) over and above the amount of variance in overall performance explained by critical thinking.

Table 3
Hierarchical Linear Regressions: Critical Thinking and Numerical Reasoning as Predictors of Overall Performance.

Variable		Standardized Beta Coefficient (β)	R Square	Adjusted R Square	R Square Change
Model 1	Critical Thinking	.29**	.09	.07	
Model 2	Critical Thinking	.03			
	Numerical Reasoning	.35*	.14	.12	.06*

Note. Overall performance was based on supervisor ratings of a single-item measure of overall performance. * p < .05 ** p < .01 (two-tailed).

•

As can be seen in Table 4 regarding the variance in overall potential, numerical reasoning explained a statistically significant 9% of incremental variance over and above the variance explained by critical thinking.

Table 4.
Hierarchical Linear Regressions: Critical Thinking and Numerical Reasoning as Predictors of Overall Potential

Variable		Standardized Beta Coefficient (*β*)	R Square	Adjusted R Square	R Square Change
Model 1	Critical Thinking	.18	.03	.02	
Model 2	Critical Thinking	-.15			
	Numerical Reasoning	.45**	.12	.10	.09**

Note. Overall potential was based on supervisor ratings of a single-item measure of overall potential. ** p< .01 (two-tailed).

The partial correlations in Table 5 also show that, separate from critical thinking, numerical reasoning shares some significant unique relationships with the criteria of overall performance and overall potential.

Table 5.
Partial Correlations of each Predictor (Critical Thinking and Numerical Reasoning) with the Criteria

Variable	Average Job Performance	Overall Job Performance	Overall Potential
Critical Thinking *(Controlling for Numerical Reasoning)*	.11	.02	-.11
Numerical Reasoning *(Controlling for Critical Thinking)*	.18	.25*	.30**

Note. Listwise. N = 84. * p< .05 ** p< .01 (two-tailed).

Discussion and Practical Implications

The results suggest that numerical reasoning explained a significant amount of incremental variance in two of the three criteria studied, over and above the variance explained by critical thinking. The addition of numerical reasoning to critical thinking explained an additional 6% of the variance in overall job performance ratings (see Table 3), and an additional 9% of the variance in overall potential (Table 4). On the other hand, the incremental variance of 2% explained by numerical reasoning over and above critical thinking with regard to average job performance (see Table 2) was not significant ($p > .05$).

Overall, the significant incremental validity of numerical reasoning over critical thinking in predicting overall performance (Table 3) and overall potential (Table 4) suggests that adding a measure of numerical reasoning to critical thinking would be beneficial in some situations. For example, in general, the typical hiring professional may consider it useful to add a test of numerical reasoning if the expected benefit is 6% increase in predicting job performance or 9% increase in predicting job potential.

One should note, however, that the incremental validity of a given test should not be the only basis for a decision whether to use that particular test or another test. Relative cost of the predictor (s) and the importance of the criterion influence the type of decisions we make with incremental validity results (Haynes & Lench, 2003). In situations where the nature of the job performance involves a lot of numerical reasoning and decision making, using a numerical reasoning test is likely to cover more job-applicable facets than would otherwise be captured by a verbal critical thinking test. For example, inferences regarding the incremental validity of a test are typically influenced by such other factors as the objectives the test user wants to accomplish in the assessment process, the predictors and criteria of interest, the particular sample or norm group, and the relative cost of acquiring the additional test (Haynes & Leach, 2003). As a result, a test of numerical reasoning is likely to be more applicable when assessing candidates for positions in finance-related occupations. On the other hand, when assessing for positions where critical thinking using language is a key skill, or positions that do not require a great deal of numeracy, a critical thinking test is likely to be more applicable.

Given the strong correlation between critical thinking and numerical reasoning as measured in this study ($r = .75$, $p < .01$), the relatively small magnitudes of the increments in explained variance are not altogether surprising. We used a stringent approach (hierarchical linear regression) to assess incremental validity in this study given that any shared variance between the predictors is taken by the first predictor (critical thinking) entered into the regression equation (Hunsley & Meyer, 2003). Critical thinking is a higher-order skill that is relevant and applicable in a variety of situations, including situations that involve proper use of language, making inferences, calculating likely outcomes, making decisions, and solving problems (Paul & Nosich, 2004).

Limitations and Suggestions for Future Research

One of the challenges we encountered in this study was getting otherwise busy employees to take the tests for research purposes and for the supervisors of these employees to also independently provide performance ratings on their subordinates. Consequently, the sample used in this study was composed of available volunteers rather than a stratified sample of job incumbents. Additional research should be done with larger and stratified samples that can afford the researcher the opportunity for cross-validation, especially given the fact that inferences regarding the incremental validity of two strongly correlated predictors are less likely to be robust across samples (Dawes, 2001).

Conclusion

This study is one of the first empirical studies that we are aware of that evaluated the incremental validity of a cognitive ability measure in addition to a critical thinking measure in relationship to job performance. This particular study showed that incremental validity could be gained, and that incremental validity gained was probably enough to justify the time and costs associated with administering an additional test. The researchers believe that more of these types of studies, looking at other jobs and other types of predictors, are essential for providing hiring professionals with the information necessary to decide which cognitive ability predictors to include in their selection batteries.

References

Dawes, R. M. (2001). Incremental validity of the Ego Impairment Index: It's fine when it's there. Psychological Assessment, 13, 408-409.

Ejiogu, K.C., Yang, Z., Trent, J., & Rose, M. (2006). Understanding the relationship between critical thinking and job performance. Poster Presentation at the 21st Annual SIOP Conference, May 5–7, 2006, Dallas, TX.

Gaston, A. (1993). Recognizing potential law enforcement executives. (Reports No. NCJ 131646) Washington, D.C.: National Institute of Justice/NCJRS.

Goleman, D. (1998). Working with emotional intelligence. New York: Bantam Books.

Haynes, S.N., & Lench, H.C. (2003). Incremental validity of new clinical assessment measures. Psychological Assessment, 15(4), 456-466.

Hunsley, J., & Meyer, G.J. (2003). The incremental validity of psychological testing and assessment: conceptual, methodological, and statistical issues. Psychological Assessment, 15, 446-455.

Kudish, J. D. & Hoffman, B. J. (2002, October). Examining the relationship between assessment center final dimension ratings and external measures of cognitive ability and personality. Paper presented at the 30th International Congress on Assessment Center Methods, Pittsburgh, PA.

References

O*Net OnLine (2005). Searches for jobs requiring critical thinking and mathematical reasoning. Occupational Information Network: O*Net OnLine. Retrieved September 8, 2006 via O*Net OnLine Access: http://online.onetcenter.org/skills/result?s=2.A.2.a&g=Go

Paul, R., & Nosich, G.M. (2004). A Model for the National Assessment of Higher Order Thinking. Retrieved July 13, 2006, from http://www.criticalthinking.org/resources/articles/a-model-nal-assessment-hot.shtml

Rust, J. (2006). Advanced Numerical Reasoning Appraisal manual. San Antonio, TX:

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.

Spector, P. A., Schneider, J. R., Vance, C. A., & Hezlett, S. A. (2000). The relation of cognitive ability and personality traits to assessment center performance. Journal of Applied Social Psychology, 30(7), 1474–1491.

Watson, G., & Glaser, E. M. (2006). Watson-Glaser Critical Thinking Appraisal, Short Form manual. San Antonio, TX: