Technical Report 1242

Assessing Professional Competence by Using Occupational Judgment Tests Derived From Job Analysis Questionnaires

Peter Legree
Joseph Psotka
U.S. Army Research Institute

Tiffany M. Bludau
Dawn Gray
Consortium Research Fellows Program
George Mason University

January 2009

United States Army Research Institute for the Behavioral and Social Sciences

U.S. Army Research Institute
for the Behavioral and Social Sciences

A Directorate of the Department of the Army
Deputy Chief of Staff, G1

Authorized and approved for distribution:

Michelle Sams

MICHELLE SAMS, PhD.
Director

Technical reviews by:

Richard Hoffman, U.S. Army Research Institute
Trueman Tremble, U.S. Army Research Institute

NOTICES

DISTRIBUTION: Primary distribution of this Technical Report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn DAPE-ARI-ZXM, 2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926.

FINAL DISPOSITION: This Technical Report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Technical Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy):	2. REPORT TYPE:	3. DATES COVERED (from. . . to)
January 2009	Final	06/01/2006 to 04/01/2008

4. TITLE AND SUBTITLE
Assessing Professional Competence by Using Occupational Judgment Tests Derived From Job Analysis Questionnaires

5a. CONTRACT OR GRANT NUMBER

5b. PROGRAM ELEMENT NUMBER
611102

6. AUTHOR(S)
Peter Legree, U.S. Army Research Institute
Joseph Psotka, U.S. Army Research Institute
Tiffany M. Bludau, Consortium Research Fellows Program, George Mason University
Dawn Gray, Consortium Research Fellows Program, George Mason University

5c. PROJECT NUMBER
B74F

5d. TASK NUMBER
2902

5e. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Institute for the Behavioral and Social Sciences
2511 Jefferson Davis Highway
Arlington, VA 22202-3926

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Institute for the Behavioral and Social Sciences
2511 Jefferson Davis Highway
Arlington, VA 22202-3926

10. MONITOR ACRONYM
ARI

11. MONITOR REPORT NUMBER
Technical Report 1242

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
Subject Matter Expert POC: Paul A. Gade

14. ABSTRACT *(Maximum 200 words)*:
Based on the historical success of job analysis questionnaires and the related expectation that respondents with technical expertise are required to obtain valid job analysis ratings data, we hypothesized that these questionnaires can be converted into judgment tests to measure individual differences in occupational expertise. As an initial test of this hypothesis, *Occupational Judgment Tests (OJTs)* were derived from job analysis questionnaires, and job incumbents were asked to objectively rate the frequency of job tasks and the importance of employee attributes to occupational performance. The OJTs required 3 minutes to complete, were administered to 302 job incumbents from four diverse occupations, and were scored using consensually derived standards and through factor analysis. As hypothesized, OJT consensus-based scores were valid against measures of incumbent job knowledge (ρ = .34 to .35), cognitive aptitude (ρ = .17 to .25), and career attitudes (ρ = .19). OJT factor scores were valid against career attitudes ($\bar{r}$ = .21 to .29). This method provides broadly sensitive and inexpensive measures of job competence that could expand the predictor and criterion space in personnel selection studies for many occupations.

15. SUBJECT TERMS
Situational Judgment Tests (SJTs), Job Analysis Questionnaires, Consensus Based Assessment, Performance Measurement

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	47	Diane Hadjiosif Technical Publication Specialist, 703-602-8047

Technical Report 1242

Assessing Professional Competence by Using Occupational Judgment Tests Derived From Job Analysis Questionnaires

Peter Legree & Joseph Psotka
U.S. Army Research Institute

Tiffany M. Bludau & Dawn Gray
Consortium Research Fellows Program
George Mason University

Basic Research Unit
Paul A. Gade, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926

January 2009

Army Project Number
611102.B74F

Personnel, Performance and Training

ASSESSING PROFESSIONAL COMPETENCE BY USING OCCUPATIONAL JUDGMENT TESTS DERIVED FROM JOB ANALYSIS QUESTIONNAIRES

EXECUTIVE SUMMARY

Research Requirement:

The U.S. Army must ensure that it continues to acquire, train, and utilize Soldiers to enable high levels of performance across a wide range of military occupations. Competent performance is highly dependent on job knowledge; conventional measures of job knowledge and situational judgment have been used to develop, maintain, and validate the U.S. Army personnel selection and classification system. However, these conventional measures can be prohibitively expensive to develop and time-consuming to administer. This project explored the hypothesis that job analysis questionnaires can be converted to judgment tests that measure individual differences in occupational expertise. This report focuses on the technical adequacy of these measures as indices of occupational competence.

Procedure:

The research team modified job analysis questionnaires that had been developed for four military occupational specialties (MOSs) to create corresponding *Occupational Judgment Tests (OJTs)*. The OJTs were administered to 302 Soldiers who were assigned to four diverse MOSs, and the scales were scored using consensually derived standards and through factor analysis. The OJT scores were validated against theoretically relevant criteria including measures of job knowledge, cognitive aptitude, and career attitudes.

Findings:

The OJTs used test administration time very efficiently, requiring between 3 and 4 minutes to complete. The OJT consensus-based scores correlated with measures of job knowledge ($\rho = .34$ to .35), cognitive aptitude ($\rho = .17$ to .25), and career attitudes ($\rho = .19$). OJT factor scores correlated with the career attitudes measure ($\bar{r} = .21$ to .29). These results show that the OJT method can provide valid measures of job competence for many occupations. In addition, the OJTs can be inexpensively developed by modifying job analysis questionnaires.

Utilization and Disseminations of Findings:

This approach promises to provide judgment tests for a wide variety of military occupations at minimal cost. This method has the potential to guide the development of new predictors of occupational performance as well as the refinement of occupational criteria. The results support the view that judgment is central to occupational performance and may provide better matching between applicants and MOS requirements. The approach is being used to create judgment scales that predict ROTC performance.

ASSESSING PROFESSIONAL COMPETENCE BY USING OCCUPATIONAL JUDGMENT TESTS DERIVED FROM JOB ANALYSIS QUESTIONNAIRES

CONTENTS

LIST OF TABLES

INTRODUCTION

We propose and evaluate the radical concept that job analysis questionnaires can be easily and simply converted into *Occupational Judgment Tests (OJTs)* in order to measure knowledge and expertise that is usually acquired through professional experience across a wide array of occupations. This expectation is broadly based on common practice and the widespread belief that respondents with technical expertise (i.e., job incumbents and technical experts) are essential to obtain valid job analysis data (Brannick & Levine, 2002). Yet the converse of this reasoning also seems sensible: that professional competence and technical expertise can be measured by analyzing response patterns on these job analysis questionnaires. This has never been systematically explored.

Our approach also reflects the view that job performance and expertise are functions of declarative knowledge, procedural knowledge and skill, and motivation (Campbell, McCloy, Oppler & Sager, 1993). This view is based partly on meta-analysis findings that have confirmed that job knowledge is highly correlated with occupational performance, $\rho = .80$ (Hunter, 1986). We view measures of job knowledge and performance as highly redundant, if not quite synonymous. However, the use of objective work samples and even conventional knowledge tests to assess job knowledge and performance can be prohibitively expensive for many applications. Instead, we suggest that the OJT approach may be used to supplement these standard techniques (i.e., performance samples and knowledge tests) at minimal cost for many applications.

As detailed in the Method section, job analysis questionnaires can be altered to create maximal performance tests. These alterations are remarkably minimal. As a concrete example, we used standard job descriptions—which were pre-existing for Army occupations and were provided by employment specialists—and placed them in the context of a simple instruction of one or two sentences. The rating instruction asked respondents for objective psychophysical judgments of task frequency and the importance of employee knowledge, skills, and abilities (KSA). Not only can such scales be created efficiently and simply, but they also extend job analysis theory and methods by adding a novel converging step of verification: the correlation between job knowledge, cognitive aptitude, and attitudes with these new kinds of judgment tests.

Structurally and semantically, these scales resemble situational judgment tests (SJTs) because a general situation is specified and followed by a list of options for the respondents to assess. (Refer to Table 1 for a description.) While most SJTs have been developed by analyzing tremendous amounts of critical incident data (Weekley, Ployhart, & Holtz, 2006), these judgment tests were created with minimal effort (less than one day per scale) by modifying existing job analysis questionnaires. These scales also differ from conventional SJTs in that each item requires encoding only a few words, thereby minimizing test administration time and confounds with reading ability.

We used consensually derived standards to score the OJTs and assess individual differences in the capacity of respondents to provide sensible ratings (Legree, Psotka, Tremble, & Bourne, 2005). We validated these measures against conventional measures of job knowledge, cognitive aptitude, and career attitudes. While empirical or expert-based scoring procedures have the potential to refine the scoring standards and improve the utility of these scales, we lacked the resources to explore these possibilities and leave these issues to future investigations. However, if there are limitations with the consensual scoring procedures, then we believe those limitations would tend to overestimate lower scoring individuals and underestimate higher scoring individuals, and so reduce the range and the potential correlations with our criterion measures. Therefore, our conclusions may understate the utility of these methods.

Because this proposal is unique and has never been formally described, we review theory and findings regarding job knowledge, situational judgment, job analysis techniques, and consensus based assessment (CBA) to describe the theoretical basis for the OJT method.

Knowledge, Occupational Performance, and Situational Judgment

Our expectations for OJTs were framed by theory, data, and common practices that have demonstrated the value of learning and knowledge to job performance. The classic theory of learning and performance, largely based on the theoretical framework developed by Edward L. Thorndike and his contemporaries (1935), proposed that (a) learning occurs in both formal and informal settings, and (b) knowledge resulting from learning bounds performance. Both the classic theory of learning and performance, as well as modern instructional design methods (e.g., Gagne, Briggs, & Wager, 1988), have proposed that learning principles, such as the laws of exercise and effect, support learning declarative and procedural knowledge, as well as learning broadly defined attitudes, such as courage, sociability, and dependability. These perspectives are largely consistent with the view that job performance is a function of declarative knowledge, procedural knowledge and skill, and motivation, especially if motivation partially reflects relevant attitudes (cf. Campbell, McCloy, Oppler, & Sager, 1993).

According to this general view, if a worker has not acquired the relevant knowledge, then the individual cannot respond appropriately in a specific situation, regardless of whether the knowledge should have been acquired formally or informally, or would have corresponded to formal facts, specific procedures, or general attitudes. Hunter (1986) confirmed this view through meta-analysis by showing that job knowledge is highly correlated with performance, $\rho = .80$.

However, even this .80 correlation may understate the relationship between knowledge and performance because the source studies may have overemphasized the role of declarative knowledge and may not have been sufficiently sensitive to the role of tacit knowledge, which tends to be procedurally oriented, informally acquired, or that relates to general attitudes and motivation. Often this type of knowledge is poorly documented and difficult to assess (Brown & Duguid, 1991). We view occupational performance as highly dependent on knowledge, and we suggest that the OJTs in this research primarily assess procedural knowledge and relevant attitudes. We propose that while declarative knowledge may be directly assessed using conventional job knowledge tests, OJTs may be better suited to assess knowledge that underlies procedural performance, career continuance attitudes, and employee retention.

Meta-analysis has demonstrated that situational judgment tests (SJTs) are correlated with both job performance, $\rho = .34$, and general cognitive aptitude, $\rho = .46$ (McDaniel, Morgeson, Finnegan, Campion, & Braveman, 2001). SJTs are widely viewed as assessing job knowledge, although much of this knowledge is believed to be acquired through occupational experiences and reflections upon those experiences. This informal job knowledge is sometimes described as tacit knowledge (Schmidt & Hunter, 1993; Sternberg & Wagner, 1993). According to this view, high levels of motivation may enhance the development of tacit knowledge by supporting reflections upon job-related experiences. Sound judgment also has been identified as the critical basis of much expertise in the human factor literature (Weiss & Shanteau, 2003); therefore, these theoretical frameworks support the general inference that judgment tests provide another method to assess occupational knowledge and competence. However, conventional SJTs have been expensive to develop and time-consuming to administer (McDaniel & Nguyen, 2001). This reasoning led us to evaluate methods that might create judgment tests at reduced cost.

We coined the term *Occupational Judgment Test (OJT)* to differentiate these instruments from conventional SJTs. At a conceptual level, OJTs are similar to SJTs because both types of measures reference work-related situations and require examinees to evaluate options. However, our OJTs were created by modifying existing survey instruments, while most SJTs have been based on the collection and analysis of waves of critical incident data (McDaniel & Nguyen, 2001). Moreover, most SJTs have presented work-related problems as item scenarios with options that consist of actions proposed to resolve the problem; the OJTs created for this research reference only the general work environment with options corresponding to job tasks and employee attributes. OJTs also differ from most conventional SJTs in that they have minimal reading and encoding requirements and require little administration time. Yet while OJTs have much in common with SJTs, their production reflects a logical extension to job analysis methods and theory. In fact, the job analysis procedure requires informed opinions to provide valid results, and our approach was based on the hypothesis that the quality of these ratings has reflected individual differences in expertise and knowledge.

Job Analysis

The use of job analysis information dates to Münsterberg (1913), who believed that personnel psychologists could use scientific techniques to specify occupational knowledge and develop valid personnel management practices with this information. Münsterberg initially expected that personnel psychologists would be required to conduct job analyses in order to obtain accurate information describing occupations. However, he quickly recognized that technical experts and job incumbents (i.e., occupational practitioners) possessed professional knowledge and could provide insight into job requirements that personnel psychologists could only understand in crude terms by reviewing occupational descriptions and related information (Münsterberg, 1913, p. 123).

It is now widely accepted that respondents with professional competence and technical knowledge (i.e., subject matter experts and job incumbents) are required to obtain valid ratings on job analysis questionnaires (Brannick & Levine, 2002; Sanchez & Levine, 2002). Like most psychologists, we regard job analysis information as knowledge that is primarily obtained through occupational experiences and reflections upon those experiences. We also suggest that much of this knowledge is either critical to high levels of job performance (e.g., by allowing an employee to attend to critical tasks as required) or is incidentally acquired as a result of competent performance (e.g., by gaining exposure to and understandings across a wide array of relevant experiences). Much of this knowledge may be described as procedural and may carry implications for documenting employee motivations. Our view implies that incumbent understandings of this knowledge will be correlated with conventional indices of expertise, such as performance on job sample tasks and scores on conventional job knowledge tests, as well as measures of employee motivation. Our view also implies that performance on these OJTs reflects valid job knowledge and motivation.

Dierdorff and Wilson (2003) sought to identify practical implications associated with job analysis techniques by meta-analyzing job analysis data obtained from technical experts and job incumbents. Two conclusions from the meta-analysis are germane to the current project. First, when the interrater reliability of job analysis data was considered irrespective of the size of the analytic sample, job incumbents provided data that were more reliable, ($\bar{r}_{xx} = .77$), than data obtained from technical experts, ($\bar{r}_{xx} = .47$). However, the greater availability of incumbent data across the individual studies confounded this first conclusion, which led to the second conclusion when controlled. When the number of raters was equated through the Spearman-Brown formula, the mean weighted interrater reliability of incumbent ratings, ($\bar{r}_{xx} = .39$) was less than the reliability of expert ratings, ($\bar{r}_{xx} = .49$).

Beyond providing an assessment of current job analysis techniques, these results suggest that professional expertise and knowledge might be assessed by the quality of the ratings to job questionnaires that had been recast as OJTs, provided that appropriate scoring standards could be deduced. In addition, because job analysis procedures have been designed to specify occupational knowledge, we predicted that OJTs will provide a measure of occupational knowledge.

Consensus Based Assessment (CBA)

Consensual standards. The meta-analysis finding that technical experts have greater interrater reliability than individual job incumbents is consistent with the premises of consensus based assessment (CBA) theory (Legree, Psotka, Tremble, & Bourne, 2006). CBA is based on demonstrations that:

1. For many domains, the opinions of practitioners will become more consistent as these individuals reach higher levels of expertise. Thus, experts tend to be more consistent than journeymen, who are more consistent than initiates or novices.

2. Scoring rubrics can be computed by averaging opinion data from either large numbers of journeymen or smaller numbers of experts. This is possible because errors in opinion tend to be random over levels of expertise and not systematic for many domains.

3. Deviations in opinion from the scoring standard can be used to quantify individual differences in domain-related expertise.

CBA procedures have been used to evaluate responses on judgment tests that have been developed for many domains, for which it was impractical to develop traditional scoring rubrics using conventional reference materials. These domains have included: social intelligence (Legree, 1995), emotional intelligence (Mayer, Caruso, & Salovey, 1999), general intelligence (Legree, Martin, & Psotka, 2000), military leadership (Antonakis, Hedlund, Pretz, & Sternberg, 2002), driver safety (Legree, Heffner, Psotka, Martin, & Medsker, 2003), tacit leadership knowledge (Legree, Psotka, Tremble, & Bourne, 2005), and college performance (Cianciolo, et al., 2006). For these domains, consensually derived scoring standards have been highly consistent with expert-derived standards and have resulted in valid individual difference measures (Legree, Psotka, Tremble, & Bourne, 2005). In addition, the CBA method is consistent with demonstrations that accurate scoring standards for multiple-choice tests can be deduced from respondent data (Batchelder & Romney, 1988), and it has much in common with anthropological techniques formulated and used to identify cultural knowledge (Romney, Weller, & Batchelder, 1986).

Profile similarity. When judgment data are collected using Likert scales, a variety of algorithms may be used to assess the similarity of individual protocols to some standard (cf. Cronbach & Gleser, 1953). One approach to quantify individual differences is to correlate each set of respondent ratings with the scoring standard (i.e., item means). This scoring method controls effects due to response biases (e.g., respondents who use only one end of the scale may still obtain very high scores), and these scores are straightforward to interpret because individual differences correspond to the covariance between item responses and item means, with superior performance reflected by greater covariance. We refer to these values as component scores, and used this procedure as the principal method to quantify OJT performance.[1] Because this approach controls for respondent differences in the mean (elevation) and variance of respondent ratings, these scores can be characterized as *scale-reduced.* These scores are broadly consistent with principles of psychophysics because they remove variance based on individual differences in the mean or "modulus" of each person's judgments (Stevens, 1975).

However, expert, journeyman, and novice ratings also may differ in either central tendency or dispersion for at least some domains. For example, when all the alternatives on a judgment test item are truly poor and objectively merit low ratings, experts may be more likely than journeymen or novices to provide low ratings for all alternatives. Alternatively, when novices believe they lack the requisite knowledge, they may choose to provide moderate ratings to all the alternatives, thereby hedging their opinions but also exhibiting less variance in ratings than more knowledgeable respondents. It follows that scoring procedures that are sensitive to this information may provide additional explanatory power against various criteria, in comparison to scale-reduced component scores, and these measures might be described as *scale-loaded.*

Many judgment tests have been scored in ways that are sensitive to scale-loaded as well as scale-reduced information. These methods have included: distances between respondent ratings and a scoring standard corresponding to mean expert ratings (Wagner, 1987), and weighted, percent agreement methods (Mayer, Caruso & Salovey, 1999; Hanson & Borman, 1992). Because scientific consensus has not developed on scoring Likert-based judgment tests,

[1] These values can be computed by inverting the data matrix so that individuals correspond to columns (and items to rows) and then conducting a Q-factor analysis. The first set of component scores from a Principal Components Analysis will correspond to the set of product-moment correlations of each individual with the scoring standard (i.e., mean ratings), hence, the term component, or Q, score.

we realized that one general method to quantify this information is to factor respondent ratings, save the factor scores, and then utilize these values to quantify individual differences in use of the Likert scale. In addition to the factor scores, respondent scale central tendency (mean or elevation) and dispersion (standard deviation) scores provide supplementary metrics that quantify individual differences in use of the judgment scales (cf. Cronbach & Gleser, 1953).

From an information perspective, the factor score method utilizes information to which the component (correlation) scores are insensitive (i.e., differences in the variance or means of respondent ratings). Because factor scores are sensitive to differences in respondent mean, and variance, as well as the covariance among respondents, factor scores can be expected to correlate with the component scores and supplementary metrics. For the current project, we computed OJT factor, central tendency, and dispersion scores to supplement the component scores and determine whether they may account for additional criterion variance.

Current Project

One of the interesting implications from analyses of judgment tests created using CBA principles follows from the fact that many scales assessed with CBA have incorporated Likert scales to register responses. Because most people are most familiar with Likert scales being used for surveys or more subjective reasons, individuals often view test items that utilize a Likert response format as less “test-like” than standard multiple-choice items. Results have demonstrated that most respondents (up to 90%) categorized those scales as opinion surveys and not as tests (Legree, Martin, & Psotka, 2000; Legree, Martin, Moore, & Sloan, 2007).

We recognized that much opinion data could be analyzed using a judgment paradigm, and many opinion questionnaires might even be modified to create psychometrically sound judgment tests. Therefore, this project also addresses the more fundamental expectation that opinion questionnaires can be converted to judgment tests. While opinion questionnaires have been developed for many domains, job questionnaires are central to modern personnel practices, and the methods used to develop these scales have been carefully refined for nearly a century.

We therefore reasoned that OJTs could be created by recasting two types of surveys—job analysis questionnaires and employee attribute questionnaires—as judgment tests that would be sensitive to respondent differences. Job analysis questionnaires also could be recast into more traditional knowledge tests; for example, by asking which of two tasks was more frequent. We leave the comparison of these multiple forms to future research, but we believe that our proposed form (such as our example in Table 1) leaves the questionnaire structure intact and is easiest to create and answer, and thus ultimately the most useful.

We decided to use the list of tasks in job analysis questionnaires by asking incumbents to objectively rate these tasks on attributes such as frequency, importance, and trainability. Similarly, we modified the lists of the knowledge, skills, and abilities (KSAs) in employee attribute questionnaires, and asked incumbents to rate their degree of relevance for enabling employee productivity. The primary difference between the source questionnaires and the OJTs was that the questionnaires had asked for subjective ratings reflecting personal experiences, while the OJTs required objective ratings reflecting professional knowledge. We refer to these two types of OJTs as Job Analysis Tests (JATs) and Employee Attribute Tests (EATs). As outlined above, these scales may provide indices of procedural knowledge and attitudinal factors that underlie job performance and employee retention (cf. Campbell, McCloy, Oppler, & Sager, 1993).

Research Hypotheses

Our first and primary hypothesis is based on the understanding that the development of experientially based knowledge is closely associated with both expertise (Chi, Glaser, & Farr, 1988; Wagner & Sternberg, 1985) and high levels of occupational performance (Hunter, 1986). This hypothesis is consistent with the expectation that technical expertise is required to obtain valid job analysis ratings data (Brannick & Levine, 2002) and with the meta-analysis result that job analysis data collected from experts are more reliable than data collected from incumbents when the number of respondents is held constant (Dierdorff & Wilson, 2003). Finally, the hypothesis reflects common-sense expectations that the ability to prioritize tasks is critical to success in many professional domains.

> *Hypothesis 1: Occupational Judgment Test (OJT) component scores are positively correlated with job knowledge.*

Because general cognitive aptitude correlates with knowledge acquisition (Hunter & Schmidt, 1996; Jensen, 1980), occupational success (Schmidt & Hunter, 2004), and SJT performance (McDaniel et al., 2001), we theorize that performance on OJTs would correlate with general cognitive aptitude. This hypothesis also is based on the general expectation that knowledge will be *g*-loaded regardless of whether it is measured with conventional knowledge tests, SJTs, or OJTs.

> *Hypothesis 2: Occupational Judgment Test (OJT) component scores are positively correlated with general cognitive aptitude.*

The management of employee attrition is a continuing challenge for many organizations, including the U.S. Army, which sponsored this research. In recent years, 30 percent of first-term enlisted Soldiers have left the military prior to completing their service obligation (Putka, 2005). A meta-analysis on the effects of realistic job previews suggests that knowledge of one’s job leads to higher performance and lower attrition (Phillips, 1998). In addition, tacit knowledge theory (Sternberg & Wagner, 1993) implies, and general learning theory (Thorndike, 1935) suggests, that more motivated employees should acquire more knowledge, so there should be a relationship between knowledge and markers of occupational motivation. These considerations led us to hypothesize a relationship between performance on the OJTs and self-assessments quantifying attitudes that relate to career continuance, such as job satisfaction, career continuance expectations, and self-perceived competence.

> *Hypothesis 3: Occupational Judgment Test (OJT) component scores are positively correlated with career attitudes.*

Because scale-reduced and scale-loaded scores could be computed for our OJT databases, and because we had the opportunity to collect OJT data for multiple occupations to replicate findings (details follow), we seized the opportunity to explore implications associated with these methods by computing correlations among the OJT scale-reduced measures (component scores), scale-loaded measures (central tendency, dispersion, and factor scores), and the principal criteria (job knowledge, general cognitive aptitude, and career attitudes). While we did not have strong expectations regarding the overall preference of these scoring procedures, we suggest that if expertise generally improves the accuracy of psychophysical scaling, then these additional scoring methods may provide additional information to quantify performance on these judgment tests. Thus, we conducted exploratory analyses to determine whether scale-loaded measures derived from the OJTs (respondent factor, central tendency, and dispersion scores) would account for incremental variance in the dependent variables (job knowledge, cognitive aptitude, career attitudes) beyond the variance accounted for by the OJT component scores.

Approach

To test our research hypotheses and explore implications associated with different scoring procedures, we embedded the OJTs into an applied research project that had been designed to validate temperament and aptitude measures against a variety of criteria for the U.S. Army (Ingerick, Diaz, Putka, & Knapp, 2007). This project also provided job knowledge and career attitude data for these occupations and respondents. Cognitive aptitude data, which were obtained from Soldier recruitment records and correspond to Armed Forces Qualification Test (AFQT) scores, were merged into the occupational databases. This approach enabled OJT data to be collected for multiple occupations, thereby providing four replications to assess our hypotheses.

METHOD

Participants

OJT data were collected from 303 Soldiers from multiple occupations at four different military posts in the United States over a 6-month period. Moderate amounts of data were obtained from Soldiers assigned to the following four occupations: infantry, artillery, vehicle mechanic, and medic. Participants had been enlisted between 9 and 57 months; the mean time enlisted was 22.2 months, and the standard deviation of time enlisted was 9.6 months. The ranks of these Soldiers ranged from Private to Sergeant as follows: Private, 15%; Private First Class, 38%; Corporal, 43%; and Sergeant, 1%. Most participants were male, 95%; Caucasian, 73%; and Non-Hispanic, 84%. Minority groups included Hispanics, 16%; African Americans, 9%; Native Americans, 4%; and Asian Americans, 5%.

Occupational Judgment Tests (OJTs)

Scale construction. We adapted the Job Analysis Tests (JATs) from job analysis questionnaires that had been developed for the infantry, artillery, vehicle mechanic, and medic occupations. Each of the four JATs was specific to a single occupation and contained between 23 and 36 items that described common tasks in that single occupation. Although each of the occupations had a separate job analysis survey, we had access to only a single employee attribute questionnaire that had been developed to quantify the importance of employee knowledge, skills and abilities (KSAs) to performance in military occupations. This single Employee Attribute Test (EAT) contained 26 items that described KSAs common to many occupations.

Table 1 displays instructions, stems, and example items for two of the resulting instruments.

Table 1
Occupational Judgment Test Stems and Items

Job Analysis Items:

Instructions: Based on your experience, how frequently will each of the following tasks be performed monthly by Soldiers in your occupation at the E4/E5 level in a combat zone, not in garrison?

Please use the following scale to rate how frequently most Soldiers in your occupation perform each task. Be sure to answer each question even if you have never deployed to a combat zone. Record your rating next to each item.

1	2	3	4	5	6	7	8	9
Never Done				Occasionally Done				Very Often Done

___ 1. Secure the scene of a traffic accident
___ 2. Operate a roadblock or a checkpoint
___ 3. Supervise the establishment and operation of a dismount point

Employee Attribute Items:

Instructions: Use all your knowledge, experience and expertise to indicate how IMPORTANT the Army believes each of the following characteristics is to success in your occupation at the E4/E5 level in a combat zone.

Please use the following scale to rate the importance of each characteristic, and record your rating next to each item. Be sure to read the description of each characteristic and answer each question even if you have never deployed to a combat zone.

1	2	3	4	5	6	7	8	9
Not-at-all Important				Somewhat Important				Extremely Important

___ 1. Conscientiousness/Dependability. The tendency to be trustworthy, reliable, and willing to accept responsibility.

___ 2. General Cognitive Aptitude. The overall ability to understand information, identify problems & solutions, and learn.

___ 3. Emotional Stability**.** Acts rationally and displays a calm mood

To systematically create judgment tests from the job questionnaires that would be sensitive to individual differences, we modified the instruments in the following ways:

1. An objective point of reference was incorporated into each questionnaire so that respondents would report their understanding of task frequency and KSA importance from the perspective of junior-level employees working in their occupation, as opposed to summarizing their own idiographic experiences. The source questionnaires had incorporated self-referenced instructions in their construction.

2. A 9-point Likert scale was incorporated into all questionnaires to enable individuals to come closer to a number matching psychophysical scale (cf. Stevens, 1975). The source questionnaires had incorporated only 5-point Likert scales in their construction.

The employee attribute questionnaire was modified slightly—the only new requirement was the instruction that respondents judge the importance of KSAs against performance in the respondent's profession (i.e., one of the four occupations).

The following scores were derived from each of the OJTs. The OJT component scores were used to test the formal research hypotheses, while the factor scores and supplementary measures allowed exploration of the utility of OJT scale-loaded metrics.

Component scores. These scores were computed by transposing the data matrix in SPSS, conducting a principle components analysis, and adding the component scores into the initial database. These scores are equivalent to the correlation between each individual's set of ratings with the scoring standard, defined as the mean rating for each item. Component scores could not be calculated for respondents whose ratings did not vary over items within each scale; for these individuals, component scores were not computed, although the other, scale-loaded measures were calculated.

Factor scores. Orthogonal factor scores were computed to describe respondent ratings on each scale. Extraction of at least two factor scores was justified based on scree plots for each of the OJT databases, but to standardize the presentation across the four occupations and enable the synthesis of these results, we report results for only the first two factor scores for each OJT. For all databases, the first two factor eigenvalues were above the scree elbow. Across the OJT datasets, the range for the eigenvalues for the first two factors follows: Factor 1, 8.44 to 15.31; and Factor 2, 2.06 to 4.17.

Scale central tendency scores. For each respondent, central tendency scores were computed as the respondent mean rating for each scale. Subsequent analyses showed that this information was highly redundant with the Factor 1 scores for each OJT ($r > .99$).

Scale dispersion scores. Dispersion scores were calculated as the respondent standard deviation for each rating scale. This information was largely redundant with the Factor 2 scores as detailed below. The correlations between the Factor 2 scores and the dispersion (i.e., standard deviation) scores ranged from .41 to .79 across the OJT databases.

Dependent Variables

The following measures from the applied validation study were used to evaluate the research hypotheses. Detailed descriptions of these measures are in Ingerick, Diaz, Putka, and Knapp (2007).

Job knowledge measures. Job knowledge was assessed using knowledge tests that had been developed for each of the four occupations in an earlier project (Knapp, Sager, & Tremble, 2005). These scales contained between 39 and 58 multiple-choice items. Reliabilities for these scales were: Infantry, $r_{xx} = .64$; Artillery, $r_{xx} = .65$; Vehicle Mechanic, $r_{xx} = .86$; and Medic, $r_{xx} = .65$.

Armed Forces Qualification Test (AFQT). The AFQT is used within the military as a cognitive measure of general aptitude. The AFQT is derived from four Armed Service Vocational Aptitude Battery (ASVAB) tests: Word Knowledge, Paragraph Comprehension, Arithmetic Reasoning, and Mathematics Knowledge. The average Soldier AFQT score was 61 and the standard deviation was 19.5 points. Reliability for this measure was estimated at .92 (Welsh, Kucinkas, & Curran, 1990).

Career attitudes. This measure was derived from self-assessment ratings of: Army Satisfaction, Occupation Satisfaction, Occupational Fit, Perceived-Competence, Extent the Respondent's Chosen Occupation Exceeded Expectations at Enlistment, Career Intentions, and Attrition Cognitions. While these scales had been developed to understand personnel attrition (Ingerick, Diaz et al., 2007), our interest was simply to evaluate the relationship between the OJTs and this class of variables. Therefore, we factor analyzed the seven measures and used the Factor 1 scores as a general measure of career attitudes. The first factor accounted for 53 percent of the variance, and the first eigenvalue equaled 3.71, with all remaining eigenvalues less than .86.

Procedure

After providing consent to participate in the project, Soldiers completed the ability and performance measures in a standard order, which required up to 4 hours to complete. Because the aims of the present effort were secondary to the applied goals of the larger project, the final two tasks involved completion of the two OJTs designed for their occupation and this research. Participants first completed the Employee Attribute Test (EAT) and then the Job Attribute Test (JAT). The median OJT administration times were 3 minutes, 6 seconds for the EATs, and 3 minutes, 30 seconds for the JATs. Because Soldiers completed only the OJTs that had been designed for their own occupations, we had four separate datasets to test our hypotheses.

Analysis Approach

Our primary research goal was to determine if occupational expertise could be assessed using judgment scales that were created by modifying job analysis and employee attribute questionnaires. To test the research hypotheses, we correlated the OJT component scores with measures of job knowledge, cognitive aptitude, and career attitudes for each occupation. We then conducted a meta-analysis to average these correlations over the occupations. To estimate true-score correlations, we computed ρ-coefficients using the Hunter and Schmidt meta-analytic method (1990; 2004), which corrects the mean weighted correlation coefficients for attenuation due to low reliability. The ρ-coefficients allowed us to compare the validity of our OJTs to that estimated through meta-analysis for SJTs. We used a combination of correlation, meta-analysis, and regression techniques to explore validity implications associated with the use of scale-reduced and scale-loaded scores.

RESULTS

OJT Descriptive Statistics

Table 2 reports descriptive statistics (means, standard deviations, and reliabilities) for the OJT component, central tendency, and dispersion scores by occupation. The component and central tendency score reliabilities were coefficient alphas,[2] and the dispersion score reliabilities were stepped-up split-half correlations.

The OJT component score reliabilities were modest, (r_{xx} ranged from .51 to .69), except for the Infantry JAT, which was low (r_{xx} = .19). The mean weighted reliability estimates for the EATs ($\bar{r}_{xx}$ = .57) and the JATs ($\bar{r}_{xx}$ = .49) were considered acceptable for research purposes because the scales addressed broad knowledge domains, required only 3 to 4 minutes to complete, and could easily be lengthened to be made more reliable for applied purposes.

Because OJT component score reliability might be influenced by respondent tendencies to use only part of the rating scales, it is worthwhile to review the average values of the central tendency scores, which also correspond to the scale mean ratings and are listed in Table 2. Across occupations, mean ratings were 6.89 for the EATs and 5.79 for the JATs. In other words, respondents had a strong tendency to use the upper half of the rating scale: that is, values 5 through 9. These results are reasonable because the task and KSA lists were developed to be job-related. However, these findings also suggest that the OJT items tended to be "too positive," in the sense that the items would not have fully allowed incumbents to demonstrate their ability to accurately assess task frequency and KSA importance. These results suggest revising the OJTs to present a broader range of items and incorporate a larger response scale to allow individuals to register subtle differences in their understanding.

Research Hypotheses: Component Score x Knowledge, Aptitude, and Career Attitudes Criteria

The first section of Table 3 reports correlations between the OJT component scores, job knowledge, cognitive aptitude, and career attitudes for the infantry, artillery, vehicle mechanic, and medic occupations. Table 3 also reports the correlation between the two OJTs for each occupation.

---

[2] Computing component score reliability required converting the respondent rating distributions and the scoring key to z-scores. Then agreement for each item was computed as the squared difference between the respondent and scoring key z-scores (cf. Cronbach & Glaser, 1953). Respondent mean squared differences were inversely, yet perfectly, correlated with the component scores (r = -1.00), and the item squared differences were used to compute coefficient alpha.

To draw broad conclusions regarding the first three hypotheses and to compare these results to SJT coefficients reported by McDaniel et al. (2001), we used meta-analysis procedures to average the correlations reported for the four occupations. The last section of Table 3 reports mean weighted correlations, aggregated sample sizes, and significance levels for the mean correlations. In addition, ρ-statistics, which correct the mean correlations for attenuation of reliability, are reported to estimate true-score correlations. In the next two sections, we first summarize the occupation-level findings, and then review the meta-analysis results.

Occupation Level Analyses. The validities reported in Table 3 provided support for the first three research hypotheses. For three of the four occupations, OJT component scores correlated significantly with job knowledge. In addition, the non-significant correlation between the component score and job knowledge was in the expected direction. Hypothesis 1 was generally supported by the occupation-specific analyses: OJT component scores correlated with job knowledge.

For two of the four occupations, the OJT component scores correlated significantly with AFQT. In addition, almost all non-significant correlations were in the expected direction, although a single non-significant correlation was contraindicated, -.01. Thus, Hypothesis 2 was generally supported at the occupational level: OJT component scores correlated with cognitive aptitude.

The Employee Attribute Test (EAT) component scores were significantly correlated with career attitudes for the artillery and infantry occupations. In addition, all non-significant correlations between the OJT component scores and career attitudes were in the expected direction for the infantry, artillery, and medic occupations. While the values for the vehicle mechanic occupation were contraindicated, these values were not statistically significant and reflected smaller samples ($n = 47$ to 50). Therefore, Hypothesis 3 was partially supported: EAT component scores correlated with career attitudes for two of the occupations.

Table 2

Occupational Judgment Test (OJT) Descriptive Statistics and Reliability Estimates

	Employee Attribute Test (EAT)				Job Analysis Test (JAT)				
	Component Score	Central Tendency	Dispersion Score	(n)	Component Score	Central Tendency	Dispersion Score	(n)	# of items
Occupation	Mn (SD) r_{xx}	Mn (SD) r_{xx}	Mn (SD) r_{xx}		Mn (SD) r_{xx}	Mn (SD) r_{xx}	Mn (SD) r_{xx}		
Infantry	.51 (.28) .64	6.95 (1.48) .96	1.49 (.77) .91	82	.48 (.28) .19	5.97 (1.45) .92	2.02 (.72) .92	81	24
Artillery	.38 (.30) .57	6.66 (1.23) .95	1.37 (.64) .90	85	.51 (.29) .56	5.65 (1.42) .90	2.26 (.63) .76	78	23
Vehicle Mechanic	.34 (.30) .51	7.04 (1.30) .95	1.23 (.71) .89	51	.54 (.22) .65	6.80 (1.21) .94	1.72 (.70) .87	55	36
Medic	.53 (.21) .51	7.01 (1.10) .93	1.59 (.70) .90	66	.63 (.17) .69	4.67 (1.29) .93	2.37 (.49) .81	53	30
Weighted Mean Values	.45 .57	6.89	1.43	284	.53 .49	5.79	2.10	267	

Note. The EAT contained 26 items for all occupations.

Table 3
Correlations Among Occupational Judgment Test (OJT) Component Scores and Outcome Variables

	JK	AFQT	CA
		Infantry	
EAT – C	.19 (76)*	.26 (78)**	.21 (77)*
JAT – C	.27 (82)**	.15 (84)	.02 (81)
r EAT/JAT	.27 (81)**		
		Artillery	
EAT – C	.29 (78)**	.15 (80)	.21 (79)*
JAT – C	.21 (78)*	.22 (81)*	.14 (80)
r EAT/JAT	.37 (83)***		
		Vehicle Mechanic	
EAT – C	.08 (41)	.08 (47)	-.06 (47)
JAT – C	.20 (43)	.04 (50)	-.15 (50)
r EAT/JAT	.39 (51)**		
		Medic	
EAT – C	.25 (57)*	.20 (64)	.12 (63)
JAT – C	.08 (60)	-.01 (67)	.08 (65)
r EAT/JAT	.00 (65)		
		Synthesized Values: $\bar{r}$ (*n*), ρ	
EAT – C	.22 (252)***, .35	.18 (269)**, .25	.14 (266)**, .19
JAT – C	.20 (263)***, .34	.11 (282)*, .17	.04 (276), .05
$r_{(EAT/JAT)}$ =	.26 (280)***, .49		
$R_{(Criteria.EAT, JAT)}$ =	.27, .40	.19, .26	.14, .23
Shrunken R	.25	.17	.11

Note. * *p* ≤ .05, ** *p* ≤ .01, *** p ≤ .001 for one-tailed significance tests. Sample size, *n*, is in parentheses. *JK* = Job knowledge; *AFQT* = Armed Forces Qualification Test; *CA* = Career attitudes; *EAT – C* = Employee attributes test component score; *JAT – C* = Job analysis test component score; *r EAT/JAT* = correlation between employee attribute test and job analysis test.

Meta-analyses of OJT component score validity. Consistent with the statistical power of these four replications, support for the research hypotheses was generally strongest for the occupations with the larger sample sizes (i.e., infantry and artillery; n = 76 to 84), and weaker for the occupations with the smaller sample sizes (i.e., mechanic and medic; n = 41 to 67). In addition, the moderate reliabilities of the OJT component scores, as well as the job knowledge measures, would have depressed demonstrations of their significance. We therefore averaged the correlations for the four replications (i.e., occupations) in order to draw general conclusions.

The mean weighted correlations and inferential statistics presented in Table 3 provide strong support for hypotheses 1 and 2 and partial support for hypothesis 3 as follows. For both types of OJTs, the correlations between the component scores, job knowledge, and cognitive aptitude were statistically significant, p < .05, as hypothesized, although the OJT component score correlations with job knowledge were significant at much more stringent levels, p < .001.

In addition, the correlation between the EAT component scores and career attitudes was statistically significant, $p < .01$.

The demonstration that the EAT component scores correlated with career attitudes ($\bar{r}_{CA,\ EAT\ component} = .14$, $p < .01$, $\rho = .19$) contrasts with the non-significant correlations of career attitudes with the conventional job knowledge measures, ($\bar{r}_{CA,\ JK} = .01$, ns), and with AFQT, ($\bar{r}_{CA,\ AFQT} = -.09$, ns). Table 4 contains these values. These results show that unlike conventional job knowledge tests, the EAT assesses knowledge that relates to incumbent career attitudes, a relationship that is predicted by tacit knowledge and general learning theory (Sternberg & Wagner, 1993; Thorndike, 1935).

Table 3 reports modest true-score correlations between the OJT component scores and job knowledge ($\rho_{EAT,\ JK} = .35$ and $\rho_{JAT,\ JK} = .34$) and between the OJT component scores and AFQT ($\rho_{EAT,\ g} = .25$ and $\rho_{JAT,\ g} = .17$). The true-score correlation between the JAT and the EAT component scores was substantial, ($\rho_{JAT,\ EAT} = .49$), but far from unity, thereby indicating that the EATs and JATs corresponded to overlapping domains. The final rows of Table 3 provide multiple regression results based on the mean and corrected correlation estimates. The multiple correlation of job knowledge with the EAT and JAT component scores was equal to .27 based on the mean correlations and .40 based on the true-score correlations.

Implications. The results indicate that the OJT method has tremendous promise to support the practical goal of objectively measuring professional knowledge and expertise, at least when performance is scored as component scores. The true-score correlations between the OJTs and job knowledge ($\rho_{EAT,\ JK} = .35$ and $\rho_{JAT,\ JK} = .34$), although based on these very short, preliminary scales, were comparable to the meta-analytic validity estimates for SJTs, $\rho_{SJT,\ Job\ Performance} = .34$ (McDaniel, et al., 2001). The results also supported expectations that higher levels of OJT performance would be associated with positive career attitudes toward military service, as is predicted by the role of incumbent attitudes and motivations in theories of learning and performance (Thorndike, 1935; Hunter, 1986; Sternberg & Wagner, 1993; Gagne, Briggs, & Wager, 1988).

Table 4

Mean Weighted Correlations ($\bar{r}$), Inferential Statistics, and True-Score Correlations (ρ) Among the Occupational Judgment Test (OJT) Measures and the Criteria

	EC1	EDS	ECT	EF1	EF2	JC1	JDS	JCT	JF1	JF2	JK	AFQT
EC1	1											
EDS	.24(.34)***	1										
ECT	.20(.28)***	-.21(-.23)***	1									
EF1	.27(.36)***	-.15(-.16)**	1(1)***	1								
EF2	.63(.84)***	.65(.68)***	-.07(-.07)	0	1							
JC1	.26(.49)***	.22(.32)***	.08(.11)	.10(.14)	.17(.24)**	1						
JDS	.17(.25)**	.32(.37)***	-.02(-.03)	0(0)	.25(.27)***	.35(.55)***	1					
JCT	-.05(-.06)	-.13(-.14)*	.40(.43)***	.40(.41)***	-.01(-.01)	-.02(-.03)	-.09(-.10)	1				
JF1	-.03(-.05)	-.11(-.12)	.40(.41)***	.40***	0	0(-.01)	-.08(-.09)	1(1)***	1			
JF2	.18(.24)**	.25(.27)***	.03(.03)	.05	.20***	.71(1)***	.60(.65)***	0(0)	0	1		
JK	.22(.35)***	.13(.16)*	.02(.02)	.03(.04)	.22(.27)***	.20(.34)***	.08(.10)	-.01(-.02)	0(0)	.16(.19)**	1	
AFQT	.18(.25)**	-.01(-.01)	.03(.03)	.03(.03)	.08(.09)	.11(.17)*	.06(.07)	-.05(-.06)	-.05(-.05)	.09(.09)	.47(.58)***	1
CA	.14(.19)**	-.04(-.04)	.28(.29)***	.29***	.10	.04(.05)	.01(.01)	.21(.22)***	.21***	.03	.01(.01)	-.09(-.09)

Note. For mean correlations $\bar{r}$, * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$. n = 263-355. All coefficients represent 2-tailed tests except those involving OJT component scores against Job Knowledge, AFQT, and Career Attitudes. True-score correlations, (ρ) are in parentheses. *EC1* = EAT component score; *EDS* = EAT dispersion scores; *ECT* = EAT central tendency scores; *EF1*, *EF2* = EAT factor scores; *JC1* = JAT component score; *JDS* = JAT dispersion scores; *JCT* = JAT central tendency scores; *JF1*, *JF2* = JAT factor scores; *JK* = Job Knowledge Test; *AFQT* = Armed Forces Qualifying Test; *CA* = Career attitudes.

Comparisons Among OJT Component and Scale-Loaded Scores

To explore the utility of the scale-loaded metrics, we computed correlations among the OJT component scores, the OJT scale-loaded scores (Factor 1, Factor 2, central tendency, and dispersion scores), and the three criteria (job knowledge, AFQT, and career attitudes) for each occupation.[3] We then used meta-analysis techniques to average these correlations. The results from the meta-analysis are summarized in Table 4, which reports mean weighted correlations, significance statistics, and ρ-coefficients to estimate true-score correlations. We first describe correlations among the scale-reduced and scale-loaded scores to explore their interdependencies. We then explore correlations between the scale-loaded scores and the three criteria in light of these interdependencies.

Correlations among scale-loaded and scale-reduced scores. For both types of OJTs, the Factor 1 scores were very highly correlated with the respondent central tendency scores, ($\bar{r} \approx 1.00$). Because the OJT Factor 1 and central tendency scores were essentially redundant, conclusions regarding the Factor 1 and central tendency scores are interchangeable. The correlations between the Factor 1 scores and the component scores ranged from low to moderate, ($\bar{r}_{\text{EAT C, EAT F1}} = .27$; $\bar{r}_{\text{JAT C by F1}} = 0$). In addition, the correlations between the Factor 1 and dispersion scores were near zero ($\bar{r}_{\text{EAT F1, EAT DS}} = -.15$; $\bar{r}_{\text{JAT F1, JAT DS}} = -.08$). It follows that the Factor 1 (and central tendency) scores represent a source of variance that is largely separate from the component and dispersion scores.

For both types of OJTs, the Factor 2 and the component scores were highly correlated ($\bar{r}_{\text{EAT C, EAT F2}} = .63$; $\bar{r}_{\text{JAT C, JAT F2}} = .71$). In addition, the Factor 2 and dispersion scores were highly correlated ($\bar{r}_{\text{EAT DS, EAT F2}} = .65$; $\bar{r}_{\text{JAT DS, JAT F2}} = .60$). The true-score coefficients replicate this pattern. However, only moderate correlations were observed between the component and dispersion scores ($\bar{r}_{\text{EAT C, EAT DS}} = .24$; $\bar{r}_{\text{JAT C, JAT DS}} = .35$). All values are significant at $p < .001$. These coefficients suggest that the OJT Factor 2 scores represent a composite of the component and dispersion scores, which corresponds to two separate sources of information. It follows that correlations between the Factor 2 scores and the criteria, which are described below and listed in Table 4, largely reflect the separate and additive action of the component and dispersion scores.

Finally, moderate correlations ranging from .20 to .40 were observed between corresponding scores across the two types of OJTs, (e.g., the EAT and JAT Factor 2 scores correlated .20; $p < .001$). These parameters indicate that the EAT and JAT scale-loaded scores are assessing related dimensions, as was implied by the true-score correlation between the OJT component scores, ($\rho_{\text{EAT C, JAT C}} = .49$).

OJT scale correlations with job knowledge, AFQT, and career attitudes. In general, the correlations of the component scores with job knowledge and AFQT were more substantial than the correlations of the scale-loaded scores with these two criteria. Refer to Table 4. These results may suggest that the OJT component scores are preferable to quantify OJT performance. However, the scale-loaded scores showed very promising validities with career attitudes, and the

[3] Because factor scores are arbitrary in direction, some of these values were inverted so that OJT component and Factor 2 scores would be sensibly correlated (i.e., positively) and could be averaged. Results for the individual occupations are detailed in the Appendix, Tables A1 through A4.

results may explicate reports of low correlations between cognitive aptitude and performance on some judgment tests. Here are key findings:

1. The mean correlations between the OJT Factor 1 scores and career attitudes were much higher than those between the component scores and career attitudes ($\bar{r}_{EAT\ F1,\ CA} = .29$, $p < .001$, vs. $\bar{r}_{EAT\ C,\ CA} = .14$, $p < .01$; $\bar{r}_{JAT\ F1,\ CA} = .21$, $p < .001$ vs. $\bar{r}_{JAT\ C,\ CA} = .04$, ns). In other words, although support for the career attitudes hypothesis was marginal using the OJT component scores, this hypothesis was strongly supported using the OJT Factor 1 or central tendency scores.

2. The EAT dispersion scores correlated with job knowledge ($\bar{r}_{EAT\ DS,\ JK} = .13$, $p < .05$), although the JAT dispersion scores were not significantly correlated with job knowledge ($\bar{r}_{JAT\ DS,\ JK} = .08$, ns). The OJT dispersion scores did not significantly correlate with AFQT ($\bar{r}_{EAT\ DS,\ AFQT} = -.01$, ns; $\bar{r}_{JAT\ DS,\ AFQT} = .06$, ns).

3. OJT Factor 2 score validities against the three criteria were consistently lower than those reported for the component scores. This result is consistent with the observation that the OJT Factor 2 scores represent variance associated with the component and dispersion scores, (i.e., the lower dispersion score validities moderated the higher component score validities). This effect was minor for correlations involving job knowledge, but more substantial for cognitive aptitude. This observation is important because scoring algorithms that assess performance on judgment tests with scale-loaded information may minimize correlations with general cognitive aptitude, while maintaining relations with other criteria (cf., Sternberg & Wagner, 1993; Mayer, Caruso, & Salovey, 1999).

Multiple regression analyses. To explore further the relationship between the scale-loaded scores and the career attitude criterion, multiple regression analyses were conducted to determine if the scale-loaded metrics would account for incremental variance beyond the variance accounted for by the scale-reduced component scores against career attitudes for each occupation. In these analyses, component scores for a single OJT were entered in block 1, followed by step-wise selection of the following scale-loaded scores for that OJT: dispersion, Factor 1, and Factor 2 scores.

The regression analyses revealed that the EAT scale-loaded scores accounted for significant incremental variance in career attitudes beyond that associated with the EAT component scores for three of the four occupations. Likewise, the JAT scale-loaded scores accounted for significant incremental variance in career attitudes for two of the four occupations. Table 5 summarizes results from these five regression analyses. This effect was greatest for the artillery occupation, where the EAT factor and component scores accounted for 22 percent of the variance in career attitudes.

Table 5

Exploratory Regression Results: Variance in Career Attitude Accounted by Component Scores (Step 1) and Alternative Measures (Step 2)

	Step 1	Step 2	Variables entered in Step 2
Infantry – EAT	.05	.10*	EAT-F1
Artillery – EAT	.05	.22***	EAT-F1, EAT-F2
Mechanics – EAT	.00	$.10^{-}$	EAT-F1
Infantry – JAT	.00	$.07^{-}$	JAT-F1
Artillery – JAT	.02	.15**	JAT-F1, JAT-F2

Note. $^{-}p \leq .10$, $^{*}p \leq .05$, $^{**}p \leq .01$, $^{***}p \leq .001$.

Implications. Although the correlation and regression analyses were exploratory, they demonstrate that the OJT Factor 1 and central tendency scores have much potential to account for variance in career attitudes. These results also may reflect the fact that the OJTs contain many more positive than negative items, as was documented in Table 2. We speculate that the inclusion of a broader range of items in the OJTs would increase the validity of the OJT component scores against the career attitudes criterion.

GENERAL DISCUSSION

Theoretical Implications

At a very abstract level, job analysis techniques have been designed to specify general information and knowledge about an occupation. While this type of knowledge may not be necessary to enable job incumbents to complete well-defined tasks, broad understandings concerning the interdependencies of task and employee characteristics have been a hallmark of expertise and provide a solid foundation to excel in many domains (Chi, Glaser & Farr, 1988; Weiss & Shanteau, 2003). When conducting job analyses, it has been customary to survey job incumbents and technical experts for nearly a century (Brannick & Levine, 2002). From this perspective, it follows that the consistency of a respondent's understanding with knowledge gleaned from a job analysis should provide an index of that respondent's level of occupational knowledge. We structured the OJTs to quantify this consistency, and we provisionally interpret OJT performance as a measure of occupational knowledge that is largely acquired through professional experience.

To our knowledge, no past investigation has made direct use of job analysis questionnaires to create judgment tests in order to quantify expertise, especially not in the simple, expedient way we have made use of them in this research.[4] Based on theory and meta-analysis data, we proposed that job analysis questionnaires could be modified to create OJTs that would assess occupational knowledge and would be correlated with conceptually relevant criteria. Our hypotheses and expectations were broadly based on findings and conceptualizations

[4] Although the JATs are similar in form to the Psychology Tacit Knowledge scale (Wagner & Sternberg, 1985).

regarding SJTs (McDaniel et al., 2001), psychophysical scaling (Stevens, 1975), job analysis (Dierdorff & Wilson, 2003), and expertise (Chi, Glaser, & Farr, 1988). Largely for expediency, we used consensually derived standards to compute both scale-reduced, component scores and scale-loaded, supplementary scores (cf. Legree, Psotka, Tremble, & Bourne, 2006). We also explored the possibility that scale-reduced and scale-loaded scores could quantify information that would improve the utility of these judgment tests, and by implication, the utility of other judgment tests that have incorporated rating scales.

We had hypothesized that performance on the OJTs correlates with job knowledge, general cognitive aptitude, and career attitudes based on theory and data linking these concepts (Thorndike, 1935; Hunter, 1986; Sternberg & Wagner, 1993). While not every correlation between the OJT component scores and these criteria (i.e., job knowledge, cognitive aptitude, and career attitudes) was significant at the occupational level, all significant results were consistent with the research hypotheses. To broaden the breadth of our conclusions, we averaged the occupation-specific validities using meta-analytic techniques. The weighted mean correlations and inferential statistics conclusively demonstrated that the OJT component scores correlated with job knowledge, general cognitive aptitude and, under some circumstances, career attitudes. The OJT central tendency and Factor 1 scores had a consistent and substantial relationship with career attitudes.

Although the ρ-values between cognitive aptitude, job knowledge, and the OJT component scores were lower than might be expected for scales based on refined methodologies, these values are impressive given the exploratory methods that were adopted. In fact, the true-score correlations reported for the two classes of OJTs, $\rho_{JAT, JK} = .34$ and $\rho_{EAT, JK} = .35$, are very similar to those reported for SJTs based on meta-analysis, $\rho_{SJT, Job\_Performance} = .34$ (McDaniel et al., 2001). Moreover, when these values were regressed on job knowledge, the multiple correlation, R = .40, actually exceeded the SJT parameter. Finally, these correlations were obtained despite the samples being highly restricted in their range of military experience (i.e., mean time in the military was 22.2 months, and the standard deviation was 9.6 months). Our principal explanation for these results is that the OJTs assess occupational knowledge that is primarily obtained through professional experiences and reflections upon those experiences, as is most knowledge.

The correlations between the EAT and JAT component scores, ($\bar{r}_{EAT, JAT} = .26$; $\rho_{EAT, JAT} = .49$), and their consistent relationships with job knowledge and cognitive aptitude, suggest that the approaches address two related but separate knowledge domains. Given this moderate relationship, we expect comparable coefficients would be obtained by using other job analysis approaches, which are described below, to create additional OJTs—thereby creating a short battery of OJTs for specific occupations. It follows that a battery developed in this way would have more substantial validity against the type of criteria we used in this project. In sum, the meta-analysis results provided compelling support for the expectation that the OJT method can be used to measure occupational knowledge and related expertise efficiently.

The significant correlations between the OJT component scores and general cognitive aptitude were expected, based on demonstrations that job knowledge, performance, and cognitive aptitude are highly correlated (Hunter, 1986; Hunter & Schmidt, 1996). These findings provide conceptual support for our assertion that job knowledge and professional competence can be assessed with OJTs. Like SJTs and conventional knowledge tests, the results show the OJTs are *g*-loaded. This result is highly consistent with our supposition that the OJTs assess expertise and job-related knowledge, much like conventional SJTs and job knowledge tests.

One important difference between these OJTs and conventional job knowledge tests is reflected by their correlations with career attitudes. Unlike many conventional job knowledge tests, the OJT component and factor scores correlated with career attitudes. Conceptually this finding is consistent with general learning and tacit knowledge theory because learning, and especially the development of informally acquired knowledge, is dependent on high levels of motivation (Thorndike, 1935; Gagne, Briggs, & Wager, 1988; Sternberg & Wagner, 1993). In fact, the failure of many conventional job knowledge tests to correlate with career attitudes may reflect a fundamental limitation with these scales. Perhaps conventional knowledge tests have been too focused on formally acquired knowledge, and less sensitive to types of knowledge (such as procedural and tacit knowledge) that result in positive attitudes and relate to motivational factors. The OJT method may effectively address this limitation, thereby extending expectations regarding the importance of motivation to the development of knowledge and high levels of performance (Campbell, McCloy, Oppler, & Sager, 1993). In fact, from this perspective the OJTs provide a better indicator of a respondent's level of professional knowledge than do conventional job knowledge tests.

Thus from a theoretical level, our overarching framework justifying the deductive creation and expectations for the OJTs was supported. Much of our speculation had been derived from our understanding of job analysis procedures and reflected expectations that the knowledge underlying superior OJT performance would accrue from professional experience and reflection upon those experiences (Schön, 1983). While conventional job knowledge can be thought to address "how-to knowledge," the OJT scales might be described as assessing "knowledge of what is important." Because employees have discretion in the performance of their duties as well as responsibility for self-development, general knowledge concerning the importance of job tasks and employee characteristics may be of critical importance in determining employee productivity and career continuance. Knowing what to learn presupposes knowing what is important.

At a very broad level, the creation of these OJTs reflected the expectation that models of human performance exist for many domains that could be leveraged to create corresponding judgment tests. For example, general cultural models might be used to develop judgment tests to assess the ability either to interact with foreign cultures or to understand implications associated with American sub-cultures. Likewise, survey questionnaires have been developed for many domains that carry implications for professional performance. However, unlike the development of conventional SJTs, these judgment tests can be rapidly prototyped based on existing models and methods, and the scoring keys do not need to be coded separately because the responses can be scored consensually.

We believe the rationale for this investigation is strong and the results are compelling, which begs the question as to why this approach has not been previously adopted. We suggest this omission may have reflected:

1. Beliefs that opinion surveys cannot assess knowledge;
2. A lack of recognition that consensually derived scoring standards can be used to compute individual differences on judgment tests, at least provisionally and in lieu of more refined approaches; and
3. Failures to recognize that the quality of responses on job questionnaires may reflect individual differences in occupational knowledge and expertise.

Practical Implications for Scale-Loaded Scores

We had suspected that the scale-loaded scores might provide incremental utility in predicting at least some criteria. Support for these expectations is documented thoroughly by the correlations reported in Table 4, the multiple regression analyses summarized in Table 5, and the Appendix. In particular, the analyses show that in comparison to the other scale scores, the OJT Factor 1 (and central tendency) scores were substantially correlated with career attitudes, while the OJT component scores were more closely related to the job knowledge and cognitive aptitude criteria. It follows that the scale-loaded and scale-reduced scores have differential predictive utility against various criteria. Several implications regarding scale-reduced and scale-loaded scores merit attention.

High versus low stakes testing. The fact that component scores are not sensitive to respondent differences in rating mean or variance (i.e., central tendency and dispersion scores) may be important in high-stakes testing environments. Because scale-loaded scores are dependent on the central tendency and the dispersion of respondent ratings, these values may be manipulated by instructing respondents to alter their use of the rating scale. Simple distance scores also combine variance associated with the component, dispersion, and central tendency scores; this observation may be important due to simulations showing that these scores may be substantially manipulated by instructions to avoid extreme ratings (Cullen, Sackett, & Lievens, 2006). It follows that component scores may be preferable in high-stakes environments because they cannot be manipulated in this way.

However, in low-stakes test environments, the scale-loaded scores may provide a more comprehensive assessment of individual differences due to their sensitivity to the additional information. This is an important point that could be applied to improve the utility of many judgment tests that have incorporated ratings scales to collect respondent data. For these applications, performance on judgment tests might be decomposed into basic components and operationalized by the component and the supplementary, scale-loaded scores. These components could then be combined to optimize the utility of the instruments.

Loadings on psychometric g. The use of scale-loaded information may have inadvertently lowered the *g*-loadings of some SJTs. This result would occur when the dispersion or central tendency scores have lower correlations with cognitive aptitude than do the component scores. Therefore, when judgment tests are scored with a method that combines these sources of variance, the resultant scores may have minimal correlations with cognitive aptitude. We speculate that judgment tests, which have been scored in ways that are sensitive to these factors, may have inadvertently minimized their correlations with cognitive aptitude (cf. Mayer, Caruso, & Salovey, 1999; Sternberg & Wagner, 1993).

Enhancing OJT Psychometric Characteristics

At the outset of this project, we were not certain that job analysis questionnaires could be converted to judgment tests and provide reliable individual difference measures. It had been suggested that job incumbents might provide ratings that would not vary sufficiently, might not know enough to provide sensible ratings, might know too much to allow meaningful differences, or might be confused by a 9-point rating scale. Yet the mean weighted reliability estimates of the OJT component scores ($\bar{r}_{xx}$ = .57 & .49) suggest that this approach has much promise, especially as an initial test of this approach using scales that required approximately 3 minutes each to complete.

A fundamental limitation of these particular OJTs was the small number of items in each scale and the narrow range of their ratings (see Table 2). Because each JAT was derived from a job analysis questionnaire for the corresponding occupation, the items were all relatively high frequency tasks, so the variance in responses was curtailed. Likewise, the EATs primarily contained KSAs that were rated as moderately to highly important for each occupation. Increasing this variance may improve the OJTs' predictive power, for both the component scores and the factor scores. For the JATs, one way to accomplish this goal might be to combine items across related occupations, thereby broadening the breadth of items and allowing respondents to demonstrate their occupational knowledge. Likewise, one possible improvement in the EAT scales would be to broaden the attributes listed beyond the common attributes, to include more specific attributes appropriate to each occupation.

Both scales might be improved by increasing the size of the rating scale to allow respondents to register subtle differences in opinion as is suggested by psychophysical scaling principles (Stevens, 1975). Although a 9-point scale may seem excessive, many incumbents used only half the scale, thereby reducing its functional range to 5 points. In earlier work, we incorporated a much larger Likert scale in five judgment tests to enable respondents to register very subtle differences in their understandings (Legree, Martin, & Psotka, 2000). On those tests, the rating scale appeared continuous, although the responses were scanned and coded on a 59-point scale. Table 6 illustrates the JAT items in this format. Although of similar length, the median reliability of those five scales was substantially greater, r_{xx} = .76, than the reliabilities of the OJTs, $\bar{r}_{xx}$ = .49 to .57. We expect that more completely incorporating these psychometric principles while constructing OJTs would result in gains in test reliability without increasing administration time.

Because the OJTs exacted only a limited burden on administration time and were shorter than many questionnaires used to conduct job analyses, these instruments could easily be lengthened with additional items to improve their reliability. Finally, inaccuracies with the consensually derived standards may have limited the reliability and validity of these instruments.

Thus more sophisticated scales, as well as more carefully refined scoring standards, perhaps using elite groups of experts, appear likely to improve the utility of the OJT method.

Table 6
Occupational Judgment Test with "Continuous" Rating Scale

Job Analysis Items from Table 1 Revised as Unobtrusive Knowledge Tests:

Instructions: Based on your experience, how ***frequently*** *will each of the following tasks be performed monthly by Soldiers in your occupation at the E4/E5 level in a combat zone, not in garrison?*

Please use the following scale to rate how frequently most Soldiers in your occupation perform each task. Draw a dark  *on the spot to show your estimate. Use the entire scale.*

1. Secure the scene of a traffic accident

Never Done | Occasionally Done | Very Often Done

2. Operate a roadblock or a checkpoint

Never Done | Occasionally Done | Very Often Done

3. Supervise the establishment and operation of a dismount point

Never Done | Occasionally Done | Very Often Done

There are many other ways to structure job questionnaires that could be adapted to create OJTs. We used ratings of task frequency and KSA importance because the source questionnaires had used that format; we had no reason to believe that specific approach would be more useful than any other one. Variations to the JAT method also could ask for ratings regarding task trainability, consequences of poor task performance, safety issues, and interdependency of tasks. Modifications to the EAT could assess the utility of not only using a much wider range of KSAs, but also linking those KSAs to various facets of performance such as technical proficiency, citizenship behavior, or well-being (cf. Borman & Motowidlo, 1997; Kahneman & Krueger, 2006).

In fact, there may be as many ways to construct an OJT as there are ways to construct job analysis questionnaires. Given the enormous number of techniques that could be leveraged, we are certain that some of these methods would provide much more reliable and valid OJTs. We also suspect that many existing job analysis questionnaires may have been created to document relatively superficial aspects of jobs, as opposed to representing the complex knowledge structures of occupational practitioners and technical experts. OJT methods, which allow greater differentiation across levels of occupational expertise, also might be applied to refine job analysis methods in line with Benton Underwood's encouragements regarding the use of individual differences to refine theory (Underwood, 1975). This reasoning suggests a broad program of research to explore some of these many possibilities.

Conclusions

First, the OJT method supported the practical goal of objectively measuring occupational knowledge to assess professional expertise. Like most SJTs and job knowledge tests, these scales correlated with measures of cognitive aptitude; however, the OJTs also correlated with career attitudes and can be created and administered with minimal cost. While research is required to refine and optimize the OJT method, these results demonstrated that this approach has tremendous potential to enrich personnel selection projects.

Second, decomposing OJT performance into scale-reduced (component) and scale-loaded (factor, central tendency, and dispersion) scores provided insights into the conditions and purposes for which these sources of information may be most useful. While the composite scores were moderately correlated with job knowledge and cognitive aptitude and are ideal for high-stakes testing environments, the scale-loaded scores were more closely related to the career attitudes criterion, and provided insights to refine the measures.

Third, judgment tests may be deductively developed by leveraging existing theories, models, and methods for many realms of human behavior, thereby expanding the range and depth of cognitive abilities that can be objectively studied. In other words, the results suggest that many questionnaires, which have been developed to provide general information, could be modified to create maximal performance measures for those domains.

REFERENCES

Antonakis, J., Hedlund, J., Pretz, J., & Sternberg, R. J. (2002). *Exploring the nature and acquisition of tacit knowledge for military leadership* (Research Note 2002-04). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Batchelder, W., & Romney, A. K. (1988).Test theory without an answer key. *Psychometrika, 53,* 71-92.

Borman, W. C., & Motowidlo, S. J. (1997). Task performance and contextual performance: The meaning for personnel selection research. *Human Performance, 10,* 99-109.

Brannick, M. T., & Levine, E. L. (2002). *Job analysis: Methods, research, and applications for human resource management in the new millennium*. London: Sage Publications.

Brown, J. S., & Duguid, P. (1991). Organizational learning and communities of practice: Toward a unified view of working, learning and innovation. *Organizational Science, 2,* 40-57.

Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of job performance. In N. Schmidt & W.C. Borman (Eds.), *Personnel selection in organizations* (pp. 35-70). San Francisco: Jossey-Bass Incorporated.

Chi, M. T. H., Glaser, R., & Farr, M. J. (1988). *The nature of expertise.* Hillsdale, NJ: Erlbaum.

Cianciolo, A. T., Grigorenko, E. L., Jarvin, L., Gil, G., Drebot, M. M., & Sternberg, R. J. (2006). Practical intelligence and tacit knowledge: Advancements in the measurement of developing expertise. *Learning and Individual Differences, 16*, 235-253.

Cronbach, L. J., & Gleser, G. C. (1953). Assessing similarity between profiles. *Psychological Bulletin, 50,* 456-473.

Cullen, M. J., Sackett, P. R., & Lievens, F. (2006). Threats to the operational use of situational judgment tests in the college admission process. *International Journal of Selection and Assessment, 14*, 142-155.

Dierdorff, E. C., & Wilson, M. A. (2003) A meta-analysis of job analysis reliability. *Journal of Applied Psychology, Vol. 88*, 635-646.

Gagne, R. M., Briggs, L. J., & Wager, W. W. (1988). *Principles of instructional design*, 3rd edition. New York: Holt, Rinehart and Winston.

Hanson, M. A., & Borman, W. C. (1992). *Development and construct validation of a situational judgment test.* (PDRI Report #230). Minneapolis, MN: Personnel Decisions Research Institute.

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. (1990). *Methods of meta-analysis: Correcting error and bias in research findings.* Thousand Oaks, CA: Sage Publications, Inc.

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

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings.* (2nd ed.) Thousand Oaks, CA: Sage Publications, Inc.

Ingerick, M., Diaz, T., Putka, D., & Knapp, D. J. (2007). *Investigations into Army enlisted classification systems (Army Class): Concurrent validation report.* Alexandria, VA: Human Resources Research Organization (HumRRO).

Jensen, A. R. (1980). *Bias in mental testing.* New York: Free Press.

Kahneman, D., & Krueger, A. B. (2006). Developments in the measurement of subjective well-being. *Journal of Economic Perspectives, 20,* 3-24.

Knapp D. J., Sager C. E., & Tremble T. R. (2005). *Development of experimental army enlisted personnel selection and classification tests and job performance criteria*. (ARI Technical Report 116). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Legree, P. J. (1995). Evidence for an oblique social intelligence factor established with a Likert based testing procedure. *Intelligence, 21*, 247-266.

Legree, P. J., Heffner, T. S., Psotka, J., Martin, D. E., & Medsker G. J. (2003). Traffic crash involvement: Experiential driving knowledge and stressful contextual antecedents. *Journal of Applied Psychology, 88,* 15-26.

Legree, P. J., Martin, D. E., Moore, C. F., & Sloan, L. R. (2007, Fall). Personnel selection: An application of the unobtrusive knowledge test. *Journal of Business and Behavioral Science*, *16*, 4-15.

Legree, P. J., Martin, D. E., & Psotka, J. (2000). Measuring cognitive aptitude using unobtrusive knowledge tests: A new survey technology. *Intelligence*, *28*, 291-308.

Legree, P. J., Psotka, J., Tremble, T., & Bourne, D. (2005). Using consensus based measurement to assess emotional intelligence. In R. Schulze & R. D. Roberts (Eds.), *Emotional intelligence: An international handbook* (pp. 155-180). Berlin, Germany: Hogrefe & Huber.

Mayer, J. D., Caruso, D. R., & Salovey, P. (1999). Emotional intelligence meets traditional standards for an intelligence. *Intelligence, 27*, 267-298.

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

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

Münsterberg, H. (1913). *Psychology and industrial efficiency.* Boston: Houghton Miffin.

Phillips, J. M. (1998). Effects of realistic job previews on multiple organizational outcomes: A meta-analysis. *Academy of Management Journal, 41*(6), 673-690.

Putka, D. J. (2005). Composition and prediction of attrition through 48 months of service. In W. J. Strickland (Ed.), *A longitudinal examination of first-term attrition and reenlistment among FY99 enlisted accessions* (ARI Technical Report 1172, pp 45-78). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Romney, A. K., Weller, S. C., & Batchelder, W. H. (1986). Culture as consensus: A theory of culture and informant accuracy. *American Anthropologist*, *8*, 313-338.

Sanchez, J. I., & Levine, E. L. (2002). The analysis of work in the 20th and 21st centuries. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work and organizational psychology, Volume 1: Personnel psychology* (pp. 71-89). Thousand Oaks, CA: Sage Publications Ltd.

Schmidt, F. L., & Hunter, J. (1993). Tacit knowledge, practical intelligence, general mental ability, and job knowledge. *Current Directions in Psychological Science, 2,* 8-9.

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

Schön, D. A. (1983). *The reflective practitioner.* New York: Basic Books.

Sternberg, R. J., & Wagner, R. K. (1993). The *g*-ocentric view of intelligence and job performance is wrong. *Current Directions in Psychological Science, 2,* 1-5.

Stevens, S. S. (1975) *Psychophysics: Introduction to its perceptual, neural and social prospects.* Oxford, England: John Wiley & Sons.

Thorndike, E. L. (1935). *The psychology of wants, interests and attitudes.* New York: Appleton Century Crofts.

Underwood, B. J. (1975). Individual differences as a crucible in theory construction. *American Psychologist, 30*, 128-134.

Wagner**,** R. K. (1987). Tacit knowledge in everyday intelligent behavior. *Journal of Personality and Social Psychology*, 52, 1236-1247.

Wagner**,** R. K., & Sternberg, R. J. (1985). Practical intelligence in real-world pursuits: The role of tacit knowledge. *Journal of Personality and Social Psychology, 49*, 436-458.

Weekley, J. A., Ployhart, R. E., & Holtz, B. C. (2006). On the development of situational judgment tests: Issues in item development, scaling, and scoring. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 157-182). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Weiss, D. J., & Shanteau, J. (2003). Empirical assessment of expertise. *Human Factors, 45,* 104-114.

Welsh, J. R., Kucinkas, S. K., & Curran, L. T. (1990). *Armed Services Vocational Aptitude Battery (ASVAB): Integrative review of validity studies.* (AFHRL-TR-90-22) San Antonio, TX: Manpower and Personnel Division, Brooks Air Force Base.

APPENDIX

Correlations by Occupation

Table A1

Correlations Among the Occupational Judgment Test (OJT) Measures and the Criteria for the Infantry Occupation

	EC1	EDS	ECT	EF1	EF2	JC1	JDS	JCT	JF1	JF2	JK	AFQT
EC1	1											
EDS	.18	1										
ECT	.19	-.16	1									
EF1	.26*	-.11	1***	1								
EF2	.66***	.77***	-.07	.00	1							
JC1	.27*	.19	.15	.17	.28**	1						
JDS	.13	.40***	-.12	-.10	.36**	.27**	1					
JCT	.08	-.12	.57***	.57***	-.05	.03	-.09	1				
JF1	.09	-.11	.59***	.59***	-.03	.06	-.07	1***	1			
JF2	.31**	.26*	.06	.09	.34**	.81***	.51***	-.01	.00	1		
JK	.19*	.28*	-.04	-.01	.31**	.27**	.08	.03	.03	.23*	1	
AFQT	.26**	-.01	-.04	-.03	.07	.15	-.06	.00	.00	.06	.48***	1
CA	.21*	-.12	.22*	.22*	.06	.02	-.04	.26*	.26*	.00	-.12	-.11

Note. * $p \le .05$, ** $p \le .01$, *** $p \le .001$, $n = 76\text{-}89$. All coefficients represent 2-tailed tests except those involving OJT component scores against Job Knowledge, AFQT, and Career Attitudes. *EC1* = EAT component score; *EDS* = EAT dispersion scores; *ECT* = EAT central tendency scores; *EF1*, *EF2* = EAT factor scores; *JC1* = JAT component score; *JDS* = JAT dispersion scores; *JCT* = JAT central tendency scores; *JF1*, *JF2* = JAT factor scores; *JK* = Job Knowledge Test; *AFQT* = Armed Forces Qualifying Test; *CA* = career attitudes.

Table A2

Correlations Among the Occupational Judgment Test (OJT) Measures and the Criteria for the Artillery Occupation

	EC1	EDS	ECT	EF1	EF2	JC1	JDS	JCT	JF1	JF2	JK	AFQT
EC1	1											
EDS	.29**	1										
ECT	.34***	-.13	1									
EF1	.41***	-.07	1	1								
EF2	.66***	.53***	-.08	.00	1							
JC1	.37***	.12	.17	.19	.18	1						
JDS	.15	.12	.10	.12	.12	.48***	1					
JCT	.00	-.23*	.38***	.36***	-.05	-.09	-.05	1				
JF1	.00	-.24*	.38***	.37***	-.05	-.11	-.06	1***	1			
JF2	.26*	.13	.14	.16	.14	.88***	.72***	.02	.00	1		
JK	.29**	-.08	.07	.07	.16	.21*	.12	-.03	-.04	.17	1	
AFQT	.15	-.13	.07	.07	.08	.22*	.05	-.16	-.17	.18	.43***	1
CA	.21*	.05	.39***	.40***	.24*	.14	-.03	.23*	.24*	.04	.04	-.10

Note. * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$, n = 78-89. All coefficients represent 2-tailed tests except those involving OJT component scores against Job Knowledge, AFQT, and Career Attitudes. *EC1* = EAT component score; *EDS* = EAT dispersion scores; *ECT* = EAT central tendency scores; *EF1*, *EF2* = EAT factor scores; *JC1* = JAT component score; *JDS* = JAT dispersion scores; *JCT* = JAT central tendency scores; *JF1*, *JF2* = JAT factor scores; *JK* = Job Knowledge Test; *AFQT* = Armed Forces Qualifying Test; *CA* = career attitudes.

Table A3

Correlations Among the Occupational Judgment Test (OJT) Measures and the Criteria for the Vehicle Mechanic Occupation

	EC1	EDS	ECT	EF1	EF2	JC1	JDS	JCT	JF1	JF2	JK	AFQT
EC1	1											
EDS	.15	1										
ECT	.16	-.45**	1									
EF1	.18	-.43**	1***	1								
EF2	.30*	.50***	.00	.00	1							
JC1	.39**	.40**	-.10	-.08	.05	1						
JDS	.31*	.61***	-.28*	-.26	.20	.55***	1					
JCT	-.18	-.29*	.53***	.52***	.13	-.19	-.38**	1				
JF1	-.13	-.20	.52***	.51***	.17	-.07	-.24	.99***	1			
JF2	.26	.61***	-.13	-.11	.28*	.75***	.79***	-.15	.00	1		
JK	.08	.31*	-.06	-.07	.25	.20	.24	.10	.17	.32*	1	
AFQT	.08	.09	-.11	-.11	.01	.04	.22	-.18	-.16	.16	.40***	1
CA	-.06	-.21	.32*	.31*	-.10	-.15	-.11	.17	.17	-.03	.06	-.13

Note. * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$, $n = 41$-89. All coefficients represent 2-tailed tests except those involving OJT component scores against Job Knowledge, AFQT, and Career Attitudes. *EC1* = EAT component score; *EDS* = EAT dispersion scores; *ECT* = EAT central tendency scores; *EF1*, *EF2* = EAT factor scores; *JC1* = JAT component score; *JDS* = JAT dispersion scores; *JCT* = JAT central tendency scores; *JF1*, *JF2* = JAT factor scores; *JK* = Job Knowledge Test; *AFQT* = Armed Forces Qualifying Test; *CA* = career attitudes.

Table A4

Correlations Among the Occupational Judgment Test (OJT) Measures and the Criteria for the Medic Occupation

	EC1	EDS	ECT	EF1	EF2	JC1	JDS	JCT	JF1	JF2	JK	AFQT
EC1	1											
EDS	.31*	1										
ECT	.08	-.21	1									
EF1	.18	-.10	.99***	1								
EF2	.80***	.75***	-.12	.00	1							
JC1	.00	.23	.01	.03	.11	1						
JDS	.14	.25	.15	.20	.31*	.14	1					
JCT	-.16	.12	.11	.12	-.03	.15	.09	1				
JF1	-.16	.10	.10	.10	-.05	.10	.02	.99***	1			
JF2	-.14	.12	-.01	.00	.04	.36**	.41**	.09	.00	1		
JK	.25*	.07	.08	.10	.17	.08	-.10	-.13	-.11	-.07	1	
AFQT	.20	.06	.15	.17	.15	-.01	.09	.11	.12	-.04	.55***	1
CA	.12	.07	.20	.22	.11	.08	.20	.17	.17	.09	.07	-.01

Note. * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$, n = 57-91. All coefficients represent 2-tailed tests except those involving OJT component scores against Job Knowledge, AFQT, and Career Attitudes. *EC1* = EAT component score; *EDS* = EAT dispersion scores; *ECT* = EAT central tendency scores; *EF1*, *EF2* = EAT factor scores; *JC1* = JAT component score; *JDS* = JAT dispersion scores; *JCT* = JAT central tendency scores; *JF1*, *JF2* = JAT factor scores; *JK* = Job Knowledge Test; *AFQT* = Armed Forces Qualifying Test; *CA* = career attitudes.