Journal of Applied Psychology
1994, Vol. 79, No. 4, 518–524

Predicting Job Performance: Not Much More Than *g*

Malcolm James Ree, James A. Earles, and Mark S. Teachout

The roles of general cognitive ability (*g*) and specific abilities or knowledge (*s*) were investigated as predictors of work sample job performance criteria in 7 jobs for U.S. Air Force enlistees. Both *g* and *s* (the interaction of general ability and experience) were defined by scores on the first and subsequent principal components of the enlistment selection and classification test (the Armed Services Vocational Aptitude Battery). Multiple regression analyses, when corrected for range restriction, revealed that *g* was the best predictor of all criteria and that *s* added a statistically significant but practically small amount to predictive efficiency. These results are consistent with those of previous studies, most notably Army Project A (J. J. McHenry, L. M. Hough, J. L. Toquam, M. A. Hanson, & S. Ashworth, 1990). The study also extends the findings to other jobs and uses traditionally more acceptable estimates of *g*, application of effective sample size in cross-validation estimation, and new performance criteria.

Spearman (1904) proposed a two-factor theory of abilities, including general cognitive ability (*g*) and specific abilities (*s*). The relative importance of *g* and *s* in the prediction of criteria has been and remains the center of controversy (Calfee, 1993; Jensen, 1993; McClelland, 1993; Ree & Earles, 1992, 1993; Schmidt & Hunter, 1993; Sternberg & Wagner, 1993). The purposes of this study were to use traditionally accepted statistical estimates of *g*, to provide better estimates of validity by including effective sample size when computing multiple correlations corrected for restriction of range, and to extend previous research to new job performance criteria and additional jobs.

g and *s* Controversy

Early test developers, such as Binet and Simon (DuBois, 1970), were influenced by the concept of *g*, but eventually the influence of multiple ability theorists such as Thurstone (1938) was pervasive. This led to the development of multiple aptitude test batteries, such as the Differential Aptitude Tests (DAT), the General Aptitude Test Battery (GATB), and the Armed Services Vocational Aptitude Battery (ASVAB). These were designed to measure specific abilities and to make specific predictions about employment or educational success. Sets of test scores would be differentially selected or differentially weighted for each situation, fulfilling a proposal by Hull (1928) that specific abilities could compensate for or substitute for general ability. The different composites of subtests used by the military for job placement (Earles & Ree, 1992) or for interpreting score profiles in counseling (McDermott, Fantuzzo, Glutting, Watkins, & Baggaley, 1992) are current examples of the application of multiple ability theory. The use of differential weighting and different composites led to multiple aptitude theory being termed a theory of "differential validity" (Brogden, 1951).

Jensen (1980) identified *s* with specific experience rather than with specific ability. In this same vein, Cattell (1971, 1987) posited his investment theory, which proposes that initially there is a general ability (called *fluid g* or g_f) that is invested in specific experiences and crystallizes to specific skills (called *crystallized g* or g_c). This means that *s* is *g* modified by experience. It implies that, for an individual, the best estimate of *g* can be made from testing content in which the individual has invested his or her ability (g_c) or from tests that require little or no prior special experience from training, interest, motivation, or exposure (g_f; see also Jensen, 1992). An example of the former is that unsatisfactory estimates of *g* would be obtained by administering a French test to a sample, half of which has studied French and half of which has not. The estimates of *g* for the half that did not study French would be unsatisfactory; the estimates for the other half would be more satisfactory. To rectify this problem, Raven (1938), a student of Spearman, developed the Progressive Matrices Test, which measured *g* through a series of abstract diagrammatic problems (Vernon, 1960, p. 19) that did not require special investment of *g* but, rather, that *g* be used to solve nonverbal problems.

The primacy of *g* as a predictor has again become the subject of many studies. The December 1986 issue of the *Journal of Vocational Behavior* (Gottfredson, 1986) documented the renewed interest, as did the evidence emerging from validity generalization studies (Hunter, 1983, 1984a, 1984b, 1984c; Hunter, Crosson, & Friedman, 1985).

ASVAB and Validity of *g*

The ASVAB is a state-of-the-art aptitude battery (Jensen, 1985; Murphy, 1985) and an excellent source of data for investigating the value of *g* as a predictor, with about 1 million ad-

Malcolm James Ree, James A. Earles, and Mark S. Teachout, U.S. Air Force Armstrong Laboratory, Human Resources Directorate, Brooks Air Force Base, Texas.

The opinions expressed in this article are ours and are not necessarily those of the Department of the Air Force, Department of Defense, or U.S. government.

We thank A. R. Jensen, F. Schmidt, and H. Wainer for discussions and comments on this research topic.

Correspondence concerning this article should be addressed to Malcolm James Ree, Department of the Air Force, AL/HRMA, 7909 Lindbergh Drive, Brooks Air Force Base, Texas 78235-5352.

ministrations and about 200,000 selections to job training each year. Although several studies have investigated the incremental validity of *s* beyond *g*, using training grades as criteria, studies relating *g* to job performance criteria have typically not investigated the incremental validity of *s*.

Jones (1988) correlated the average validity of the ASVAB subtests for predicting training performance with the *g* saturation of the subtests. For each subtest, she corrected the training validities for range restriction and averaged them over 37 diverse air force technical training courses. These averages were subject weighted over 24,482 technical training students. For each subtest, the *g* saturation was measured by its loading on the unrotated first principal component (see Jensen, 1987; Ree & Earles, 1991b). Jones found a rank order correlation of .75, demonstrating a strong positive relationship between *g* and predictive validity. This was found across all jobs, and comparable values were found within four air force job families: Mechanical, Administrative, General (Technical), and Electronics. Following Jensen (1980), Ree and Earles (1992) corrected the *g*-factor loadings for subtest unreliability and calculated the Jones rank order correlation as .98.

Ree and Earles (1990) investigated the predictive utility of both the general and specific components of the ASVAB by regressing air force technical school grades on the unrotated principal-component scores of the ASVAB. Psychometric *g* was represented by the first principal component, and *s* was represented by the remaining principal components. Across 89 jobs (individual sample sizes ranged from 274 to 3,939), the average correlation of *g* and the training criterion was .76, corrected for range restriction. When the specific ($g \times$ Experience) components were added to the regressions, the multiple correlation increased an average of .02.

Using a linear models approach, Ree and Earles (1991a) evaluated the nature of the relationships of *g* and *s* to 82 air force job-training criteria. They found statistically significant but small contributions (an average gain of .02) for *s* to the regressions.

These three studies examined the predictive utility of *g* and the contribution of *s*, but none used job performance measures as criteria. Jones (1988) observed that although measures of job performance were the preferred criteria, they were frequently unavailable because of their cost. Several studies have used job performance measures as criteria in the evaluation of the predictiveness of *g*; however, only McHenry, Hough, Toquam, Hanson, and Ashworth (1990) examined the incremental validity of specific measures.

For example, psychometric *g* as measured by the Army General Classification Test (Stewart, 1947) was found to be related to the preservice occupations of soldiers serving in World War II. In as far as individuals sort themselves into jobs on the basis of their ability to perform, job incumbency becomes a form of job performance. Among the jobs with highest average-estimated intelligence were accounting, engineering, and medicine. Jobs with middling average-estimated intelligence were police officer, electrician, and meat cutter. Jobs with the lowest average-estimated intelligence included laborer, farm worker, and lumberjack. The distribution of within-job intelligence scores did not overlap for the very highest and the very lowest jobs. This study did not consider specific or invested abilities.

Additionally, Hunter (1986) reviewed hundreds of studies showing that *g* predicted job performance criteria, including training success, supervisory ratings, and content-valid hands-on work samples for both civilian and military jobs. However, direct tests of the incremental contribution of *s* for the prediction of job performance criteria were not made.

Finally, McHenry et al. (1990) evaluated *g*, defined as the sum of several simple weighted composites rather than as a principal component, principal factor, or hierarchical factor—the traditionally accepted statistical estimates of *g* (Jensen, 1980). They also evaluated noncognitive variables, such as temperament, vocational interest, and reward preference in an incremental validity design in Army Project A (McHenry et al., 1990). For predicting core job proficiency as measured by job performance, no predictor added more than .03 to the predictiveness provided by *g*, about the same increment found by Ree and Earles (1991a) for training grades.

In the current study, we replicated some aspects of McHenry et al.'s (1990) study by determining if measures of *g* and *s* were differentially (Brogden, 1951) useful predictors of job performance. We extend the results to several additional jobs as well as to a new criterion measure, the Walk-Through Performance Test (WTPT; Hedge & Teachout, 1992). This study also demonstrates the use of effective sample size in estimating multiple correlations corrected for the effects of range restriction.

Method

Subjects

The subjects were 1,036 non-prior-service enlistees in the U.S. Air Force who had entered from 1984 through 1988; had tested with ASVAB parallel forms 11, 12, or 13; had completed both basic military training and technical training; and were, for the most part, working in their first term of enlistment. They were predominantly White (78.1%), male (83.2%), 17- to 23-year-old graduates of high school or better (99.1%), with an average job tenure of about 28 months.

Predictors

The ASVAB is a multiple aptitude test battery composed of 10 subtests (see Earles & Ree, 1992) that is carefully crafted to represent major cognitive abilities, which facilitates the estimation of *g*. The ASVAB is used for enlistment qualification and initial job assignment in the armed services. The Numerical Operations and Coding Speed subtests are speeded, and the others are power tests. The ASVAB has been used in this subtest configuration since 1980. Its reliability has been studied (Palmer, Hartke, Ree, Welsh, & Valentine, 1988), and it has been validated for many military occupations (Earles & Ree, 1992; Welsh, Kucinkas, & Curran, 1990; Welsh, Trent, et al., 1990; Wilbourn, Valentine, & Ree, 1984).

In a reification of differential aptitude theory, the air force aggregates the subtests into Mechanical, Administrative, General, and Electronics composites (Earles & Ree, 1992).

There are three generally accepted ways of estimating *g* from a set of variables (Jensen, 1980). Ree and Earles (1991b) have shown that for the ASVAB, estimates of *g* from these three methods, principal components, principal factors, and hierarchical factor analysis all correlated greater than .996. Because of high correlations among the various *g* estimates and the mathematical simplicity of the principal components, the principal components were chosen to represent the *g* and *s* measures of the ASVAB. The first unrotated principal component serves as a measure of g (Jensen, 1980). Group factors from common-factors analyses often represent s, with g ineluctably distributed through them from rotation (Jensen, 1980). The g can be removed from the lower order factors through Schmid and Leiman's (1957) procedure. However, those theories that require common-factors procedures (Cattell, 1971, 1987; Vernon, 1960) do not account for all of the variance in the variables and put the specific variances at a relative statistical disadvantage in comparison with principal-components procedures, which do account for all of the variance and provide maximum advantage for s.

To determine the maximal predictive efficiency (Brogden, 1946) of s, one's best choice is the procedure that most fully represents the non-g portions. Therefore, we used the nine remaining unrotated principal components as the measures of s (s_1 to s_9). These are mathematically defined measures and do not necessarily represent identifiable or nameable psychological concepts. The principal components have the additional benefit of being orthogonal (Hotelling, 1933a, 1933b), which circumvents the problems of collinearity and enhances their usefulness in regression (Kendall, Stuart, & Ord, 1983).

Jobs

Data were collected for eight jobs as part of the joint-services (Wigdor & Green, 1991) and air force job performance measurement project (Hedge & Teachout, 1986, 1992). Each job had a minimum requirement on one of the four aptitude composites derived from the ASVAB. Jet Engine Mechanic and Aerospace Ground Equipment Mechanic were selected by the Mechanical composite; Information Systems Radio Operator and Personnel Specialist by the Administrative composite; Air Traffic Control Operator and Aircrew Life Support Specialist by the General composite; and Precision Measurement Equipment Laboratory Specialist and Avionic Communications Specialist by the Electronics composite. The Aircrew Life Support Specialist job was found not to be predictable by aptitude in this and in a previous study (Teachout & Pellum, 1991). Therefore, the Aircrew Life Support Specialist job was removed from the study.

WTPTs

The criterion measures used in the present study were hands-on performance tests (HOPTs) and interview work sample tests (INT), and their combination, a WTPT (Hedge & Teachout, 1992). For each job, there were tasks unique to the hands-on format, tasks unique to the interview format, and overlapping tasks measured by both formats. The hands-on criterion was the sum of all hands-on task scores for a job; the interview criterion was the sum of all interview task scores for a job; and the WTPT was the sum of all hands-on task scores and unique interview task scores for a job. Both formats required the examinee to accomplish the tasks, either manually or verbally, at the work site under the observation of a trained administrator. The HOPTs and INTs were constructed for each job to assess proficiency on representative job tasks. The task domains for each job were identified and defined from the Air Force Occupational Survey database (Christal, 1974). A domain sampling plan was developed, and tasks were selected with a stratified random sampling procedure (Lipscomb & Dickinson, 1988).

For each task, work sample developers used technical descriptions of work procedures (U.S. Air Force technical orders and manuals) as well as input from subject matter experts to define and describe the procedural steps required for successful task completion. The work sample tests were constructed for each task, reviewed by subject matter experts, and field tested at several air force bases.

The work sample tests were administered to the subjects and scored by active-duty or retired, noncommissioned officers with extensive job experience. The administrators received 1–2 weeks of observation and scorer accuracy training (Hedge, Lipscomb, & Teachout, 1988). Using videotapes of work sample test performance with known target scores, administrators discussed the key work behaviors to perform or avoid for successful task completion. These procedures have been shown to produce accurate and reliable work sample test ratings (Hedge, Dickinson, & Bierstedt, 1988). The raters demonstrated high average agreement ($r = .81$) and high average correlational accuracy ($r = .85$) between their ratings and videotape target ratings.

In addition, a shadow scoring technique was used during a portion of data collection on 58 subjects, requiring two test administrators to observe and rate task performance. The technique was effective in maintaining agreement in the scoring of the work sample tests. The average agreement of the scorer with the shadow scorer was 95% across the 58 subjects.

Procedures

Data collection. In a group orientation session, the research project was described and participation conditions were explained. Subsequent to the group session, job-incumbent subjects were individually administered both HOPTs and INTs. Time limits were specified for each WTPT, ranging from 4 hr to 7 hr, depending on the job.

Analyses. We used a linear-models approach. Two linear models were computed for each criterion. The first linear model used only g to predict the criterion, whereas the second used g and all of the s values to predict the criterion. This gives maximal advantage to the measures of specific ability; it was accomplished for the correlations artifactually depressed by prior selection. To make better estimates of the correlations in the unrestricted population, we also computed the regressions in matrices after multivariate correction for range restriction (Lawley, 1943). The Type I error rate was set at $p < .05$ for the regressions computed in the uncorrected data. No statistical tests were performed in the regressions on the basis of the corrected correlations.

We used Wherry's (1975) procedure to adjust the observed multiple correlations, the multiple correlations after correction for range restriction, and the multiple correlations computed after estimating the effective sample size for the increase in sampling variance due to correcting the correlations for range restriction (Schmidt, Hunter, & Larson, 1988). More information on the estimation of effective sample size can be found in the Appendix.

Failure to apply Schmidt et al.'s (1988) procedure would lead to an overestimate in the Wherry-adjusted correlations. Although often called a *correction for cross-validation*, Wherry's procedure actually provides better parameter estimates of the multiple correlation. It is particularly useful in theoretically oriented studies like ours because it effectively estimates the correlation as if the population regression weights were applied in the population. The sample-weighted differences between the range-restriction-corrected correlations of g with the criteria and the Wherry-adjusted, range-restriction-corrected correlations of g and s_1–s_9 with the criteria provide the most informative outcome measure for incremental validity. These values were computed both across jobs, to yield an average for the criterion type, and within jobs across the three criteria to yield a within-job average.

Results and Discussion

These analyses indicated that g and s_1–s_9 were useful in predicting the job performance criteria, as has been found for training criteria (Ree & Earles, 1990, 1991a). As represented by the 2nd through 10th principal components, s added to the accuracy of prediction, but only by a small amount. The efficiency of the predictors (uncorrected correlations, r and R) in this study was smaller than in a previous study in which technical training criteria were used (Ree & Earles, 1991a). The sample sizes in this study were much smaller, so some portion of the

Table 1
Results of Validity and Incremental Validity Analyses for the Jobs Studied

Criterion	N	r_g	r_g^{a}	R_{g+s}	R_{g+s}^{b}	R_{g+s}^{a}	${}^{wc}R_{g+s}^{a,b}$	N_{eff}	$R_{g+s}^{a,b,c}$
Air Traffic Control Operator									
HOPT	164	.127	.134	.328	.222	.391	.312	116	.267
INT	164	.104	.251	.302	.179	.406	.332	91	.268
WTPT	164	.141	.255	.311	.195	.407	.333	96	.279
Precision Measurement Equipment Laboratory Specialist									
HOPT	126	.343	.691	.497	.427	.780	.757	63	.730
INT	126	.322	.752	.406	.304	.774	.751	35	.689
WTPT	126	.355	.713	.503	.434	.793	.772	51	.732
Avionics Communications Specialist									
HOPT	74	.343	.717	.504	.369	.795	.757	30	.683
INT	74	.262	.607	.554	.443	.765	.720	39	.661
WTPT	74	.335	.713	.583	.484	.825	.793	37	.757
Aerospace Ground Equipment Mechanic									
HOPT	211	.294	.424	.457	.411	.608	.581	120	.558
INT	211	.191	.307	.357	.289	.478	.435	118	.405
WTPT	211	.262	.384	.432	.382	.570	.539	121	.513
Jet Engine Mechanic									
HOPT	178	.134	.251	.319	.219	.415	.350	106	.291
INT	178	.253	.429	.343	.254	.508	.462	82	.406
WTPT	178	.192	.354	.340	.250	.471	.418	93	.370
Information Systems Radio Operator									
HOPT	111	.222	.278	.413	.296	.476	.386	84	.375
INT	111	.319	.317	.480	.391	.549	.491	85	.454
WTPT	111	.274	.341	.442	.340	.520	.444	80	.405
Personnel Specialist									
HOPT	172	.220	.487	.323	.220	.538	.495	63	.390
INT	172	.112	.455	.308	.196	.549	.507	55	.402
WTPT	172	.206	.529	.345	.254	.595	.560	58	.465

Note. HOPT = hands-on performance test; INT = interview work sample test; WTPT = walk-through performance test; g = general cognitive ability; s = specific abilities or knowledge; r_g = correlation of g and the criterion; R_{g+s} = multiple correlation of g and s with the criterion; N_{eff} = effective sample size.
[a] Corrected for range restriction. [b] Wherry's (1975) correction for degrees of freedom was applied. [c] Effective sample size used in applying Wherry's (1975) correction.

increases resulting from s was likely to be the result of overfitting and, therefore, was likely to diminish on cross-validation. These regression results are reported in Table 1. The last column shows the adjusted correlations; these were estimated by Wherry's (1975) procedure, using effective sample size (Schmidt et al., 1988).

The regressions in the uncorrected correlations showed low to moderate correlations of g and s with the criteria. Results of the regressions in the corrected matrices, the better parameter estimates, were closer to previous findings indicating an increment of only .02 (McHenry et al., 1990; Ree & Earles, 1991a). The reason for the disparity between these corrected findings and the findings in the uncorrected (incorrect) correlations is the artifact of range restriction.

HOPTs. Across all jobs, the average observed validities of g and of g plus s_1–s_9 for the HOPT criterion were .229 and .394, respectively. Corrected for restriction of range, the average correlation of g and the criterion was .396. The average Wherry-corrected correlation coefficient estimate adjusted by Schmidt et al.'s (1988) effective sample size was .431 for the multiple correlation of g and s with the criteria.

INTs. Similar validity results were found for this criterion measure. The average observed validities were .209 and .370 for g and for g and s_1–s_9 as predictors. Corrected for range restriction, the average correlation of g and the criterion was .420, whereas the average correlation corrected for range restriction and the Wherry-adjusted correlation of g and s with the criterion was .427 when effective sample size was used.

WTPTs. The WTPT is a combination of the HOPT and the INT and provides a more thorough content sampling of the jobs' tasks (Hedge & Teachout, 1992) and higher reliability (Kraiger & Teachout, 1991). The average observed validity of g

Table 2
Averaged Results of the Validity Analyses by Job and Criterion

Measure	r_g[a]	R_{g+s}[b]	Δ[c]
	Averaged within 3 criteria within job		
Averaged within 3 criteria within job			
Air Traffic Control Operator	.213	.271	.058
Precision Measurement Equipment Laboratory Specialist	.718	.721	.003
Avionics Communications Specialist	.679	.700	.021
Aerospace Ground Equipment Mechanic	.371	.492	.121
Jet Engine Mechanic	.344	.350	.006
Information Systems Radio Operator	.312	.411	.099
Personnel Specialist	.490	.418	.000
Averaged across jobs			
Hands-on performance test	.396	.431	.035
Interview work sample test	.420	.427	.007
Walk-through performance test	.441	.462	.021

Note. All averages were weighted by effective sample size. g = general cognitive ability; s = specific abilities or knowledge.
[a] Correlation of g with criterion, corrected for range restriction.
[b] Multiple correlation of g and s with criterion, corrected for range restriction and using effective sample size when applying Wherry's (1975) correction for degrees of freedom. [c] The difference between a and b.

was .205, whereas the average for g and s_1–s_9 was .403. Corrected for range restriction, the correlation of g and the criterion became .441, and the average corrected and Wherry-adjusted correlation of g and s_1–s_9 with the WTPT was .462. These average coefficients are presented in Table 2.

In the current study, the difference between the average correlations of g corrected for range restriction and the criteria and the fully corrected multiple correlation of g and s_1–s_9 across all criteria and all jobs was an increment of .021. Adjusted by the Wherry (1975) procedure, the incremental validities of g plus s in the current study were very similar to those reported by McHenry et al. (1990, p. 343) for the job performance factors of Core Technical Proficiency and General Soldiering Proficiency. No increments reported by McHenry et al. exceeded approximately .03 for these job performance factors.

In previous studies (Ree & Earles, 1990, 1991a), when g and s_1–s_9 were used to predict training grades, it was found that g was the most potent predictor and that s added little to prediction. The same was true for predicting the job performance criteria in the current study.

The criterion that provided the greatest coverage of tasks and highest reliability was the WTPT. The average increment to g by measures of s was .021, about the same found in previous studies for both training criteria (Ree & Earles, 1991a) and job performance criteria (McHenry et al., 1990). It is also consistent with the estimate provided by Hunter and Hunter (1984). Finally, this difference is consistent with several related studies of incremental validity in performance prediction. Carey (1994) studied the predictive efficiency of adding new tests to a highly g-saturated test battery for the prediction of both job performance and training criteria; he found increments averaging .02 across these criteria. Likewise, Morales and Ree (1992) found similar incremental differences for predicting pilot and navigator performance that included work sample criteria. The consistency of results across these studies is remarkable.

Summary

The current study has extended the finding of statistically significant but practically small incremental validity for specific measures to seven additional jobs and to new criteria. It has also shown that the incremental value of the specific measures was small for all three criteria and has demonstrated the application of estimates of effective sample size in the computation of adjusted multiple correlation coefficients.

References

Brogden, H. E. (1946). On the interpretation of the correlation coefficient as a measure of predictive efficiency. *Journal of Educational Psychology, 37,* 65–76.

Brogden, H. E. (1951). Increased efficiency of selection resulting from replacement of a single predictor with several different predictors. *Journal of Educational Psychology, 11,* 173–196.

Calfee, R. (1993). Paper, pencil, potential, and performance. *Current Directions in Psychological Science, 2,* 6–7.

Carey, N. (1994, August). Computer predictors of mechanical job performance: Marine Corps findings. *Military Psychology, 6,* 1–30.

Cattell, R. (1971). *Abilities: Their structure, growth and action.* Boston: Houghton Mifflin.

Cattell, R. (1987). *Intelligence: Its structure, growth and action.* New York: Elsevier Science.

Christal, R. E. (1974). *The United States Air Force occupational research project* (AFHRL-TR-73-75). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Occupational Research Division.

Dubois, P. H. (1970). *A history of psychological testing.* Boston: Allyn & Bacon.

Earles, J. A., & Ree, M. J. (1992). The predictive validity of the ASVAB for training grades. *Educational and Psychological Measurement, 52,* 721–725.

Gottfredson, L. S. (1986). Foreword: The g factor in employment. *Journal of Vocational Behavior, 29,* 293–296.

Hedge J. W., Dickinson, T. L., & Bierstedt, S. A. (1988). *The use of videotape technology to train administrators of Walk-Through Performance Testing* (AFHRL-TP-87-71). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Training Systems Division.

Hedge, J. W., Lipscomb, M. S., & Teachout, M. S. (1988). Work sample testing in the Air Force job performance measurement project. In M. S. Lipscomb & J. W. Hedge (Eds.), *Job performance measurement: Topics in the performance measurement of Air Force enlisted personnel* (AFHRL-TP-87-58). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Training Systems Division.

Hedge, J. W., & Teachout, M. S. (1986). *Job performance measurement: A systematic program of research and development* (AFHRL-TP-86-37). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Training Systems Division.

Hedge, J. W., & Teachout, M. S. (1992). An interview approach to work sample criterion development. *Journal of Applied Psychology, 77,* 453–462.

Hotelling, H. H. (1933a). Analysis of a complex of statistical variables with principal components. *Journal of Educational Psychology, 24,* 417–441.

Hotelling, H. H. (1933b). Analysis of a complex of statistical variables with principal components (continued). *Journal of Educational Psychology, 24,* 498–520.

Hull, C. (1928). *Aptitude testing.* New York: World Book.

Hunter, J. E. (1983). *Validity generalization of the ASVAB: Higher validity for factor analytic composites.* Rockville, MD: Research Applications.

Hunter, J. E. (1984a). *The prediction of job performance in the civilian sector using the ASVAB.* Rockville, MD: Research Applications.

Hunter, J. E. (1984b). *The validity of the Armed Services Vocational Aptitude Battery (ASVAB) high school composites.* Rockville, MD: Research Applications.

Hunter, J. E. (1984c). *The validity of the ASVAB as a predictor of civilian job performance.* Rockville, MD: Research Applications.

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

Hunter, J. E., Crosson, J. J., & Friedman, D. H. (1985). *The validity of the Armed Services Vocational Aptitude Battery (ASVAB) for civilian and military job performance.* Rockville, MD: Research Applications.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors across studies. *Psychological Bulletin, 96,* 72–98.

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

Jensen, A. R. (1985). Armed Services Vocational Aptitude Battery. *Measurement and Evaluation in Counseling and Development, 18,* 32–37.

Jensen, A. R. (1987). Editorial: Psychometric *g* as a focus of concerted research effort. *Intelligence, 11,* 193–198.

Jensen, A. R. (1992). Commentary: Vehicles of *g. Psychological Science, 3,* 275–278.

Jensen, A. R. (1993). Test validity: *g* versus "Tacit Knowledge." *Current Directions in Psychological Science, 2,* 9–10.

Jones, G. E. (1988). *Investigation of the efficacy of general ability versus specific ability as predictors of occupational success.* Unpublished master's thesis, St. Mary's University, San Antonio, TX.

Kendall, M., Stuart, A., & Ord, J. K. (1983). *The advanced theory of statistics* (Vol. 3, 4th ed.). New York: Macmillan.

Kraiger, K., & Teachout, M. S. (1991). *Application of generalizability theory to Air Force's job performance measurement project: A summary of research results* (AFHRL-TP-90-85). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Training Systems Division.

Lawley, D. N. (1943). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh, Section A, 62, Part 1,* 28–30.

Lipscomb, M. S., & Dickinson, T. L. (1988). The Air Force domain specification and sampling plan. In M. S. Lipscomb & J. W. Hedge (Eds.), *Job performance: Topics in the performance measurement of Air Force enlisted personnel* (AFHRL-TP-87-58). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Training Systems Division.

McClelland, D. C. (1993). Intelligence is not the best predictor of job performance. *Current Directions in Psychological Science, 2,* 5–6.

McDermott, P. A., Fantuzzo, J. W., Glutting, J. J., Watkins, M. W., & Baggaley, A. R. (1992). Illusions of meaning in the ipsative assessment of children's ability. *Journal of Special Education, 25,* 504–526.

McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A, & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology, 43,* 335–354.

Mendoza, J. L., Hart, D. E., & Powell, A. (1991). A bootstrap confidence interval based on a correlation corrected for range restriction. *Multivariate Behavioral Research, 26,* 255–269.

Morales, M., & Ree, M. J. (1992, June). *Intelligence predicts academic and work sample training performance.* Paper presented at the annual meeting of the American Psychological Society, San Diego, CA.

Murphy, K. (1985). Armed Services Vocational Aptitude Battery. In D. J. Keyser & R. C. Sweetland (Eds.), *Test critiques* (Vol. 1). Kansas City, MO: Test Corporation of America.

Palmer, P., Hartke, D. D., Ree, M. J., Welsh, J. R., & Valentine, L. D., Jr. (1988). *Armed Services Vocational Aptitude Battery (ASVAB): Alternate forms reliability, Forms 8, 9, 10, and 11* (AFHRL-TP-87-48). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Raven, J. (1938). *Progressive matrices: A perceptual test of intelligence.* London: H. K. Lewis.

Ree, M. J., & Earles, J. A. (1990). *The differential validity of a differential aptitude test* (AFHRL-TR-89-59). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Ree, M. J., & Earles, J. A. (1991a). Predicting training success: Not much more than *g. Personnel Psychology, 44,* 321–332.

Ree, M. J., & Earles, J. A. (1991b). The stability of *g* across different methods of estimation. *Intelligence, 15,* 271–278.

Ree, M. J., & Earles, J. A. (1992). Intelligence is the best predictor of job performance. *Current Directions in Psychological Science, 1,* 86–89.

Ree, M. J., & Earles, J. A. (1993). *g* is to psychology what carbon is to chemistry: A reply to Sternberg and Wagner, McClelland, and Calfee. *Current Directions in Psychological Science, 2,* 11–12.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22,* 53–61.

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

Schmidt, F. L., Hunter, J. E., & Larson, M. (1988, August). *General cognitive ability vs. general and specific aptitudes in the prediction of training performance: Some preliminary findings.* Paper presented at the 96th Annual Convention of the American Psychological Association, Atlanta, GA.

Spearman, C. (1904). "General Intelligence," objectively determined and measured. *American Journal of Psychology, 15,* 201–293.

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.

Stewart, N. (1947). Scores of army personnel grouped by occupation. *Occupations, 26,* 1–37.

Teachout, M. S., & Pellum, M. W. (1991). *Air Force research to link standards for enlistment to on-the-job performance* (AFHRL-TR-90-90). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Training Systems Division.

Thurstone, L. L. (1938). *Primary mental abilities.* Chicago: University of Chicago Press.

Vernon, P. (1960). *Intelligence and attainment tests.* New York: Philosophical Library.

Ward, J. H., & Jennings, E. (1973). *Introduction to linear models.* Englewood Cliffs, NJ: Prentice Hall.

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

Welsh, J. R., Jr., Trent, T., Nakasone, R., Fairbank, B., Kucinkas, S., & Sawin, L. (1990). *Annotated bibliography of Armed Services Vocational Aptitude Battery (ASVAB) Validity Studies* (AFHRL-TR-89-

76). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Wherry, R. J., Sr. (1975). Underprediction from overfitting: 45 years of shrinkage. *Personnel Psychology, 28,* 1–18.

Wigdor, A., & Green, B. F., Jr. (1991). *Performance assessment for the workplace* (Vols. 1 and 2). Washington, DC: National Academy Press.

Wilbourn, J. M., Valentine, L. D., Jr., & Ree, M. J. (1984). *Relationships of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 8, 9, and 10 to U.S. Air Force technical school final grades* (AFHRL-TR-84-8). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Appendix

Schmidt, Hunter, and Larson's (1988) Formula for the Estimation of Effective Sample Size

The usual formula for the standard error (*SE*) of a correlation is

$$SE = (1 - r^2)/(n - 1)^{1/2},$$

and application of this would typically give a smaller standard error for a corrected correlation than for that correlation's uncorrected value. (This would be so because corrected correlations typically increase and larger correlations have smaller standard errors than do smaller correlations.) However, this is neither logical nor, for that matter, empirical (Mendoza, Hart, & Powell, 1992). Therefore, for the usual standard error formula to make sense it must be assumed that the sample size (in the denominator) for the corrected correlation is smaller than the sample size for its uncorrected counterpart. It is this effective sample size that we are trying to estimate.

If the usual standard error formula (*SE*) is merged with the linear regression equation, that gives $(1 - r^2) = (SSE/SST)$, where *SSE* is the sum of squares resulting from error and *SST* is the sum of squares total. We then have

$$SE_r = (1 - r^2)/(n - 1)^{1/2} = (SSE/SST)/(n - 1)^{1/2}, \quad \text{and}$$

$$SE_{rc} = (1 - r_c^2)/(n_{\text{eff}} - 1)^{1/2} = (SSE_c/SST_c)/(n_{\text{eff}} - 1)^{1/2}.$$

There is no reason to expect that the instability of the uncorrected correlation should not carry through to the corrected correlation. If we let the inaccuracy of the prediction in the corrected variance case equal that of the restricted variance case, that is,

$$(SSE/SST) = (SSE_c/SST_c),$$

we then have

$$SE_r(n - 1)^{1/2} = SE_{rc}(n_{\text{eff}} - 1)^{1/2}, \quad \text{or}$$

$$(n_{\text{eff}} - 1) = (SE_r^2/SE_{rc}^2)(n - 1). \qquad \text{(A1)}$$

Schmidt et al. (1988) proposed that this effective sample size be used with Wherry's (1975) correction. They substituted the usual formula for SE^2 and obtained

$$(n_{\text{eff}} - 1) = (1 - (r^2))^2/SE^2_{rc}, \quad \text{or}$$

$$n_{\text{eff}} = [(1 - r^2)^2 + SE^2_{rc}]/SE^2_{rc}. \qquad \text{(A2)}$$

If several estimates of r and r_c are available, then the mean of correlations would be used along with the empirical SE^2 of the r_cs.

If only one correlation is available, then (r^2/r_c^2) is used in Equation 1 in place of SE^2_r/SE^2_{rc}. This specifies that the endpoint of a confidence interval about r would be changed with the same proportional change made by correcting from r to r_c. Therefore, the equation for effective sample size can be rewritten as

$$(n_{\text{eff}} - 1) = (r^2/r_c^2)(n - 1),$$

or the computationally simpler

$$n_{\text{eff}} = [r^2(n - 1) + r_c^2]/r_c^2. \qquad \text{(A3)}$$

Received March 29, 1993
Revision received November 22, 1993
Accepted November 23, 1993 ■