JOURNAL OF BUSINESS AND PSYCHOLOGY
Volume 13, No. 4, Summer 1999

SELECTION TESTS FOR FIREFIGHTERS: A COMPREHENSIVE REVIEW AND META-ANALYSIS

Gerald V. Barrett
University of Akron

Michael D. Polomsky
Barrett & Associates, Inc.

Michael A. McDaniel
University of Akron

ABSTRACT: This study reviewed and summarized the literature on the use of written tests in the selection of firefighters using a sample of 13,418 individuals drawn from 101 samples. For the prediction of job performance, cognitive tests showed substantial validity (.42), although mechanical comprehension tests showed even higher validity (.54). However, the best prediction was obtained by tests which were composites of cognitive and mechanical predictors (.56). Training criteria was best predicted by cognitive measures (.77), although mechanical comprehension predictors also showed substantial prediction value (.62). Tests which were composites of cognitive and mechanical measures showed validities equal to that of cognitive measures (.77).

This study was undertaken to estimate the validity of tests designed to measure cognitive and mechanical ability used in the selection of our nation's firefighters. Quality selection of firefighters is very important because of the critical job duties of firefighters during fire, hazard, and other life-threatening events. Documentation of the validity of common selection procedures for firefighters is also needed given the extensive amount of litigation in the area of fire personnel selection.

Campbell (1982) stated that one of the large voids in the selection literature was validity studies on police officers and firefighters. Hirsh, Northrop and Schmidt (1986) sought to fill part of this void through an

Address correspondence to Gerald V. Barrett, Department of Psychology, University of Akron, Akron, OH 44325-4301.

© 1999 Human Sciences Press, Inc.

evaluation of cognitive ability predictors for law enforcement officers. Employing validity generalization methods, the validity of memory, quantitative, reasoning, spatial/mechanical, verbal, and verbal plus reasoning tests for the given criteria of performance in training programs and training on the job was examined. For the training criteria, the hypothesis of situational specificity was rejected for four of the six test types and the 90% credibility value was at least 0.37 for all of the tests. For the job performance criteria, the situational specificity hypothesis was rejected for two of the seven test types. Further, a useful 90% credibility value was obtained for only the spatial/mechanical test type (0.17). The lower validities for job performance were attributed to criterion problems. Given that most police work is conducted in the field, the supervisors have very little opportunity to observe their subordinate's performance and thus have problems providing accurate ratings of performance.

The present study examines the validity of tests for firefighters. Police and firefighters both go through an extensive training period. Similar to the findings of Hirsh et al. (1986) for police officers, we anticipate the selection tests for firefighters to be useful predictors of training performance. Unlike the criterion problems experienced with police occupations, the supervisors of firefighters have substantial opportunity to observe the performance of their subordinates. Thus, in contrast to the findings of Hirsh et al. (1986) for police officers, we anticipate that the selection tests will be useful predictors of the performance of firefighters.

METHOD

Meta-analysis as a method of determining validity. The Schmidt-Hunter meta-analysis method used in this study is based on the hypothesis that much of the variation in results across studies may be due to statistical and methodological artifacts rather than to substantive differences in underlying population relationships. Some of these artifacts also reduce the correlations below their true (e.g., population) values. The method determines the variance attributable to sampling error and to differences between studies in reliability and range restriction, and subtracts that amount from the total amount of variation, yielding estimates of the true variation across studies and of the true average correlation (Hunter & Schmidt, 1990). Artifact distribution meta-analysis, using the interactive method, was employed (Hunter & Schmidt, 1990, Chapter 4). The mean observed correlation ($\bar{r}$) was used in the sampling error variance formula (Hunter & Schmidt, 1990, pp. 208–210; Law, Schmidt, & Hunter 1991; Schmidt, Law, Hunter, Rothstein, Pearlman, & McDaniel, 1993). The computer program utilized is described in McDaniel (1986). Additional detail on the program is presented in Appendix B of McDaniel, Schmidt, and Hunter (1988).

Analysis Method

Literature search. The data used in this study was compiled from criterion-related studies of cognitive ability and mechanical/spatial tests conducted over the past two decades. A conscious effort to locate as many validity studies as possible was undertaken in order to enhance the confidence one can place in these results. The analysis undertaken utilized both published selection studies of firefighter examinations as well as unpublished sources. The following references were used to locate firefighter validity studies: *Psychological Abstracts, Psychological Documents, Selected Rand Abstracts, National Technical Information Service, Government Documents, Mental Measurements Yearbook, Tests in Print,* and *Tests.* Information was also requested from test publishers who produced commercially available firefighter tests. Thus, an attempt was aggressively made to locate all published and unpublished firefighter validity studies. The search also utilized automated computer searches.

For job performance criteria, data from 73 independent samples comprising 9,515 individuals were obtained. For training performance, validity data from 28 independent samples comprising 3,903 individuals were located.

Predictors. A detailed coding scheme was established to categorize the predictors. To permit an adequate number of coefficients for each predictor category, tests were assigned to one of three categories: cognitive tests, mechanical comprehension tests, and composites of cognitive and mechanical measures. Tests which are mixtures of mechanical and spatial abilities were classified as mechanical.

Criterion. The criteria were categorized as training performance or job performance. Firefighters receive substantial training prior to being placed in operational service. Such training is a mixture of classroom learning and hands-on learning. The grades in such training served as training performance criterion. Job performance criterion consisted of supervisory ratings.

Range restriction and criterion reliability data. Range restriction, predictor and criterion reliability were obtained from the studies which provided the validity data.[1] The resulting distributions were comparable to those reported by Pearlman (1979) thus adding confidence to the appropriateness of the artifact distributions.

Decision rules. Validity coefficients were accepted for the position of entry-level firefighter. This study excluded data on paramedic personnel.

RESULTS

Table 1 present the results for job performance. Table 2 present the results for training performance. The first column of each table identi-

[1]The artifact distributions are available from the senior author.

Table 1
Meta-analysis Results for the Validity of Selection Tests for Firefighters. Criterion Is Job Performance

Distribution	Number of r	Total N	Mean Observed r	Observed σ	p	σ_p	90% CV
Cognitive tests	24	2,791	.20	.19	.42	.35	−.03
Mechanical tests	26	3,087	.26	.17	.54	.29	.17
Cognitive/mechanical tests	23	3,637	.28	.11	.56	.12	.40

Table 2
Meta-analysis Results for the Validity of Selection Tests for Firefighters. Criterion Is Training Performance

Distribution	Number of r	Total N	Mean Observed r	Observed σ	p	σ_p	90% CV
Cognitive tests	14	2,007	.50	.09	.77	.03	.73
Mechanical tests	5	869	.37	.14	.62	.17	.40
Cognitive/mechanical tests	9	1,027	.50	.12	.77	.12	.62

fies the distribution of validities analyzed. The next four columns present the total sample size, the number of validity coefficients on which each distribution was based, and the uncorrected mean and standard deviation of each distribution. The final three columns present the estimated population mean (p), the estimated population standard deviation (σ_p), and the 90% credibility value for the distribution of true validities. These population distribution estimates are for distributions in which the mean true validities are corrected for unreliability in the criterion and range restriction. Corrections to the mean do not include corrections for predictor unreliability because the tests, as used operationally, have less than perfect reliability. The variances of the true validity distributions are corrected for sampling error and for differences among the studies in predictor and criterion reliability and range restriction.

DISCUSSION

The results for the prediction of job performance demonstrate the substantial evidence for the validity of commonly used predictors of firefighter job performance. The firefighter position is unusual in that mechanical predictors have a larger mean validity (.54) than do the cognitive predictors

(.42). Typically, cognitive predictors have greater validity than any other predictor. However, the superiority of the mechanical tests is consistent with the substantial mechanical demands inherent in the successful performance of firefighter duties. Although both cognitive and mechanical predictors have substantial levels of prediction, their combination into a cognitive/mechanical composite yields the highest validity (.56).

The validities of the examined tests for the prediction of training performance are also large. Cognitive predictors yield the highest validities (.77). Mechanical predictors also show very large validities (.62). Tests which are composites of cognitive and mechanical predictors yield the same level of validity as the cognitive measures alone. The apparent lack of incremental validity for the mechanical tests over and above the validity of the cognitive tests is suspected to be due to the substantial validity of the cognitive tests. It is very difficult to improve prediction over and above that of the cognitive tests alone at .77.

One can place substantial confidence in the validity evidence for the prediction of job performance criteria. All three distributions shown in Table 1 are based on a relatively large number of coefficients (23 to 26) and the sample sizes within each distribution are also large (2,791 to 3,637). The conclusion drawn from the training criteria are more tentative. The number of coefficients in each distribution are not large (5 to 14) and the total sample sizes are smaller than some might like (869 to 2,007). Still, these results are based on data that is much more impressive than that obtained in any single validity study and represent the best estimates of validity based on data collected to date.

Collectively, the data for the cognitive and mechanical predictors suggest that these tests show substantial validity. Thus, although the continuing litigation concerns in the selection of firefighters may make it wise for organizations to conduct validity studies for each test administered, this study's results provide strong evidence that such validation efforts are largely unneeded on scientific grounds.

REFERENCES

Campbell, J.C. (1982). Editorial. Some remarks from the outgoing editor. *Journal of Applied Psychology, 67*, 671–700.

Hirsh, H.R., Northrop, L.C., & Schmidt, F.L. (1986). Validity generalization results for law enforcement occupations. *Personnel Psychology, 39*, 399–420.

Hunter, J. E. & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Beverly Hills, California: Sage Publications.

Law, K. S., Schmidt, F. L., & Hunter, J. E. (1991). *Non-linearity of range correlations in meta-analysis: A test of an improved procedure*. Submitted for publication.

McDaniel, M. A. (1986). Computer programs for calculating meta-analysis statistics. *Educational and Psychological Measurement, 46*, 175–177.

McDaniel, M. A., Schmidt, F. L., & Hunter, J. E. (1988). A meta-analysis of methods for rating training and experience in personnel selection. *Personnel Psychology, 41*, 283–314.

Pearlman, K. (1979). *The validity of tests used to select clerical personnel: A comprehensive summary and evaluation (TS-79-1).* U.S. Office of Personnel Management, Personnel Research and Development Center, (NTIS No. PB 80. 102 650).

Schmidt, F. L., Law, K., Hunter, J. E., Rothstein, J. R., Pearlman, K., & McDaniel, M. A. (1993). Refinements in validity generalization procedures: Implications for the situational specificity hypothesis. *Journal of Applied Psychology, 78,* 3–13.

APPENDIX A: PUBLISHED FIREFIGHTER STUDIES

Alvares, K.M., and Boston, D.A. (1976). *THe validation of a selection device for metropolitan firefighters: A comparison of measures of aptitude, personality measures, and measures designed to be job-related.* (Technical Report No. TCSC 76-1). Toledo, OH: Toledo Civil Service Commission.

Barrett, G.V., Alexander, R.A., Byers, P., and Klein, I. (1981). *Development of an entrance level selection program for firefighters.* Akron, Ohio.

Bullock, C. (1981). *A firefighter test transportability study: Comparison of Robbins Hose Company (Dover, Delaware) and Maryland counties firefighter jobs.* Philadelphia, PA: U.S. Office of Personnel Management, Mid-Atlantic Region.

Civil Service Board, City of Tampa (1977). *A validation report: Entry-level firefighters.* Tampa, FL: Civil Service Board, City of Tampa.

Ford, S.F., and Sullivan, S.M. (1982). *Firefighter validation study: A concurrent study of the entry-level position.* NJ: New Jersey Department of Civil Service, Division of Examinations.

Henderson, N.D. (1986). *A concurrent validation study of the validity and utility of the 1983 Cleveland firefighter Entry Level examination.* Oberlin, OH, Oberlin College, Department of Psychology.

Kriska, S.D., and Hines, C.V. (1984). *Firefighter selection-test validation study for the City of Columbus.* Columbus, OH: Municipal Civil Service Commission.

Manpower Administration (1958). Technical report on standardization of the General Aptitude Test Battery. Manpower Administration (DOL). Washington, D.C. U.S. Training and Employment Service. June 58.8P.

McCann, F.E., Zupkis, R., Howeth, W.F. and Nichols, G.V. (July, 1975). *The validation of McCann Associates ESV Entrance firefighter written test.* Huntington Valley, PA: McCann Associates.

MacNaughton, J.F., Richardson, P. and Mellon, S.J., Jr. (1978). *Job and worker characteristics of the entry-level firefighter position in the Houston Fire Department.* Houston, TX: City of Houston, Legal Department.

Miller, J.M. (1976). *Validation of a test for selecting firefighters.* Unpublished doctoral dissertation, Marquette University.

Murdy, L.B., and Norton, R.P. (1972). *Fire private test validation study: City of Fort Worth Fire Department* (IBR Technical Report No. 72-2). Forth Worth, TX: Texas Christian University, Institute of Behavioral Research.

Payne, S.S. (1979). *Validity of Test 21 for selection of entry level firefighters in the D.C. Fire Department* (Personnel Research and Development Center).

Psychological Services, Inc. (1983). *A description of the PSI entry-level Firefighter Selection Test.* Washington, D.C.: Psychological Services, Inc.

Rosenfeld, M., and Thornton, R.F. (1976). *The development and validation of a firefighter selection examination for the City of Philadelphia.* Princeton, NJ: Educational Testing Service, Center for Occupational and Professional Assessment.

Ruedebusch, V.M., Vaaler, T.P., Keck, M., Powers, M., Mendenhall, M., Bayreder, C., and Black, P. (1975). *A study of the validity of proposed written examinations for the selection of entry-level firefighters in Des Moines.* St. Louis, MO: U.S. Civil Service Commission, Intergovernmental Personnel Programs, St. Louis Region.

Shores, J.A. (1984). *Validation of a civil service selection procedures for firefighters.* Unpublished master's thesis, California State University, Long Beach, 1984. *Masters Abstracts International, 23 (01), 229.*

Skinas, N., and Goldstein, L.S. (1975). *Construction and validation of an entry-level firefighter examination.* Trenton, NJ: New Jersey Department of Civil Service, Division of Examinations, Test Validation and Staff Development Unit.

Strand, T. (1975). *A comparative study of two entry level firefighter selection tests.* Unpublished doctoral dissertation, Wayne State University.

Sulzer, J.L. (1977). *Final report. Test development project for entrance exam for firefighter.* New Orleans, LA: City of New Orleans, Department of city civil service.

Tice, T. E. (1970). *Selection systems and performance appraisal in the fire service: A study of criteria development and test validation.* Unpublished master's thesis, Iowa State University of Science and Technology, Ames, Iowa.

Tyler (1980) study (cited in Bullock, 1981).

van Rijn, P., and Payne, S.S. (1980). *Criterion-related validity research base for the D.C. Firefighter Selection Test* (Personnel Research Report 80-28). Washington, D.C.: Office of Personnel Management, Personnel Research and Development Center, Examination Services Branch. (NTIS No. PB-81-122 087).

Waibel, J.J., Billingsley, W., and Thorsen, S. (1974). *The validation of entry-level firefighter examinations in the states of California and Nevada.* Sacramento, CA: Selection Consulting Center. (NTIS No, P-231-997).

Wetrogan, L. I., and Schemmer, F. M. (1986). *Development and validation of the B-3 and B-4 entry-level firefighter examinations.* (Tech. Rep.) Alexandria, Virginia: International Personnel Management Association.

APPENDIX B

List of Contributing Studies, the Number of Samples per Study, and Number of r's per Sample

Author(s) (year)	*N* of samples	*N* of *rs*
Alvares & Boston (1976)	1	1
Barrett et al. (1981)	2	7
Bullock (1980)	1	5
Civil Service Board, City of Tampa (1977)	1	2
Ford & Sullivan (1982)	1	7
Henderson (1986)	1	3
Kriska & Hines (1984)	1	19
Manpower Administration (1958)	1	7
McCann et al. (1975)	1	1
MacNaughton et al. (1978)	1	3
Miller (1976)	1	4
Murdy & Norton (1972)	1	9
Payne (1979)	1	1
Psychological Services, Inc. (1983)	1	9
Rosenfcld & Thornton (1976)	1	1
Ruedebusch et al. (1975)	1	6
Shores (1984)	1	1
Skinas & Goldstein (1975)	1	3
Sulzer (1977)	1	3
Tice (1970)	1	1
Tyler (1980)	1	4
van Rijn & Payne (1980)	1	3
Waibel et al. (1974)	1	3
Wetrogan (1986)	1	2