PREDICTING LONG TERM JOB PERFORMANCE USING A COGNITIVE ABILITY TEST

Sandra G. Alexander, B.A., M.S.

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

August 2007

APPROVED:

Linda Marshall, Major Professor and Chair of the Department of Psychology
Michael Beyerlein, Minor Professor
Paul Lambert, Committee Member
Damon Bryant, Committee Member
Sandra L. Terrell, Dean of the Robert B. Toulouse School of Graduate Studies

Alexander, Sandra G., Predicting long term job performance using a cognitive ability test. Doctor of Philosophy (Industrial Organizational Psychology), August 2007, 31 pp., 9 tables, references, 25 titles.

This study focuses on the relationship of one cognitive ability test on long-term job performance as measured by personnel data. Archival data from over 3,000 employees at an international technology company were used to assess how aptitude test scores relate to both objective and subjective job performance measures. Supervisory performance ratings, level of promotion, and salary increase significantly contributed to variance in test scores; however, these results were inconsistent. Number of training courses did not have a significant relationship with test scores. Additionally, type of turnover did not moderate the relationship between aptitude test scores and job performance. These results indicate that although aptitude test score is related to long term job performance factors, other factors account for the majority of the variance. The implication is that aptitude should not be the sole consideration when predicting long term job success.

Job or work analyses in organizations are conducted to highlight key skills and abilities that are necessary for successful job performance (Brannick & Levine, 2002). These skills and abilities are often used to identify tests that may be used to screen job applicants. The use of tests should be guided by the knowledge, skills, abilities, and other characteristics that are identified as being critical to the job and job level. Research has shown that specific types of tests such as ability and personality tests are likely to be related to job performance (Gatewood & Field, 2001; Hough & Oswald, 2000; Robertson & Smith, 2001). One type of test that is related to multiple job roles is cognitive ability because most jobs require a baseline of intelligence (Hough & Oswald, 2000; Outtz, 2002; Ree, Earles, & Teachout, 1994; Schmidt, 2002; Schmidt & Hunter, 1998; Tenopyr, 2002). Despite wide acceptance of the efficacy of cognitive ability screening for applicants, there have been relatively few studies to determine the relationship between this screening and long term job performance outcomes. This study addresses the relationship of one such screening test to outcomes across a period of up to 8 years.

The purpose of this study is to examine the direct relation of one cognitive ability test with long-term job performance among workers in a technology company. First, the issues associated with validating a cognitive ability test will be discussed before describing this validation effort. Specifically, the controversy over using cognitive ability tests for predicting potential for all groups of people will be examined. Next, the measurement issues associated with any validation effort will be addressed, including the use of appropriate criteria, accounting for the dynamic nature of criteria, and the most appropriate method for handling missing data. Also, a brief review of how turnover may impact validation efforts is also discussed. Finally, three hypotheses about cognitive ability test results will be presented.

Organizations often make hiring decisions based on an individual's test performance. The assumption is that tests are objective, standard measures that can be used to assess an applicant's likelihood of success in a particular job. Although test performance can provide an indication of potential job performance, multiple factors (e.g., work experience) often make important contributions (Arvey & Murphy, 1998). In a review of performance evaluations in work settings, Arvey and Murphy (1998), found the domain of job performance was expanding, indicating that task proficiency may no longer be a sufficient measure of performance. They suggest that because work environments are moving towards flexible roles and jobs more focus should be put on measuring personal competencies rather than specific tasks.

Cognitive Ability Testing

Although there is agreement that cognitive ability testing can be expected to predict performance in many jobs (Gatewood & Field, 2001; Schmidt & Hunter, 1984), there has been controversy over the widespread use of these tests for selection. Some of the controversy focuses on ability tests used as the sole basis for hiring decisions and the potential for excluding historically low scoring groups (Bobko, Roth, & Potosky, 1999; Kehoe, 2002; Murphy, Cronin, & Tam, 2003) which are most frequently racial or ethnic minorities. This leads to the possibility of adverse impact, discrimination in hiring that occurs when members of a subgroup are selected disproportionately less frequently than members of another group (Robertson & Smith, 2001). This unintentional discrimination can have a strong impact on members of the EEOC protected groups (e.g., minority groups, women, people over the age of 40, and people with disabilities).

There is a history of cognitive ability testing causing adverse impact (Hough & Oswald, 2000; Robertson & Smith, 2001). Terpstra, Mohamed, and Kethley (1999) reviewed court cases

involving different methods of selection and found ability tests were the most frequently challenged in court. In this review, cognitive ability tests were found to be nondiscriminatory in 67% of the observed cases. The Uniform Guidelines on Employee Selection Procedures (1978) state that if a test of equal or greater validity exists that has less adverse impact, that measure should be used. Because of the increased potential for adverse impact with cognitive ability tests, it is crucial they exhibit strong job relatedness.

One of the most common forms of assessment, cognitive ability testing, addresses general knowledge and capabilities. These tests are often characterized as measures of aptitude used to determine a person's ability to learn (Hunter, 1986; Hunter& Schmidt, 1996; Schmidt, 2002). Cognitive ability tests often include measures of verbal, mathematical, memory, and reasoning abilities (Gatewood & Field, 2001). The test examined in this study measures mathematical reasoning, numerical reasoning, and data manipulation. Reasoning skills are commonly measured (Gatewood & Field, 2001), but data manipulation is less commonly assessed. The three measures in this test are combined to yield an overall aptitude evaluation.

In comparison to other types of assessments, validity coefficients for cognitive ability tests show they are the single most effective predictor of job performance across all job types (Hough & Oswald, 2000; Schmidt & Hunter, 1998, 2004; Viswesvaran & Ones, 2002). Ree, Earles, and Teachout (1994) examined how both general cognitive ability and specific abilities predicted job performance criteria. Their results indicated that both general cognitive ability and specific abilities predicted performance criteria, however, specific abilities only added a small amount of predictive power above general cognitive ability. When both performance criteria were combined for the job of personnel specialist the observed correlation for general cognitive

ability was .53. When the specific abilities were added to general cognitive ability the observed correlation increased to .60.

In a review of the literature, Schmidt and Hunter (2004) found that the correlation between general mental ability and performance on the job ranged from .31 to .73. They also found that validity coefficients between cognitive ability and job performance were strongest for jobs high in complexity. Validity generalization allows justification for the use of a test in a new setting if there is accumulated validity evidence for the same or similar type of job (Gatewood & Field, 2001). Cognitive ability testing has been shown to have validity generalization in predicting performance (Schmidt & Hunter, 1998; Viswesvaran & Ones, 2002). The strongest validity for ability tests has been found for jobs that are complex and require high levels of information processing (Gatewood & Field, 2001). Thus, selection testing and its relevance to later job performance in a complex technology organization with cognitively demanding jobs, is appropriate to address the effectiveness of an aptitude test.

Measurement Issues

An important consideration when determining the usefulness of any selection test is the criteria the test is measured against. An important distinction should be made between using objective and subjective criteria. For the purposes of validation, objective criteria are often seen as superior because potential bias is minimized. Additionally, the value of objective criteria to the organization is often evident. On the other hand, subjective measures of performance have the potential to capture facets of performance that may not be included in objective measures. Many times validation criteria are chosen based on convenience, which may underestimate the relationship between selection tests and important performance criteria (Robertson & Smith, 2001). In addition, error variance may be misinterpreted as true variance. There may be bias or

systematic error variance driving the relationship between the construct of interest and the criterion (Arvey & Murphy, 1998). This extraneous component could be contaminating any observed relationship. For example, in a well known series of studies, the Hawthorne Studies, researchers originally attributed increases in productivity to changes in the plant's physical environment. However, when these changes were removed the increases in productivity continued, allowing the researchers to conclude that social factors were impacting the performance of the plant employees. If the researches had sustained the changes in the physical environment, the increases in productivity may have been incorrectly assigned as the cause for improved performance.

The criteria often used in validation studies are supervisory ratings of job performance (Robertson & Smith, 2001). However, these ratings often serve multiple purposes (e.g., pay increases, promotion, goal setting). When multiple outcomes are dependent on these ratings, supervisors may be more lenient than if the ratings were for research purposes only (Schultz & Schultz, 1998). Another problem is the subjective way ratings are assigned. They are based on one individual's perception of employee performance and often do not account for the amount of interaction between the rater and the individual being evaluated, which can affect accuracy. Additionally, many organizations determine performance ratings on an annual basis. If relevant information during the year is not recorded, the likelihood of error is increased. The accuracy of supervisory ratings is increased when notes about performance are kept during the time period of the performance review (Gatewood & Field, 2001). The performance ratings used in this study are based on the annual evaluation of overall performance. Although these ratings are used for multiple purposes (e.g., pay increases, goal setting) all employees are rated based on a consistent rating scale.

Objective measures of employee performance should also be considered when determining the validity of selection tests. Many organizations keep records of alternative measures of performance that can be considered as acceptable performance criteria. For example, increases in salary are normally indicative of strong performance. For the purposes of this study, salary increase will be determined by calculating the difference between an employee's starting base salary and base salary level at time of the data collection.

A criterion issue that is often overlooked is the changing nature of job performance. The determinants of performance, such as knowledge and motivation, are often treated as static when they are actually dynamic (Hough & Oswald, 2000; Robertson & Smith, 2001). Thus, studies need to consider these types of changes. One way to control for the dynamic nature of criterion measures is to examine multiple performance variables. An alternative would be to use criterion measures from multiple points in time. This study will address these issues by using multiple criteria, several of which have been measured more than once.

Another issue that commonly arises in validation research is that of missing data. Performance criteria are often based on incomplete organizational records. Consequently, Hough and Oswald (2000) advocate pairwise deletion and estimating missing scores with regression analyses. With listwise deletion, subjects and perhaps valuable information could be lost.

The use of multiple job performance criteria was previously dismissed because it was perceived as impractical and potentially confusing when different validities were calculated for the same predictor (Robertson & Smith, 2001). Criterion measures often fall into one of three groups: production output, personnel data, or ratings based on personal judgment. This study uses a combination of personnel data and subjective job performance ratings as the criteria for

validating a selection test. Although performance ratings are important criteria when examining test performance, organizations may also consider other relevant criteria.

Tenure and Turnover

Most of the research on tenure focuses on why employees stay with a company, rather than addressing pre-hire factors that may contribute to the likelihood of remaining once hired. The predictive validity of cognitive ability for tenure seems to be inconsistent. Griffeth et al. (2000) updated a meta-analysis on correlates of employee turnover. In the original study cognitive ability was associated with tenure, but this relationship was not replicated in the updated study. However, Mount et al. (2000) found that general mental ability, along with tenure, biodata, and the big five personality predictors, accounted for 17% of the variance in retention probability. General mental ability and tenure will be included in this study's analyses. An important distinction is between voluntary and involuntary turnover. When appropriate, the analyses conducted in this study will control for type of turnover.

Present Study

The aptitude test examined in this study is a standardized paper and pencil measure developed and validated for use as part of the selection process of entry-level applicants to technical positions in a large technology company. The test was originally designed in 1984 to measure reasoning skills for individuals that were applying for positions requiring complex logical analyses. Generally, these entry-level applicants had recently completed college and/or had less than two years of work experience. The technical positions that require the test include technical sales and services representatives, information technology specialists, software engineers, and programmers. No personal or motivational characteristics are assessed. The test

was originally validated with supervisory ratings and training performance as criteria measures of early job performance (within the first 6 months). These supervisory performance ratings were collected to be used for the purposes of research only. Managers were asked to rate employees on competencies such as technical knowledge, ability to rapidly learn, and problem solving. No organizational decisions were made from these ratings and were only available to the research team. The correlations between test performance and both criteria were in the moderate range which is consistent with previous research (Schmidt & Hunter, 1998, 2004) where the validity coefficients range from +.25 to +.50. The test has been shown to be a superior predictor of early job performance than grade point average, type of school attended, level of education, and number of job related courses completed (IBM, 1984). The test has been shown to have predictive validity for all groups and be free of adverse impact.

The purpose of this study is to determine the usefulness of an aptitude test for predicting long-term job performance. The potential maximum length of employment for the participants in this study is 8 years. Thus, this study improves on previous validation efforts that have focused on short term job performance, collecting criterion measures only after several months on the job. Extending the length of time before collecting criterion data should provide an accurate reflection of employee performance.

Because the majority of previous research (Hough & Oswald, 2000; Schmidt & Hunter, 1998, 2004; Viswesvaran & Ones, 2002) has found a positive relation between scores on aptitude tests and job performance, this outcome is expected to be consistent for a long term measure of job performance. Due to interrelations among job performance, promotions and salary (Gatewood & Field, 2001), aptitude testing results should also be positively related to promotions and raises. A review of longitudinal studies of aptitude test scores by Schmidt and

Hunter (2004) found that aptitude predicted both movement in job hierarchy and income. These considerations lead to

Hypothesis 1: Aptitude test scores will be positively related to supervisory ratings of job performance, promotions and salary increases.

In addition to being positively related to job performance, aptitude test scores are often shown to be related to performance in training courses (Hough & Oswald, 2000; Schmidt & Hunter, 1998, 2004; Viswesvaran & Ones, 2002). Although previous studies have looked at training performance, few have examined the number of training courses attended. One might expect those with high aptitude test scores to be proactive in acquiring knowledge and subsequently have a larger number of training courses completed. Thus,

Hypothesis 2: Aptitude test scores will be positively related to the number of completed training courses.

Although multiple job performance measures may be available, much of the existing research examines performance criteria on an individual basis (Arvey & Murphy, 1998; Schmidt & Hunter, 1998, 2004). Despite the distinction between subjective and objective performance criteria, they may be combined to yield a more accurate overall performance rating. Because multiple performance criteria will be available, a factor analysis will determine whether there is an overall underlying structure of performance. Finally,

Hypothesis 3: There will be one overall job performance factor that will be predicted by test performance.

Method

Participants

Archival data from over 3,000 current and past employees who completed an aptitude test as part of the employment selection process at an international technology company are included in this study. Data were gathered for job applicants who took the aptitude test and were hired between January of 1997 and December of 1998. These archival data were collected between 1997 and 2005. The aptitude test was administered during the application process, after an initial screening but prior to a formal interview with a hiring manager. The archival data include demographic information and organizational data. Of the participants, the majority of the sample was male (67.7%), Caucasian (60.8%), and under the age of 40 (87.4%).

Measures

Aptitude test scores from 1997 and 1998 were used to generate a list of employees on whom to gather performance data. This list of employees was sent to a central data warehouse where performance data from each employee's history were compiled. The variables included were aptitude test scores, supervisory job performance ratings, number of awards received, number of training courses, salary increase, promotions, management status, and tenure with company.

Test scores. As part of the application process, entry level job applicants to technical positions completed a paper and pencil cognitive aptitude test that measures their ability to learn and problem solve. This 70 minute test consists of three parts: a series of matrix manipulations, number series completion, and mathematical problem solving. (See Appendix A for sample questions.) Each subtest contains multiple choice questions with five answer options. For the

matrix manipulations, test takers have 25 minutes to answer 30 items. For the number series completion, test takers have 15 minutes to answer 40 items. For mathematical problem solving, test takers have 30 minutes to answer 25 items. A corrected score is created for each test part. For two of the three test parts, the corrected score is calculated by subtracting $1/4^{th}$ the number incorrect from the number of correct answers. The corrected scores are then equated and summed. This total equated score can range from -15 to 104.

Job performance. Supervisory performance ratings are based on goals that each employee sets with their manager. Managers evaluate employee performance according to how these preset goals are achieved. As part of the annual performance evaluation procedures, each employee's performance is rated by her or his manager on a 4-point scale with one indicating the highest level of performance. An unsatisfactory rating is the lowest point on this scale. This rating scale is standard and used by all managers providing evaluations. For each employee, the three most recent performance ratings are kept on file. For individuals who were no longer employed, performance ratings were gathered for the last three years that each person was employed with the company. A job performance rating was computed by averaging the available job performance ratings. The average performance ratings ranged from 1 to 3.33, with a mean of 2.16 (SD = 0.49). The analyses will control for the number of performance ratings available when appropriate. Additionally, an average of the available performance ratings was computed.

Tenure. The length of time each individual had been employed with the company was also gathered. To compute the length of tenure for each employee, the date of hire was subtracted from the date the data were collected. In years, tenure had a mean of 4.78 (SD = 3.03), with a range of .01 to 8.85.

Turnover. A measure of employment status was also computed for each employee. If a separation date was available, the employee was coded as 1, if currently employed with the organization the employee was coded as 2. Those employees that were no longer employed are identified as having either voluntary or involuntary turnover. Of the 1,618 no longer employed, 1,080 (66.7%) had voluntarily left the company.

Level of Promotion. The level of promotion was determined by calculating the difference between the position level when each employee was hired and the current or highest level attained.

Salary increase. Salary increase was calculated using the difference in the starting salaries for each employee and comparing them with the current or highest salary achieved.

Training courses. The number of training courses completed by each employee was obtained from organizational records. The training topics cover a wide range of topics from developing managerial skills to specific computer technology courses. The data does not identify if individuals received certificates of completion for the training courses.

Awards received. The number of awards received by each employee was obtained from organizational records. These awards include any patents issued.

Table 1 provides the frequency, means, and standard deviations for the criteria measures. When both predictor and criterion data sets were combined, the resulting sample included 3,001 cases. The overall sample mean aptitude test score was 57.81 (*SD*=16.05), with a range of 7.00 to 98.00.

Table 1

Frequency, Means, and Standard Deviations of Criteria

Criteria	*N*	Mean	Std. Deviation
# of invention awards*	232	4.45	6.75
# of patent awards*	91	2.92	4.06
# of training courses	3,176	12.88	10.37
Performance rating average	2, 785	2.16	0.49

*Excludes those with 0 awards received

Analyses

To determine the relation between aptitude test scores and job performance, multiple regression procedures were used. The aptitude test score was regressed on supervisory job performance ratings, number of awards received, number of training courses, salary increase, and level of promotion. This analysis was used determine the variance that can be accounted for by changes in test scores. In addition, exploratory factor analyses were conducted to determine whether there was overall job performance factor from the individual performance variables.

Additionally, an exploratory analysis was conducted using multiple regression procedures. Type of turnover (voluntary/involuntary) was expected to moderate the relationship between aptitude test scores and job performance such that those cases with voluntary turnover are expected to have the strongest relationship. Any relationship between employee performance and aptitude test scores may have important implications for organizations. Weighted effects codes were created for the turnover variable as suggested by West, Aiken, & Krull (1996). Interactions variables were created by multiplying the centered aptitude test scores by the

computed weighted effects codes for turnover. Significant betas for the interactions will indicate moderation. The standard of evidence for this study will be $p \leq .05$.

Before any analyses were conducted, the available data were screened to determine that the variables were in an appropriate format. First, univariate descriptive statistics were examined for accuracy. The performance criteria variables and test score data were reviewed individually for univariate outliers. Twenty-six outliers were found and were excluded from subsequent analyses. The data were also examined for missing data. For the aptitude test score variable, 16 cases were missing data. Given the small percentage of data that were missing for this variable, nothing was done to replace these missing values and the participants were dropped. A larger amount of missing data was found in the three variables that measured supervisory ratings of job performance. For the first job performance rating 243 cases were missing. 630 were missing for the second rating. 1,048 were missing for the third rating. Because no pattern to these missing data values was observed nothing was done to replace these missing values. The number of performance ratings available for each case was computed to use as a control variable in the analyses as appropriate.

The variables were also examined for normality by reviewing the skewness, kurtosis, and histograms of the individual variables. Aptitude test score and all supervisory performance ratings fit a normal curve pattern. The tenure variable exhibited a U-shaped pattern with peak numbers at both the low and high ends of the distribution. Increase in salary, number of training courses, promotions and number of awards were each positively skewed. Given that the data was provided from organizational records, only those variables that exhibited extreme deviations from normality were transformed. Following the guidelines suggested by Tabachnick and Fidell (2001) the variables for number of training courses and number of awards received were each

corrected for substantial positive skewness by applying a logarithmic transformation. The variables for promotion and salary increase were each corrected for moderate positive skewness by applying a square root transformation.

Table 2

Descriptive Statistics for Included Variables

Variable	*N*	Mean	Standard Deviation	Skewness (Std. Error)	Kurtosis (Std. Error)	Min	Max
# of training courses	3001	12.47	10.31	1.86 (.04)	6.97 (.09)	0.00	106.00
Tenure (years)	3001	4.74	2.95	-0.06 (.04)	-1.62 (.09)	0.01	8.85
Promotions	3001	0.82	0.86	0.84 (.04)	0.40 (.09)	0.00	6.00
Salary increase	3001	1647.02	1504.94	0.93 (.04)	0.66 (.09)	0.00	9636.70
# of awards received	3001	5.23	6.32	2.67 (.04)	16.84 (.09)	0.00	82.00
Aptitude test score	3001	57.81	16.05	-0.01 (.04)	-0.47 (.09)	7.00	98.00
Performance rating average	2785	2.16	0.49	0.14 (.05)	-0.53 (.09)	1.00	3.33

*Note: Values are prior to any applied transformations

Results

Before testing the hypotheses, correlations were calculated to determine the strength of relationships between the cognitive ability test score and the available performance criteria. These are shown in Tables 3 and 4. Although the majority of the individual performance criteria are significantly related to aptitude test score, the strength of these relations is small in

Table 3

Correlation Matrix of Included Variables

	1	2	3	4	5	6	7	8	9
# training courses (1)									
Employment status (2)	.18**								
Tenure (3)	.27**	.82**							
Salary increase (4)	.28**	.70**	.87**						
# awards received (5)	.11**	.71**	.73**	.70**					
Test score (6)	-.02	.06**	.03	.12**	.10**				
Performance rating avg. (7)	-.07**	-.49**	-.50**	-.65**	-.50**	-.15**			
Performance rating sum (8)	.19**	.12**	.37**	.21**	-.01	-.09**	.26**		
Turnover (9)	.10**	.a	.29**	.11**	.10**	-.23**	.13**	.24**	
Promotions (10)	.23**	.68**	.79**	.85**	.64**	.09**	-.60**	.17**	.10*

* Correlation is significant at the 0.05 level 2-tailed.
** Correlation is significant at the 0.01 level 2-tailed.
a. Cannot be computed because at least one of the variables is constant.
b. Lower performance ratings indicate better job performance.

magnitude. Tenure, length of time with the organization, and number of training courses completed were the only variables that were unrelated to aptitude test score. Salary increase r = .12, promotions r = .09, and number of awards received r = .10 each had a significant correlation with aptitude test score. With the exception of turnover, the strongest correlations between aptitude test score and the available criteria were found for the supervisory performance ratings r = -.15.

Table 4

Zero Order Correlations of Aptitude Test Score with Performance Criteria

Performance Criteria	*r*	*n*
Tenure (days)	.03	3001
Turnover	-.23**	1618
Promotion level increase	.09**	2777
Salary increase	.12**	3001
# of training courses	-.02	2911
# of awards received	.10**	2273
# of invention awards	.10**	3001
# of patent awards	.08**	3001
Average performance rating	-.15**	2785

** Correlation is significant at the 0.01 level (2-tailed).

To test Hypothesis 1, that aptitude test scores will be positively related to supervisory ratings of job performance, promotions and salary increases, a regression analysis was conducted. Aptitude test score was the dependent variable with performance rating average, salary increase, and number of promotion levels as independent variables. Tenure was controlled

for by entering this variable as the first step in a sequential regression model. Results indicate that both steps in this regression model are significantly different from zero as shown in Table 5. In the first step of the regression F (1, 2587) = 7.44, $p < .01$ tenure was a significant predictor β = .054, $p < .01$. For the second step in this regression F (4, 2584) = 30.70, $p < .001$ all but one of the variables were significant predictors. Tenure β = -.246, $p < .001$, performance rating average β = -.056, $p = .03$, and salary increase β = .363, $p < .001$ were significant predictors of aptitude test score, but promotions β = -.046 did not reach significance. Approximately 5% of the variance in test scores was accounted for in this regression model.

Table 5

Hierarchical Multiple Regression of Job Performance Variables on Aptitude Test Score

Variables	*B*	*SE B*	β	*t*	R^2	Δ R^2
Step 1					.003	.003
Tenure	.001	.000	.054	2.73*		
Step 2					.045	.042
Tenure	-.004	.001	-.246	-6.62*		
Average Performance Rating	-1.81	.842	-.056	-2.14*		
Promotion transformed	-1.23	.969	-.046	-1.27		
Salary increase transformed	.291	.037	.363	7.85*		

*$p < .05$. $N = 2589$

To test the second hypothesis, that aptitude test scores will be positively related to the number of completed training courses, an additional regression analysis was conducted. Tenure was controlled for by entering this variable as the first step in a sequential regression model. Results indicate that the second step in this regression model is significantly different from zero as shown in Table 6. In the first step of the regression F (1, 2909) = 3.80, p = .051 tenure was not a significant predictor β = .036. For the second step in this regression F (2, 2908) = 3.02, p < .05 only one of the variables was a significant predictor. Tenure β = .044, p < .05 was a significant predictor of aptitude test score, but number of training courses β = -.029 did not reach significance. Although this regression model was significant, results indicate it explained only a negligible amount of variance in aptitude test score.

Table 6

Hierarchical Multiple Regression of Number of Training Courses on Aptitude Test Score

Variables	*B*	*SE B*	β	*t*	R^2	Δ R^2
Step 1					.001	.001
Tenure	.001	.000	.036	1.95		
Step 2					.002	.001
Tenure	.001	.000	.044	2.28*		
# of Training Courses Completed Transformed	-1.23	.824	-.029	-1.50		

*p < .05. N = 2911

To test the third hypothesis, that one overall job performance factor existed, a factor analysis was conducted. A principal factors extraction with varimax rotation was performed on the available performance criteria. Only one factor had an eigenvalue larger than one. This factor

accounted for 60% of the variance. As shown in Table 7, the variables that loaded onto this factor include salary increase, promotion level, tenure, number of awards received, and performance rating average. Number of training courses did not load onto this factor.

Table 7

Principal Factors Loadings of Job Performance Variables

	Factor 1
Salary increase	.946
Promotions	.848
Tenure	.848
# of awards received	.801
Average performance rating	-.597
# of training courses	.146

**Note*: Lower performance ratings indicate better job performance; Factor 1 accounts for 60.89% variance.

Based on the findings from the factor analysis, an addition regression was conducted with each of the performance criteria that loaded onto the identified factor entered as independent variables. Again, tenure was entered in the first step in the regression to control for length of time with the organization. Results indicate that the second step in this regression model is significantly different from zero as shown in Table 8. In the first step of the regression F (1, 2076) = 3.78, p = .052 tenure was not a significant predictor β = .043. For the second step in this regression F (5, 2072) = 20.90, p < .001 three of the five variables were significant predictors. Salary increase had the strongest association β = .354, p < .001, followed by tenure β = -.243, p < .001, and promotions β = -.075, p = .05. Neither performance rating average β = -.032 nor

number of awards received β = .056 reached significance. Approximately 5% of the variance in test scores was accounted for by this model.

Table 8

Hierarchical Multiple Regression of Job Performance Variables on Aptitude Test Score Based on Factor Analysis Results

Variables	*B*	*SE B*	*B*	*t*	R^2	ΔR^2
Step 1					.002	.002
Tenure	.001	.000	.043	1.95		
Step 2					.048	.046
Tenure	-.004	.001	-.243	-5.89*		
Average Performance Rating	-1.134	.986	-.032	-1.15		
# of awards transformed	2.096	1.30	.056	1.61		
Salary increase transformed	.315	.042	.354	7.45*		
Promotions transformed	-2.164	1.10	-.075	-1.97		

*$p < .05$. $N = 2078$

An additional exploratory analysis was conducted to determine if type of turnover (voluntary/involuntary) moderated the relationship between aptitude test score and job performance. It was expected that voluntary turnover would have a stronger relation than those with involuntary turnover. Performance rating average was the dependent variable, with centered aptitude test score, two weighted effects codes for turnover, and two interaction terms of test score and turnover were entered as one step in a regression. Results indicate that type of turnover

did not moderate the relationship between aptitude test score and job performance. Table 9 shows the regression model was significantly different from zero, F (5, 2754) = 196.43, $p <$ *.001*, with test scores β = -.101, $p < .001$, and the weighted effects codes for both voluntary β = .297, $p < .001$ and involuntary turnover β = .294, $p < .001$ making the contribution. This model accounted for 26.3% of the variance in performance rating average. The interaction terms were not significant predictors of supervisory performance ratings. Thus, type of turnover was not a moderator of test scores and job performance ratings.

Table 9

Hierarchical Multiple Regression to Examine Type of Turnover as a Moderator of Aptitude Test Score on Performance Rating Average

Variables	*B*	*SE B*	β	*t*
Aptitude test score	-.003	.001	-.101	-6.04*
Weighted effect code-voluntary turnover (C1)	.175	.011	.297	16.51*
Weighted effect code-involuntary turnover (C2)	.280	.018	.294	15.63*
Interaction of test score and C1	.000	.001	-.008	-.434
Interaction of test score and C2	-.001	.001	-.011	-.569

Note: $R^2 = .263$; $\Delta R^2 = .263$ *$p < .05$. $N = 2759$

Discussion

The present study sought to extend the use of cognitive ability to predicting long-term job performance. The results supported the expectation that aptitude test score was positively related

to job performance indicators; however, these results were not consistently significant. The number of training courses and tenure were the only performance variables with no significant correlation with aptitude test score.

The first hypothesis was partially supported. When the performance criteria of supervisory performance rating average, promotions, and salary increase were entered as predictors of aptitude test score, a significant model was observed. With tenure entered as a control variable, all criteria, except promotion, significantly contributed to variance in test scores. This model accounted for approximately five percent of the variance.

While aptitude test scores do not provide a complete picture of future employee performance, they can provide some insight into the probability of success on the job. The findings from this study are consistent with previous research (Hough & Oswald, 2000; Schmidt & Hunter, 1998, 2004; Viswesvaran & Ones, 2002) that has found a relationship between aptitude and job performance. However, the strength of these relations was smaller in magnitude than previous findings (Schmidt & Hunter, 2004). This smaller relation with long term job performance could possibly be indicative of the importance of other factors in maintaining job success. While cognitive ability is important to initial success, perhaps this declines as work experience is established. Additionally, once an employee is on the job the supervisor has more information on which to base performance ratings so other factors likely become more important.

The second hypothesis was not supported. Although the regression model was significant, no variance in aptitude test score was accounted for by the number of training courses. Although previous research has suggested that training performance does have a positive significant relationship with aptitude test score (Hough & Oswald, 2000; Schmidt & Hunter, 1998, 2004; Viswesvaran & Ones, 2002), this type of relationship does not extend to the

amount of training courses completed. This finding suggests that training course enrollment cannot be predicted based on aptitude. In regards to training, aptitude should only be used to predict training performance rather than the amount of training. Various extraneous factors likely affect the decision to take training courses. For example, if courses are offered against competing deadlines, interest in or use for the topic, or perceived likelihood of rewards and/or consequences would affect whether an employee participates in training courses. These considerations suggest type of training should be examined.

The third hypothesis was supported. One overall performance factor was identified as accounting for the majority of variance, 60.9%. The variables that loaded onto this factor include salary increase, promotions, tenure, number of awards received, and supervisory performance rating average. Number of completed training courses did not load onto this overall performance factor. This is not surprising given that there was no relationship between aptitude test score and the number of training courses. While previous research has examined performance criteria individually (Arvey & Murphy, 1998; Schimdt & Hunter, 1998, 2004), the results of this factor analysis suggest that groupings of performance criteria may be appropriate.

Based on the findings of the factor analysis, an additional regression was conducted with the variables that contributed to the identified factor input as the predictors of aptitude test score. This regression model was significant; however, the significant predictors were not consistent with those found in support of the first hypothesis. In both models, tenure and salary increase were significant predictors of aptitude test score. Although performance rating average was a significant predictor in support of the first hypothesis, this finding did not extend to the additional regression. Conversely, promotion was not a significant predictor in the regression conducted for hypothesis 1, but did reach significance in the additional model. The number of

awards received was included in the additional regression analysis, but was not a significant predictor of test scores. These results indicate the relationships between performance rating average and promotions with aptitude test scores is not as consistent as the relationships between tenure and salary increase with aptitude test scores.

An additional exploratory analysis was conducted to determine if type of turnover (voluntary/involuntary) moderated the relationship between aptitude test score and job performance. A moderator effect for turnover was not observed. Although this is contrary to the expectations, this result is positive from an organizational perspective. The lack of moderation indicates that the validity of the aptitude test is not different for those who have voluntarily left and those whose turnover was involuntary.

Although type of turnover was not a moderator, this analysis did produce several main effects. Aptitude test score was a significant predictor of performance ratings such that higher test score was indicative of better job performance. Additionally, the two weighted effects codes that were computed for turnover were both positive, significant predictors of performance rating.

Additionally, several of the individual performance criteria exhibited significant relationships with the aptitude test score. Supervisory performance ratings, salary increase, and turnover were the performance criteria that had the strongest relationship with test scores. The relationship between aptitude test score and performance rating was stronger for those cases with two or three ratings available than those with only one rating available. Promotion level increase also had a significant correlation with test score; however, the correlation was stronger between test score and salary increase. Another positive significant relationship was found between aptitude test score and the number of awards received. When looking at specific types of awards received, there was a slightly stronger relationship between the number of invention awards and

test score than the number of patent awards received. Tenure, or length of time with the organization, and number of training courses were not significantly related to test score. These findings indicate that cognitive ability test score can help predict successful job performance.

Limitations.

Although a significant relationship between aptitude test score and employee performance was observed, the strength of this relationship may have been weakened by the availability of data. Because archival records were used to complete this study the amount of control of the data was minimal. Also, the organizational records used in this study were recorded for purposes other than test validation. The available supervisory performance rating data was used for multiple purposes therefore; job performance may not have been the sole consideration when these ratings were assigned.

Research suggests that organizational data may not be appropriate for validation efforts (Robertson & Smith, 2001). One of the performance criteria analyzed in this study was supervisory ratings of performance. These are subjective ratings and may not be completely accurate reflections of performance. To minimize this potential for error a sum of three performance ratings was created, with the goal that multiple measures would lead to more accuracy for this performance rating. Often, the use of data that are gathered for research purposes only, that has no connection to organizational decisions (e.g., promotions), is advised.

The results from this study show that cognitive ability test score is useful for predicting job performance; however, this is often only one portion of any selection process. Many organizations use multiple methods to determine if a job applicant has a good probability of being a successful addition. Future research should examine how the combination of all tools

used during the hiring process predicts future job success. For example, given the widespread use of interviews, future research could examine how the combined results from cognitive ability tests and interviews could aid in the prediction of future successful job performance.

References

Arvey, R. D., & Murphy, K. R. (1998). Performance evaluation in work settings. *Annual Review of Psychology, 49,* 141-168.

Bobko, P., Roth, P. H., & Potosky, D. (1999). Derivation and implications of a meta-analytic matrix incorporating cognitive ability, alternative predictors, and job performance. *Personnel Psychology, 52,* 561-589.

Brannick, M.T. & Levine, E.L. (2002). *Job analysis: Methods research and applications*. Thousand Oaks, CA: Sage Publications.

Gatewood, R. & Field, H. (2001). *Human resource selection*. Orlando, FL: Harcourt Brace & Company.

Griffeth, R. W., Hom, P. W., & Gaertner, S. (2000). A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium. *Journal of Management, 26*(3)*,* 463-488.

Hough, L. M., & Oswald, F. L. (2000). Personnel selection: Looking toward the future—remembering the past. *Annual Review of Psychology, 51,* 631-664.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior, 29,* 340-362.

Hunter, J.E., & Schmidt, F.L. (1996). Intelligence and job performance: Economic and social implications. *Psychology, Public Policy, and Law, 2,* 447-472.

International Business Machines. (1984). *Development, validation, and use of the Information Processing Aptitude Test.* NY: Author.

Kehoe, J. F. (2002). General mental ability and selection in private sector organizations: A commentary. *Human Performance, 15*(1/2*),* 97-106.

Mount, M. K., Witt, L. A., & Barrick, M. R. (2000). Incremental validity of empirically keyed biodata scales over GMA and the five factor personality constructs. *Personnel Psychology, 53,* 299-323.

Murphy, K. R., Cronin, B. E., & Tam, A. P. (2003). Controversy and consensus regarding the use of cognitive ability testing in organization. *Journal of Applied Psychology, 88*(4*),* 660-671.

Outtz, J. L. (2002). The role of cognitive ability tests in employment selection. *Human Performance, 15*(1/2*),* 161-171.

Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than *g*. *Journal of Applied Psychology, 79*(4*),* 518-524.

Robertson, I. T., & Smith, M. (2001). Personnel Selection. *Journal of Occupational and Organizational Psychology, 74,* 441-472.

Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance, 15*(1/2*),* 187-210.

Schmidt, F. L., & Hunter, J. E. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology, 86*(1*),* 162-173.

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*(2*),* 262-274.

Schultz, D. P., & Schultz, S. E. (1998). *Psychology and work today*. Upper Saddle River, NJ: Prentice Hall.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics.* Boston, MA: Allyn and Bacon.

Tenopyr, M. L. (2002). Theory versus reality: Evaluation of *g* in the workplace. *Human Performance, 15*(1/2*),* 107-122.

Terpstra, D. E., Mohamed, A. A., & Kethley R. B. (1999). An analysis of federal court cases involving nine selection devices. *International Journal of Selection and Assessment, 7*(1*),* 26-34.

U.S. Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice. (1978). Uniform guidelines on employee selection procedures. *Federal Register, 43,* 38290-38315.

Viswesvaran, C. & Ones, D. S. (2002). Agreements and disagreements on the role of general mental ability (GMA) in industrial, work, and organizational psychology. *Human Performance, 15*(1/2*),* 211-231.

West, S. G., Aiken, L. S., & Krull, J. L. (1996). Experimental personality designs: Analyzing categorical by continuous variable interactions. *Journal of Personality, 64*(1), 1-48.

APPENDIX

SAMPLE QUESTIONS

Example matrix manipulation question:

	Column				
	1	2	3	4	5
Row 1	A	*	C	3	E
Row 2	E	A	*	C	3
Row 3	3	E	A	*	C
Row 4	C	3	E	A	*
Row 5	*	C	3	E	A

Q: What character would appear above the letter A in Row 2, if the characters in Row 1 were written in reverse order?

(A) A (B) * (C) C (D) 3 (E) E

Example number completion question:

Complete the following number sequence.

3 1 4 1 5 1 6 1

(A) 4 (B) 5 (C) 6 (D) 7 (E) 8

Example math reasoning question:

An office uses 2 kinds of forms: Deposit and Withdrawal. The office has a total of 1,200 forms. The number of deposit forms is twice the number of withdrawal forms. How many withdrawal forms are in the office?

(A) 300 (B) 400 (C) 600 (D) 800 (E) 900