

## PREDICTING TRAINING SUCCESS: NOT MUCH MORE THAN *g*

MALCOLM JAMES REE, JAMES A. EARLES  
Armstrong Laboratory  
United States Air Force

The roles of general ability (*g*) and specific abilities ( $s_1 \dots s_9$ ) were investigated in prediction of job-training-school grades. Subjects were 78,041 Air Force enlistees in 82 jobs. General ability and specific abilities were defined by scores on the first and subsequent unrotated principal components of the enlistment selection and classification test, the Armed Services Vocational Aptitude Battery. Linear models analyses revealed that  $s_1 \dots s_9$  added little to the prediction afforded by *g*. It was also determined that a common prediction equation for all jobs was almost as predictive as an equation for each job.

The concept of general cognitive ability or psychometric *g*, first proposed by Galton (1883), appeared in analyses early in this century (Spearman, 1904). It has been, and remains, the center of a controversy as to whether only general ability is sufficient or whether specific abilities are required to adequately predict criteria of importance, such as training success or job performance.

Early intelligence test developers, such as Binet and Simon, were proponents of *g* (see DuBois, 1970), but eventually the influence of multiple-ability theorists (Thurstone, 1938) was pervasive. This led to the development of multiple-aptitude batteries. The Differential Aptitude Tests (DAT), the General Aptitude Test Battery (GATB), and the Armed Services Vocational Aptitude Battery (ASVAB) were designed to measure specific abilities ( $s_1 \dots s_n$ ) and to make specific predictions about job or training performance. Sets of test scores could be differentially selected or differentially weighted for each situation, fulfilling a proposal by Hull (1928). An extreme position on this is called the hypothesis of situational specificity, which maintains that a specific set

---

The authors wish to thank three anonymous reviewers for extensive comments and insights, which have improved the paper. We also are indebted to W. Alley, J. Skinner, and L. Valentine of the Armstrong Laboratory, to Howard Wainer, Educational Testing Service, Inc., and to Arthur R. Jensen, University of California, Berkeley.

The views expressed herein are those of the authors and do not necessarily reflect those of the United States Air Force or the Department of Defense.

Correspondence and requests for reprints should be addressed to Malcolm James Ree, Armstrong Laboratory, Human Resources Directorate, HRMA, Brooks AFB, TX 78235-5601.

of tests should be differentially weighted for the same job, when applied across organizations or across time within the same organization (Schmidt & Hunter, 1984). The different composites of subtests used by the military for job placement and the interpretation of score profiles in counseling are current examples of the application of multiple-aptitude theory.

The use of differential weighting or different composites led to multiple-aptitude theory being termed a theory of differential validity (Brogden, 1951). The basic idea was that there were only specific abilities (Thurstone, 1938), or that lack of general ability could be replaced or compensated by specific abilities (Hull, 1928). Belief about the efficacy of specific abilities for validity became so widespread that Ghiselli (1973) was able to write that "there are few today who believe that success on each and every job can be predicted by any single type of test, or even by a small number of them" (p. 119). Following the proposals of Hull, the American military, one of the largest users of tests and one of the largest employers, still forms composites of from three to six subtests in the belief that specific abilities are superior in prediction to general ability.

However, there were dissenters, such as McNemar (1964), who found highly  $g$ -saturated tests to be the best predictors in 4,096 validity studies of the Differential Aptitude Tests. Schmidt and Hunter (1977) observed that the apparent differences in validity, frequently attributed to the effects of specific ability, were likely due to sampling and other sources of error. This led to a body of research from which they concluded that tests of general ability were valid for all jobs, for all job criteria, and that specific ability tests had little utility for prediction (Hunter, 1986). Further, Jensen (1984) arrived at the same conclusion: "For most jobs,  $g$  accounts for all of the significantly predicted variance; other testable ability factors, independent of  $g$ , add practically nothing to the predictive validity" (p. 101).

Support for this position was also provided by Thorndike (1985), who found that  $g$  measures played a central role in prediction of both training and job performance criteria, especially when cross validation was considered. He later observed that when sample sizes exceed approximately 100 to 200, "specialized weighting systems may begin to do better than a common factor score by a modest amount" (Thorndike, 1986, p. 338). Thorndike further reflected that pooling all jobs and calculating a universal set of weights would likely come quite close to creating a general factor score. Thus, Thorndike has addressed two issues: the role of  $g$  in prediction with large sample sizes, and pooling of data across jobs, which would compensate for small sample sizes per job. Linear models analysis provides an appropriate method to investigate these issues.

This study addressed Thorndike's issues with a large sample of jobs and subjects. The first research question was, with very large samples, how much predictive efficiency did measures of specific ability add to the prediction afforded by  $g$ ? The second research question was, how much predictive efficiency was lost by pooling all jobs in one regression equation?

### *Method*

#### *Subjects*

The subjects were 78,041 Air Force enlistees, recruited from 1984 through 1988, who had been tested with ASVAB parallel Forms 11, 12, or 13 and had completed both basic military training and a job training course. They were white (80%), male (83%), 17- to 23-years-old (86%), and high school graduates (99%). The sample sizes per job ranged from 274 to 3,930, with an average size of 952.

#### *Variables*

*Ability measures.* The Armed Services Vocational Aptitude Battery is a multiple-aptitude test (Department of Defense, 1984) used for selection and classification, and composed of the 10 subtests shown in Table 1. Except for the speeded Numerical Operations and Coding Speed, all are power tests. It is normed on a weighted, nationally representative sample of 18- to 23-year-old youths (Maier & Sims, 1986; Ree & Wegner, 1990). The battery has been used in this current content and subtest configuration since 1980. Its parallel forms reliabilities (Palmer, Hartke, Ree, Welsh, & Valentine, 1988) are presented in Table 1. It has been validated for hundreds of military jobs (Welsh, Kucinkas, & Curran, 1990; Welsh, Trent, Nakasone, Fairbank, Kucinkas, & Sawin, 1990; Wilbourn, Valentine, & Ree, 1984).

There are three generally accepted ways of estimating the  $g$  component of a set of variables (Jensen, 1980). For the test battery, Ree and Earles (in press) have shown that estimates of  $g$  from these three methods—principal components, principal factors, and hierarchical factor analysis—all correlated greater than .996. The mathematically simple principal components (Hotelling, 1933a, 1933b) were chosen to represent  $g$  and  $s_1 \dots s_n$  of the test. These components are orthogonal and, as Kendall, Stuart, and Ord (1983) noted, such uncorrelated scores avoid problems of colinearity and are useful for regression analyses. Table 2 gives the 10 sets of principal component score weights and the eigenvalues derived on the normative sample (Ree & Earles, in press).

TABLE 1  
*Subtests of the ASVAB*

Subtests	Number of items	Time in minutes	Reliability
General science (GS)	25	11	.80
Arithmetic reasoning (AR)	30	36	.87
Word knowledge (WK)	35	11	.88
Paragraph comprehension (PC)	15	13	.67
Numerical operations (NO)	50	3	.72
Coding speed (CS)	84	7	.77
Auto & shop information (AS)	25	11	.82
Mathematics knowledge (MK)	25	24	.84
Mechanical comprehension (MC)	25	19	.77
Electronics information (EI)	20	9	.71

*Note:* Reliabilities estimated by correlations of parallel forms. Taken from Table C-1 of Palmer, Hartke, Ree, Welsh, & Valentine, 1988.

TABLE 2  
*Principal Component Weights and Eigenvector for ASVAB Subtests*

	Principal Components				
	1	2	3	4	5
GS	.13808	-.11244	-.21982	-.29416	.19523
AR	.13715	.03854	-.39912	.54694	-.02066
WK	.13736	.06649	-.21381	-.64261	-.08976
PC	.12778	.16656	-.31273	-.71570	-.02359
NO	.11291	.38342	.42663	.23843	-1.36760
CS	.09956	.44464	.75816	.03679	1.11560
AS	.10878	-.43374	.60474	-.00918	-.34001
MK	.12965	.12086	-.61486	.64452	.20353
MC	.12448	-.30623	.21087	.39938	.36281
EI	.12857	-.29635	.14351	-.13640	-.00001
	6	7	8	9	10
GS	-.88893	-1.05107	.56764	.46367	-1.25618
AR	.26159	.58641	.25640	-1.51740	-1.06178
WK	-.20343	-.35471	.19392	-1.22910	1.53259
PC	1.10958	.48914	-.18581	.83254	-.55741
NO	-.11449	-.39672	-.29306	.20266	-.11527
CS	-.14894	.21734	.13184	-.06193	-.04099
AS	.22086	.62982	1.28388	.27471	.26269
MK	-.26607	.28551	.29615	1.16925	1.09690
MC	.89768	-1.19071	-.72807	-.02996	.28081
EI	-.78167	.90823	-1.43032	.09391	-.06884
Factor/Eigenvector					
	1/6.39	2/1.28	3/.52	4/.50	5/.29
	6/.27	7/.21	8/.20	9/.16	10/.14

These 10 unrotated principal component scores were the measures of ability under investigation. The first unrotated principal component served as a measure of  $g$ , and the other nine as the measures of specific abilities ( $s_1 \dots s_9$ ). Each subtest contributed substantially to the estimation of  $g$ , and all specific abilities measured by the test were contained in  $s_1 \dots s_9$ .

This is a mathematical definition of  $g$  (Thomson, 1939) and the specific abilities, and is necessary for analyses to proceed. That  $g$  is more than a mathematical artifact can be seen by its relationship with physiological measures, such as reaction time and evoked cortical potential, as well as with practical measures, such as academic accomplishment and job performance (Jensen, 1987). In practice, every estimate of  $g$  is bound by the composition of the group of variables from which it was estimated. But the invariance of  $g$  across groups of variables becomes greater as a function of the number, diversity, and cognitive complexity of the variables (Jensen, 1987).

*Criteria.* The criteria were final school grades (FSG) received by the subjects in one of 82 job training courses. The FSG was typically an average of four multiple-choice technical knowledge and procedures tests.

FSG was reported as a numerical grade from a lowest passing of 70 to a highest of 99, although some courses credited 60 as passing. The selectivity and difficulty of the courses were not reflected in the criterion measures. For example, the easier courses, which received the lowest aptitude students, gave grades with the same effective range of scores as the more difficult courses, which received the highest aptitude students.

*Categorical job variables.* Each job was represented by a categorical variable taking values of 1 or 0. Every subject in job A has a 1 in the categorical variable for that job and a 0 in all the other categorical job variables. Similarly, every subject in job B has a 1 in job B's categorical variable and 0s elsewhere. There were 82 categorical job variables. In some texts (e.g., Cohen & Cohen, 1983, p. 183) these are referred to as "dummy" variables.

*Job-by-ability interaction variables.* These variables were the product of multiplying the continuous variables for ability by the categorical job variables. For each of the 82 jobs there were 10 ability variables, producing a total of 820 interaction variables. For a person in job A, the values in the job A interaction variables were their  $g$  and  $s_1 \dots s_9$  scores. For the other 740 interaction variables (from the other 81 jobs the person was not in) the values were all 0s.

*Procedures*

Linear models analyses (Ward & Jennings, 1973) were used to assess the contribution of  $g$ ,  $s_1 \dots s_9$ , and job pooling in predicting training success. The testing of linear models is based on establishing a "full" model, which contains a given set of information, and then evaluating the loss of predictive accuracy resulting from the elimination of portions of that information. The simplified model, which comes from the elimination of portions of the information, is often called a "restricted" model. Removing information from the equations decreases prediction. The full and restricted models are compared with an  $F$  test to determine the significance of the loss of information. Table 3 gives the linear models and variables used.

Three pairs of linear models were tested to evaluate the relative predictive efficiency of general and specific abilities. These pairs were models 1 and 2, models 3 and 4, and models 5 and 6. The first of each pair, the full model, contained both  $g$  and  $s_1 \dots s_9$  ability information, but the second of the pair, the restricted model, contained only  $g$  information. Comparison of the models allowed the assessment of whether removing the specific abilities reduced prediction significantly.

The first of the three pairs of models contained job category variables and the interaction variables of job category-by-ability. There were 902 variables in the full model. This model is only appropriate with more subjects per job than are often available in industrial research. The second pair of models contained job category variables and ability variables, but no job category-by-ability interaction variables. The second pair of models requires many fewer subjects than the first pair. The third pair of models contained only ability information, and requires sample sizes likely to be obtainable in any industrial research.

Thorndike (1986) also speculated about the utility of having a single equation for abilities, what he referred to as "a kind of universal set of predictor weights that would be applied to all jobs" (p. 337). Analyses of two pairs of linear models addressed the predictive efficiency of universal weights. These pairs were models 1 and 3, and models 2 and 4. The full models of these pairs had as many sets of ability weights as there were jobs. The restricted models applied a single set of ability weights across all jobs. Comparison of these models allowed the assessment of whether removing the job-specific weights reduced prediction significantly.

Both the full and restricted models of the first pair used  $g$  and  $s_1 \dots s_9$ , and both the full and restricted models of the second pair used only  $g$ . These two pairs of models represented the possible outcomes from the investigation of specific ability added to  $g$  (i.e.,  $s_1 \dots s_9$  adds to  $g$ , or  $g$  alone is sufficient).

TABLE 3  
*Linear Models Used in the Statistical Tests*

Name	Description
Model 1	FSG = 82 jobs binary variables + 820 JP1 to JP10 product variables
Model 2	FSG = 82 jobs binary variables + 82 JP1 product variables
Model 3	FSG = 82 jobs binary variables + P1 to P10
Model 4	FSG = 82 jobs binary variables + P1
Model 5	FSG = P1 to P10 + intercept
Model 6	FSG = P1 + intercept
Variables used in the linear models	
FSG	Final school grade, a continuous variable
JOBS	Categorical variable, 1 if in job, 0 otherwise
P1 to P10	Scores on principal components 1 to 10, continuous variables
JP1 to JP10	Product variables of categorical variable JOBS and P1 to P10

*Note:* The intercept is only shown for models 5 and 6 because the job binary weights in the other models are effectively 82 intercepts. P1 is the estimate of  $g$ , and P2 through P10 are estimates of specific abilities.

Model 1 contained 902 independent variables, allowing each job to have its own intercept, and 10 ability slopes, one for each of the 10 principal component scores. There were 820 variables associated with the principal components (10 per job times 82 jobs) and 82 job categorical variables.

Model 2 contained 82 job categorical variables and the 82 interaction variables of job category with  $g$  (JP1). This allowed each job to have its own intercept and its own slope or relationship to the criteria for  $g$ . No other ability information was included, and there were 164 variables in this model.

Model 3 contained intercepts for each of the 82 jobs and the 10 principal component scores (P1 to P10) for a total of 92 predictors. In this model, the relationship (slope) between each ability predictor and FSG was constrained to be the same for all jobs. That is,  $g$  had a single slope across all jobs,  $s_1$  had a single slope across all jobs, as did  $s_2$ , and so on.

Model 4, with 83 predictors, had 82 job categorical variables and the scores on principal component 1 ( $g$ ), which was constrained to have the same slope across all jobs.

Model 5 contained only the 10 principal components (P1 to P10) as variables, with constant slopes across all jobs and the regression intercept. This model removed all job information.

Model 6 contained only principal component 1 ( $g$ ), with a common slope across all jobs and the regression intercept.

When information about specific abilities was removed, model 1 was reduced to model 2, model 3 to model 4, and model 5 to model 6. Comparisons of these models answered questions about the capacity of specific ability to add to the predictive efficiency of  $g$ . Further, model 1 was reduced to model 3, with the imposition of a common slope on  $g$  instead of 82 slopes, and a common slope for  $s_1$  instead of 82, and so on through  $s_9$ . Likewise, model 2 was reduced to model 4, with the restriction of one slope for  $g$  instead of 82 slopes. Imposition of common slopes creates universal prediction equations and permits evaluation of their predictive efficiency.

Each statistical test compared two models to answer questions about the predictive efficiency of general and specific ability or about pooling of jobs for universal weights. All statistical tests were evaluated at the  $p < .01$  Type I error rate.

While the linear models'  $F$  tests compared the difference between error sums of squares (or  $R^2$ s), the relative predictive efficiency of the models was evaluated using  $R$ s. Brogden (1946) presents a proof which demonstrates that predictive efficiency is linearly and directly related to  $R$  (or  $r$  in the bivariate case). A correlation of .40 is half as efficient in prediction as a correlation of .80. The predictive efficiency of restricted models was expressed as a percent of the predictive efficiency of their full models.

### Results

The predictive efficiency of each of the six linear models was determined to be significantly different from zero. Table 4 shows the correlations and the respective  $F$  statistics.

Each restricted model was tested against its respective full model, and all were found to be statistically different.

The test of model 2 against model 1, the first test of the predictive utility of specific abilities beyond  $g$  alone, yielded a statistically significant but small difference ( $F = 4.31$ ,  $df_1 = 731$ ,  $df_2 = 77139$ ). There was an increase of .08 in the  $S_e$ . The  $R$  difference of .02019 (.62831 - .60812 = .02019) represented a loss in relative predictive efficiency of 3% (.02019/.62831 = .03213). This restricted model contained all job information (the job categorical variables) and variables to allow  $g$  to have a different slope for each job. No information on specific abilities was included, and the loss in relative predictive efficiency was very small.

As in the first statistical test, the second test of the utility of  $s_1 \dots s_9$  over  $g$  found a significant difference between model 3 and model 4 ( $F = 88.88$ ,  $df_1 = 9$ ,  $df_2 = 77948$ ). There was a very small  $R$  difference of .00529, a loss of .8% relative predictive efficiency, and a commensurately



TABLE 4

*Correlations, Squared Multiple Correlations, and F Tests for the Models*

Model	<i>R</i>	<i>R</i> <sup>2</sup>	<i>df</i> <sub>1</sub>	<i>df</i> <sub>2</sub>	<i>F</i>	<i>S</i> <sub><i>e</i></sub>
M1	.62831	.39477	901	77139	55.77900	5.09
M2	.60812	.36981	163	77877	280.37153	5.17
M3	.60863	.37042	91	77949	503.98888	5.16
M4	.60334	.36402	82	77958	544.17176	5.19
M5	.42814	.18330	10	78030	1751.36070	5.88
M6	.41803	.17475	1	78039	16525.06542	5.91

Note: The *F*s in this table test the null hypothesis of no relationship. All are significant at  $p < .01$ . *S*<sub>*e*</sub> is the standard error of estimate of the regression.

small increase in *S*<sub>*e*</sub> from 5.16 to 5.19. Model 3 created parallel regression lines for each principal component across jobs, but allowed each of the 10 principal components to have its own slope. For example, the slope for principal component 1 was the same for each job, and the slope for principal component 2 was the same for each job, but the slopes of principal components 1 and 2 were not necessarily the same. This model specified that the predictive value of ability was the same for each job, differing only in level. Again the loss in predictive efficiency caused by omitting the specific ability predictors was all but meaningless.

The test of model 6 against model 5, the final test of the utility of specific ability predictors, produced a very small but statistically significant result ( $F = 4.78$ ,  $df_1 = 819$ ,  $df_2 = 77139$ ). The difference in correlations was .01, yielding a 2% loss in relative predictive efficiency. The increase in *S*<sub>*e*</sub> was a trivial .03. Using only *g* with the same slope for all jobs led to a very small reduction in predictive efficiency.

The linear models comparison of model 3 against model 1 investigated the utility of a universal set of weights on *g* and *s*<sub>1</sub>...*s*<sub>9</sub> (10 weights), as opposed to 82 job-unique sets of weights. The statistical test found the models to be significantly different ( $F = 3.83$ ,  $df_1 = 810$ ,  $df_2 = 77139$ ). The *R* difference was .01968, with a loss of relative predictive efficiency of 3% and an increase in *S*<sub>*e*</sub> of .07. Using a universal set of weights would degrade prediction a small amount.

A test of the difference between models 2 and 4 was conducted to determine the statistical difference and the relative predictive efficiency loss from using a single slope (a universal weight) for *g*, as opposed to using a different slope for each job. It was statistically significant ( $F = 8.83$ ,  $df_1 = 81$ ,  $df_2 = 77877$ ), and there was an *R* difference of .00478 and a .02 *S*<sub>*e*</sub> increase. While there was not precisely the same relationship between *g* and each job, the differences among these relationships were so trivial that the loss in relative predictive efficiency was less than 1%.

Finally, it may be noted that models 5 and 6 were less predictive than models 1 through 4, because 5 and 6 lacked information on jobs. The first four models had job binary variables, which allowed the regression for each job to have its intercept adjusted to minimize the errors of prediction. The substantial drop in  $R$ s was a result of using the same grading range for all job training courses, regardless of whether the courses were easier and received somewhat lower ability students or more difficult and received somewhat higher ability students. This situation should be expected in all studies with more than one job, and job categorical variables should be used.

### *Discussion*

The three pairs of models tested to investigate the incremental validity of specific ability, beyond that of general ability, showed statistical significance. However, the practical contribution of specific ability measures was trivial, adding an average of .01186 to predictive efficiency. The comparison of model 1 and model 2 was particularly informative because model 1 allowed  $s_1 \dots s_9$  to be best weighted for each job. This placed specific abilities at maximal advantage, removing the objection that their contributions were lost in averaging across jobs. Even in this comparison, the gain in predictive efficiency was a small .02019, an estimate of Thorndike's "modest amount" (1986, p. 338). While model 1 is likely to be unobtainable in industrial settings because of small sample sizes, the current study has enough subjects to make it appropriate (86.5 subjects per predictor in the model with the largest predictor set).

The two pairs of models tested to examine the utility of a set of universal weights showed statistical significance, but little practical importance. An average loss of only .01223 in predictive efficiency was found. The first of the two pairs comparing model 3 to model 1 was again most informative. The predictive efficiency loss was only .01968 for pooling all jobs, substantiating Thorndike's (1986) second speculation.

The test battery used had seemingly specific measures of automotive knowledge, shop information, word knowledge, reading, mathematics, mechanical principles, electronic and scientific facts, as well as clerical speed, yet its predictive power was derived from psychometric  $g$ . The training courses prepared students for seemingly different job performance, such as handling police dogs, clerical filing, jet engine repair, administering injections, and fire fighting, yet a universal set of weights across all jobs was as good as a unique set of weights for each job.

Two conclusions emerged. First, differential prediction in this differential aptitude battery is largely illusory. Psychometric  $g$  was the best predictor, and measures of specific ability were not needed to predict

training success. Second, the pooling of jobs in validity studies to compensate for small sample sizes and using common regression equations will not cause much of a reduction in prediction. However, this pooling should be accomplished with job categorical variables to allow for adjustment of criterion means. The practical implication is that when sample sizes are so small as to prohibit accurate within job criterion prediction, using  $g$  and a universal equation will still yield good prediction for training success.

It would be appropriate to extend this study to other criteria, such as job performance, absenteeism, or accident rates, to further bolster understanding of the roles of aptitudes in prediction.

## REFERENCES

- Brogden HE. (1946). On the interpretation of the correlation coefficient as a measure of predictive efficiency. *Journal of Educational Psychology*, 37, 65–76.
- Brogden HE. (1951). Increased efficiency of selection resulting from replacement of a single predictor with several differential predictors. *Educational and Psychological Measurement*, 11, 173–196.
- Cohen J, Cohen P. (1983). *Applied regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Department of Defense. (1984). *Test Manual for the Armed Services Vocational Aptitude Battery*. North Chicago, IL: United States Military Entrance Processing Command.
- DuBois PH. (1970). *A history of psychological testing*. Boston: Allyn and Bacon.
- Galton F. (1883). *Inquiries into human faculty and its development*. London: Macmillan.
- Ghiselli EE. (1973). The validity of aptitude tests in personnel selection. *PERSONNEL PSYCHOLOGY*, 26, 461–477.
- Hotelling HH. (1933a). Analysis of a complex of statistical variables with principal components. *Journal of Educational Psychology*, 24, 417–441.
- Hotelling HH. (1933b). Analysis of a complex of statistical variables with principal components (continued). *Journal of Educational Psychology*, 24, 498–520.
- Hull C. (1928) *Aptitude testing*. Great Britain: World Book.
- Hunter JE. (1986). Cognitive ability, cognitive aptitudes, job knowledge and job performance. *Journal of Vocational Behavior*, 9, 340–362.
- Jensen AR. (1980). *Bias in mental testing*. New York: The Free Press.
- Jensen AR. (1984). Test validity:  $g$  versus the specificity doctrine. *Journal of Social and Biological Structures*, 7, 93–118.
- Jensen AR. (1987). The  $g$  beyond factor analysis. In Ronning RR, Glover JA, Conoley JC, Diwitt JC (Eds.), *The influence of cognitive psychology on testing and measurement*. Hillsdale, NJ: Erlbaum.
- Kendall M, Stuart A, Ord JK. (1983). *The advanced theory of statistics Volume 3* (4th ed). New York: Macmillan.
- Maier MH, Sims WH. (1986). *The ASVAB score scales: 1980 and World War II* (CNR 116). Alexandria, VA: Center for Naval Analyses.
- McNemar Q. (1964). Lost: our intelligence? Why? *American Psychologist*, 19, 871–882.
- Palmer P, Hartke DD, Ree MJ, Welsh JR, Valentine LD Jr. (1988). *Armed Services Vocational Aptitude Battery (ASVAB): Alternate forms reliability* (Forms 8, 9, 10 and 11) (AFHRL-TP-87-48, AD-A191 658). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

- Ree MJ, Earles JA. (in press). The stability of convergent estimates of *g*. *Intelligence*.
- Ree MJ, Wegner AG. (1990). Correcting differences in answer sheets for the 1980 Armed Services Vocational Aptitude Battery reference population. *Military Psychology*, 2, 157-169.
- Schmidt FL, Hunter JE. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, 62, 529-540.
- Schmidt FL, Hunter JE. (1984). A within setting empirical test of the specificity hypothesis in personnel selection. *PERSONNEL PSYCHOLOGY*, 37, 317-326.
- Spearman C. (1904). "General Intelligence," objectively determined and measured. *American Journal of Psychology*, 15, 201-293.
- Thomson GH. (1939). *The factorial analysis of human ability*. London: University of London Press.
- Thorndike RL. (1985). The central role of general ability in prediction. *Multivariate Behavioral Research*, 20, 241-254.
- Thorndike RL. (1986). The role of general ability in prediction. *Journal of Vocation Behavior*, 29, 332-339.
- Thurstone LL. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.
- Ward JH, Jennings E. (1973). *Introduction to linear models*. Englewood Cliffs, NJ: Prentice-Hall.
- Welsh JR Jr, Kucinkas SK, Curran LT. (1990). *Armed Service Vocational Aptitude Battery (ASVAB): Integrative review of validity studies* (AFHRL-TR-90-22, AD-A225 074). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.
- Welsh JR Jr, Trent LM, Nakasone RI, Fairbank BA Jr, Kucinkas SK, Sawin LL. (1990). *Annotated bibliography of Armed Services Vocational Aptitude Battery (ASVAB) validity studies* (AFHRL-TR-89-76, AD-A220 375). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.
- Wilbourn JM, Valentine LD Jr, Ree MJ. (1984). *Relationships of the Armed Services Vocational Aptitude Battery: (ASVAB) Forms 8, 9, and 10 to Air Force technical school final grades* (AFHRL-TR-84-8, AD-A144 213). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Copyright of Personnel Psychology is the property of Blackwell Publishing Limited and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.