The Power to Predict Performance.™

Understanding the Relationship Between Critical Thinking and Job Performance

Kingsley C. Ejiogu

Zhiming Yang

John Trent

Mark Rose

May 2006

PEARSON

Understanding the Relationship Between Critical Thinking and Job Performance

Abstract

This study was conducted to evaluate the relationship between a measure of critical thinking ability and job performance as measured by supervisors' ratings. Results indicated that the measure of critical thinking ability is related to several important aspects of job performance.

This paper presents the results of a study examining the relationship between critical thinking ability, as measured by total scores on the Watson-Glaser Critical Appraisal Short Form (Watson-Glaser), and job-related performance. A review of research literature suggests that the bulk of published studies on the Watson-Glaser relate to its use to predict performance in a variety of educational settings. For example, Gadzella, Stacks, Stephens, & Masten, (2005) found the Watson-Glaser to be "a good instrument to measure critical thinking for students pursuing the teaching career" (p.12). In the study by Gadzella et. al., the researchers found a correlation of .31 between total scores on the Watson-Glaser and course grades. In their studies of three freshmen classes in a Pennsylvania nursing program, Behrens (1996) found that Watson-Glaser scores correlated .59, .53, and .51 respectively, with semester GPA. Similarly, in a study of 428 educational psychology students, Williams (2003) found that Watson-Glaser total scores correlated .42 and .57 with mid-term and final exam scores, respectively.

Ejiogu, Kingsley C., Yang, Zhiming, Trent, John, and Rose, Mark (2006). Understanding the Relationship Between Critical Thinking and Job Performance. Poster session presented at the 21st annual conference of the Society for Industrial and Organizational Psychology, Dallas, TX, May 5-7.

Email: Kingsley_Ejiogu@harcourt.com

The authors gratefully acknowledge the assistance of Nishidha Goel with the final version of this paper.

Critical thinking ability plays a vital role in academic instruction (e.g., College Board, as cited in Gadzella, Stacks, Stephens, & Masten, 2005), as well as in occupations that require careful analytical thinking to perform essential job functions (e.g., Spector, Schneider, Vance, & Hezlett, 2000). Kudish and Hoffman (2002), in a study of 71 participants in a leadership assessment center, reported that scores on the Watson-Glaser correlated .58 with ratings on Analysis and .43 with ratings on Judgment. The ratings on Analysis and Judgment obtained by Kudish and Hoffman were based on performance of the participants across assessment center exercises including a coaching meeting, in-basket simulation, and a leaderless group discussion. In a study of managerial and executive-level participants in an assessment center, Spector, Schneider, Vance, & Hezlett, (2000) evaluated the relationship between Watson-Glaser scores and performance in the assessment center exercises. Spector, Schneider, Vance, & Hezlett, found that Watson-Glaser scores significantly correlated with overall scores in six of eight assessment center exercises, and related more strongly to exercises involving primarily cognitive problem-solving skills (e.g., with in-basket exercise scores, $r = .26$, $p < .05$) than exercises involving a greater level of interpersonal skills (e.g., with a coaching exercise, $r = .16$, $p < .05$).

Another indication of the importance of critical thinking to effective performance across various occupations and job levels can be found in O*Net OnLine (2005). In a search of the O*Net OnLine database, the authors found that, for occupational positions like Government Service Executives and Private Sector Executives, critical thinking received standardized importance ratings of 92 and 79, respectively (on a scale of 0 – 100); the occupation Program Directors had a standardized importance rating of 76 attached to critical thinking. For various manager positions (e.g., Compensation & Benefits Managers, Construction Managers, Financial Managers, Marketing Managers, Storage & Distribution Managers, Training & Development Managers), the standardized importance rating attached to critical thinking ranged from 76 to 88. In such professional positions as Actuaries, Chiropractors, Emergency 911 Dispatchers, Industrial/Organizational Psychologists, Management Analysts, Registered Nurses, standardized importance ratings attached to critical thinking ranged from 81 to 94. Other examples from O*Net regarding the rated importance of critical thinking to a variety of occupations include Actors (73), Concierges (74), Employment Interviewers (73), Fashion Designers (79), First-Line Supervisors – Customer Service (79), Security Guards (74), and Tax Preparers (73).

In their book on Staffing Organizations: Contemporary Practice and Theory, Ployhart, Schneider, and Schmitt (2006) highlighted the need for researchers to engage in more current efforts to update cognitive ability tests and conduct new studies of the validity of such tests as an important predictor of job performance. According to Ployhart et. al., with few exceptions, “not many primary studies have added to the database that support the validity of cognitive ability since the 1970’s” (p.415). Consequently, the main purpose of this paper is to report the findings of a study on the criterion-related validity of Watson-Glaser total scores as a predictor of supervisor-rated performance of job incumbents.

Research Questions

The main research questions in this study were:

(1) is the 40-item Watson-Glaser test a reliable instrument to measure critical thinking ability in the workplace, and (2) is critical thinking ability as measured by the Watson-Glaser related to job performance as measured by supervisor ratings?

Method

Participants

The participants were 84 job incumbents working as Analysts (a professional-level individual contributor position) in a government agency. The gender composition of the participants was 49 (58%) males and 25 (29.8%) females, with 10 providing no information regarding their gender. Out of the participants that provided information regarding their highest educational qualifications, 19 reported having a Masters degree or higher qualification, 7 reported having done some post-graduate work, while 12 reported having a Bachelors degree. Seventy-nine of the participants provided ethnic group information as follows: 72 (85.7%) White (non-Hispanic), 3 (3.6%) Black/African American, 3 (3.6%) Hispanic/Latino (a), and 1 (1.2%) Asian/Pacific Islander.

Materials

The 40-item Watson-Glaser Critical Thinking Appraisal – Short Form (Watson & Glaser, 1994) was used to measure the critical thinking ability of the participants. The 40-item Watson Glaser was published in 1994 to enhance the use of the Watson Glaser Critical Thinking Appraisal for selection and developmental purposes in a range of work settings, and to assess critical thinking skills in an educational context. According to Watson and Glaser, 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. Coefficient alpha and test-retest reliability coefficients of the total score on the 40-item Watson-Glaser test had both been estimated to be .81 (Watson & Glaser, 1994).

Job performance was measured using a 21-item questionnaire independently completed on each participant by the participant's work supervisor. Nineteen items required ratings on job-relevant behaviors while two items required ratings on overall performance and overall potential. All the behaviors were derived from O*Net Online descriptions of jobs similar to the target job.

Procedure

The participants completed the computer-administered Watson-Glaser as part of a larger validation and normative study. The data were collected over a six-month period in 2004. 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 21 behavioral items. Nineteen of these items were behaviors regarding the following three composite areas that were relevant to most professional, managerial, and executive jobs: Analysis and Problem Solving, Judgment and Decision Making, and Professional/Technical 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 a single-item Overall Potential using a 5-point scale ranging from 1 = "no higher than current job" to 5 = "more than two levels above current job." The researchers also examined a composite variable – Overall Performance – that they composed by summing the ratings on 19 performance behaviors in the questionnaire.

Data were analyzed by calculating correlation coefficients for the relationships of critical thinking with job performance. Participant data were analyzed for subordinates whose supervisors reported (a) having supervised them for at least four months and, (b) that the supervisors were at least "knowledgeable" of the job performance of the subordinate. The observed criterion-related validity coefficients were corrected for unreliability in the criterion (Schmidt & Hunter, 1998).

Results

The Watson-Glaser total scores of the 84 participants ranged from 21 to 40, with a mean of 32.8 and a standard deviation of 4.6.

Evidence of Reliability

The internal consistency reliability of the total score on the 40-item Watson-Glaser test, was .85 when corrected for restriction of range using the correction formula by Allen and Yen (1979, p.195). This result suggests "good reliability" (U.S. Department of Labor, 1999, p. 3-3) of this test as a measure of critical thinking ability. The uncorrected reliability coefficient was .76.

Relationship Between Watson-Glaser Total Scores and Job Performance

Out of the 84 participants, the researchers examined the relationship between Watson-Glaser scores and on-the-job performance of 68 participants for whom there were performance ratings from their supervisors. This group of 68 participants had a mean Watson-Glaser score of 32.9 with a standard deviation of 5.8. As shown in Table 1 regarding the criterion-related validity coefficients corrected for unreliability in the criterion, the Watson-Glaser total scores correlated .52 with performance ratings on each of the two performance dimensions of (1) Analysis and Problem Solving and (2) Judgment and Decision Making. The results in Table 1 also show that the Watson-Glaser total scores correlated .48 with performance ratings of job behaviors dealing with Professional/Technical Knowledge and Expertise. The correlation of the Watson-Glaser scores with the Overall Performance composite was .51, while the correlation with the single-item Overall Potential was .32 (see Table 1). Table 1 also shows the correlation coefficients uncorrected for unreliability in the criterion.

Table 1. Means and standard deviations of criterion variables, and correlation coefficients between Watson-Glaser total scores and supervisors' job-performance ratings of participants.

Criterion Dimension (Supervisory Ratings of Job Performance Behaviors)

Criterion Dimension	*N*	*Mean*	SD	r (corrected for criterion unreliability)	r (uncorrected for criterion unreliability)
1. Analysis & Problem Solving	64	38.3	6.6	.52	.40
2. Judgement & Decision Making	59	32.8	5.8	.52	.40
3. Professional / Technical Knowledge and Expertise	65	17.1	2.4	.48	.37
4. Overall Performance	66	100.4	14.3	.51	.39
5. Overall Potential	64	3.2	1.2	.32	.25*

Note 1: * p < .05 for Overall Potential. For the other four correlation coefficients, p < .01
Note 2: The column labeled N shows the number of participants in the group or sub-sample whose data were in the results reported. In summing item scores from the performance questionnaire to yield criterion dimensions, cases with missing values were dropped.

Discussion

This study attempted to examine the internal consistency reliability and validity of the 40-item Watson-Glaser among job incumbents. The corrected internal consistency reliability of .85 shown in this study indicates that the Watson-Glaser total score possesses good internal consistency reliability for this sample of job incumbents.

Previous researchers have reported higher internal consistency reliability coefficients of the Watson-Glaser. For example, Gadzella, Baloglu, & Stephens (2002) reported internal consistency reliability coefficients of .91 for 30 men, .83 for 105 women, and .91 for 135 students majoring in Education. Since the reliability coefficient is a correlation coefficient, the relatively high and narrow range of ability among the participants in this study might have attenuated the internal consistency reliability obtained in this study, resulting in the uncorrected reliability coefficient of .76. Samples with restricted variances can lead to reliability coefficients being spuriously low (Nunnally & Bernstein, 1994).

The results indicate that critical thinking ability as measured by the 40-item Watson-Glaser is significantly related to job performance. All the criterion-related validity coefficients reported in Table 1 suggest that the 40-item Watson-Glaser could be regarded as "very beneficial" (U.S. Department of Labor, 1999, p. 3-10) for use in relating critical thinking ability to the following aspects of performance: Analysis and Problem Solving, Judgment and Decision Making, Professional/Technical Knowledge and Expertise, and Overall Performance. The results in Table 1 also indicate that the Watson-Glaser test is "likely to be useful" (U.S. Department of Labor, 1999, p. 3-10) in relating critical thinking ability to Overall Potential. There is also convergent validity evidence from previous studies relating the Watson-Glaser to other cognitive ability tests. Such evidence can be found in Watson and Glaser (1994) showing significant relationships between scores of mid-level management applicants on the Watson-Glaser and their scores on the following tests: Wesman Personnel Classification Test, Verbal (.66), Employee Aptitude Survey – Verbal Reasoning (.51), Employee Aptitude Survey – Verbal Comprehension (.50), Employee Aptitude Survey – Numerical Reasoning (.41), Employee Aptitude Survey – Space Visualization (.26). For a sample of executive management applicants, the correlation between their scores on the Watson-Glaser and their scores on Differential Aptitude Tests – Abstract Reasoning – was .40 (Watson & Glaser, 1994).

Many organizations typically use selection tests in their hiring process. The results obtained in this study suggest that in jobs such as those of Analysts where critical thinking ability is important for successful performance, the Watson-Glaser is likely to be beneficial as part of the external or internal selection process for the job. However, since successful performance typically depends on several factors and aspects related to the job, no single test is sufficient to cover all the aspects of performance. As such, combining information from the Watson-Glaser with other sources of information (for example, interviews, work samples, and records of past performance) will likely enhance the validity of the selection process.

Of course, it is important to note that the practical value of a selection test depends not only on its validity but also on such factors as the base rate

for success on the job (that is, the proportion of people who would be successful in the absence of any selection tool), the selection ratio (that is, the ratio of applicants to the number of vacancies to be filled), adverse impact associated with the test, the cost of hiring error, and the cost of the test itself (Anastasi & Urbina, 1997; Cascio, 1997; Schmidt & Hunter, 1998; U.S. Department of Labor, 1999).

Limitations and Suggestions for Future Research

One of the difficulties of conducting real-life validation studies is getting enough of the needed participants to provide the required data. Consequently, this study was challenged by the constraints of getting otherwise busy employees to take the Watson-Glaser for research purposes and for the supervisors of these employees to also independently provide performance ratings on their subordinates.

Some researchers might be interested in how examinee scores on the Watson-Glaser relate to national norms. The focus and constraints of this study, however, necessitated the use of a sample that was more occupation-specific than national in scope. When using cognitive ability tests for purposes of talent assessment, comparing scores of candidates against a norm population of relevant occupation groups is usually more applicable than using the general census-type "national norms" for such an occupation-specific purpose.

Given the relevance of critical thinking in the employment context, as well as the popularity of the Watson-Glaser as a measure of critical thinking ability, the organizational literature would benefit from more published studies that relate critical thinking ability to performance. For example, in addition to more concurrent validation studies, it also would be beneficial for future researches to try and publish predictive validation studies relating scores on critical thinking tests to subsequent performance.

References

Allen, M.J., & Yen, W.M. (1979). Introduction to measurement theory. Monterrey, CA: Brooks/Cole.

Anastasi, A. & Urbina, S. (1997). Psychological testing (7th ed.). Upper Saddle River, N.J.: Prentice Hall.

Behrens, P. J. (1996). The Watson-Glaser Critical Thinking Appraisal and academic perform Journal of Nursing Education, 35, 34–36.

Cascio, W. F. (1997). Applied psychology in human resource management (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Gadzella, B.M. Stacks, J., Stephens, R.C., & Masten, W.G. (2005). Watson-Glaser Critical Thinking Appraisal, Form S for education majors. Journal of Instructional Psychology, 32, 9-12.

Gadzella, B. M., Baloglu, M., & Stephens, R. (2002). Prediction of GPA with educational psychology grades and critical thinking scores. Education, 122(3), 618–623.

Holmgren, B. & Covin, T. (1984). Selective characteristics of preservice professionals. Education, 104, 321–328.

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.

Nunnally, J.C. & Bernstein, I.H. (1994). Psychometric theory (3rd ed.). New York: McGraw-Hill, Inc.

O*Net OnLine (2005). Skill searches for: Critical thinking. Occupational Information Network: O*Net OnLine. Retrieved April 21, 2006 via O*Net OnLine Access: http://online.onetcenter.org/skills/result?s=2.A.2.a&g=Go

Ployhart, R.E., Schneider, B., & Schmitt, N. (2006). Staffing organizations: Contemporary theory and practice (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

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.

U.S. Department of Labor. (1999). Testing and assessment: An employer's guide to good practices. Washington, DC: Author.

Watson, G., & Glaser, E. M. (1994). Watson-Glaser Critical Thinking Appraisal, Form S manual. San Antonio, TX: The Psychological Corporation.

Williams, R. L. (2003). Critical thinking as a predictor and outcome measure in a large undergraduate educational psychology course. (Report No. TM-035-016). Knoxville, TN: University of Tennessee. (ERIC Document Reproduction Service No. ED478075).

Pearson

The Talent Assessment business of Pearson develops and delivers assessments for employee selection and development that increase workforce performance. With more than 80 years of experience in the employee assessment field, Pearson offers award-winning products such as the Watson-Glaser Critical Thinking Appraisal, Bennett Mechanical Comprehension Test, and the Golden Personality Type Profiler, and has a growing global presence with offices in seven countries. Its clients include half of the Fortune 500, expanding small businesses, and all organizations that make identifying talent and unlocking employee potential a top priority. It is a part of Pearson (NYSE: PSO), the international media company. Pearson's other primary businesses include the Financial Times Group and the Penguin Group.

Learn more about Pearson's Talent Assessment products at www.AssessTalent.com.

The Power to Predict Performance.™

888-298-6227 • AssessTalent.com

PEARSON