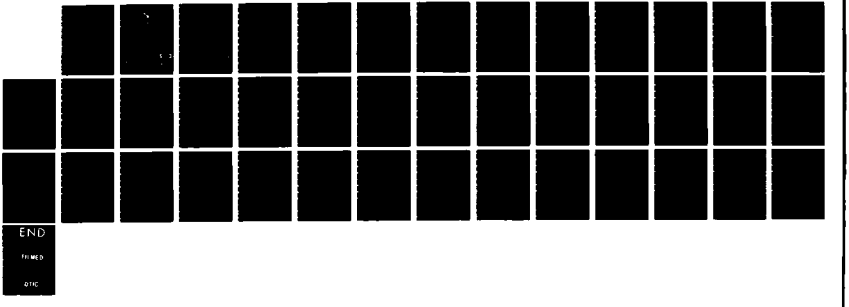
AD-A160 870 CRITERIA FOR SELECTION OF FIRST-LINE SUPERVISORS(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOOL OF SYSTEMS AND LOGISTICS F A BURKE SEP 85 1/1

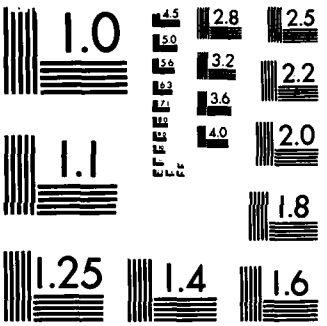
UNCLASSIFIED AFIT/GLM/LSM/85S-9 F/G 5/9 NL



END

FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(2)

AD-A160 870

CRITERIA FOR SELECTION OF FIRST-LINE SUPERVISORS

THESIS

Frances A. Burke

AFIT/GLM/LSM/85S-9

DTIC FILE COPY

DISTRIBUTION STATEMENT A
Approved for public release
Distribution Unlimited

DTIC
ELECTE
NOV 5 1985
S D
B

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY
AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

85 11 05 025

AFIT/GLM/LSM/85S-9

CRITERIA FOR SELECTION OF FIRST-LINE SUPERVISORS

THESIS

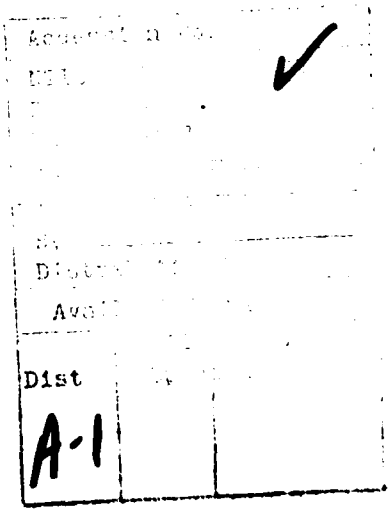
Frances A. Burke

AFIT/GLM/LSM/85S-9

DTIC
ELECTE
NOV 5 1985
S D
B

Approved for public release; distribution unlimited

The contents of the document are technically accurate, and no sensitive items, detrimental ideas, or deleterious information are contained therein. Furthermore, the views expressed in the document are those of the author(s) and do not necessarily reflect the views of the School of Systems and Logistics, the Air University, the United States Air Force, or the Department of Defense.



AFIT/GLM/LSM/85S-9

CRITERIA FOR SELECTION OF FIRST-LINE SUPERVISORS

THESIS

Presented to the Faculty

of the School of Systems and Logistics

of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science in Logistics Management

Frances A. Burke, B.A.

GS-12, DESC

September 1985

Approved for public release; distribution unlimited

Preface

The purpose of this study was to identify criteria with the greatest predictive validity for the successful performance of first-line supervisors, which could be used by middle managers to improve the percentage of successful first-time supervisors selected. A validity generalization model developed by Schmidt and Hunter was used to process data from the findings of many studies, correcting for range variation, sample size and measurement errors to ensure conservative estimates of predictor validity coefficients.

I wish to express my thanks to my thesis advisor, Dr. John A. Muller, for his guidance and continuing encouragement. I also wish to thank Dr. Guy S. Shane, my technical advisor, not only for the data base, but for his unwavering support in the face of some unexpected disappointments. Another word of thanks is due Mr. Joel Rice, a true computer expert, who made the impossible happen.

Finally, I want to express my gratitude and appreciation to my family and close friends for their patience and confidence that one day I would actually finish this thesis.

Frances A. Burke

Table of Contents

		Page
Preface		ii
List of Tables		iv
Abstract		v
I.	Introduction	1
	General Issue	1
	Specific Problem	2
	Background	3
	Scope of Research	8
	Research Objectives	9
II.	Methodology	10
	Method of Approach	10
	Methodology Literature Review	11
	Justification of the Approach	13
	Decision Rules	16
	Summary	18
III.	Findings	19
	Explanation of Output Data	19
	Analysis	22
IV.	Conclusion	25
Bibliography		27
Vita		30

List of Tables

Table		Page
I.	Comparison of Meaningful Statistics for Predictors	24

Abstract

This study identified criteria with predictive validity for the successful performance of first-line supervisors. Through meta-analysis it is possible to generalize the validity of predictors across cumulative studies. A validity generalization model, which corrects for artifactual variance, was used to process data from the findings of many previous studies based upon the job performance measurement criterion of first-line supervisors.

Analysis revealed two predictors, How Supervise? and the General Mental Ability Measures, with substantial validity for predicting successful performance of first-line supervisors. The Bennett Mechanical Comprehension Test and the Otis Mental Ability Test also showed relatively high predictive validity; however, neither form of the Leadership Opinion Questionnaire evidenced predictive validity for successful job performance by first-line supervisors. The Wonderlic Personnel Test, although not highly predictive, may be useful in the absence of the other predictors.

CRITERIA FOR SELECTION OF FIRST-LINE SUPERVISORS

I. Introduction

General Issue

"Supervision is the function of leading, coordinating, and directing the work of others to accomplish designated objectives" (3:532). "[It] is a matter of getting the job done through others" (5:xiii). "Supervisor," a generic term, "can be properly used to apply to persons in charge of production workers as well as those in charge of office workers" (3:533).

In a typical organizational structure, three divisions of management are commonly recognized--top, middle, and lower (3:533). "The first level of supervision is at the bottom of the management level" (5:3) and is one of the most critical elements in organizational effectiveness (3:532). Middle-level managers have the responsibility of selecting first-level supervisors who, generally, have no previous supervisory experience (5:4-5).

According to Bittel,

> The performance of first-line supervisors . . . is judged on two prime accounts--(1) how well the supervisor has managed the inputs or resources and (2) how

> good the results are in terms of the volume, quality, and cost of the products and services produced. (4:25)

Too often, the first-level supervisor's performance is less than satisfactory in directing his/her group of employees toward accomplishment of organizational goals.

> In one research investigation of the attitudes and motivation of 200 accountants and engineers employed in nine companies, supervision was found to be one of the chief causes for dissatisfaction. The dissatisfaction related to such aspects of supervision as incompetency, poor scheduling of work, lack of teaching ability, unfriendly relations, lack of support, and unwillingness to listen to suggestions. (3:532)

Furthermore, when management "selects a person who becomes a poor supervisor . . . it is a costly error" (5:42-43). The supervisory selection decision affects many workers, and a poor choice is both difficult and time-consuming to rectify (5:14).

Specific Problem

Much has been written identifying and discussing the characteristics of a successful supervisor. Northrup and others in The Objective Selection of Supervisors, quote Milton M. Mandell, stating

> A 'good' supervisor is one who 'generates job satisfaction, reduces absenteeism and turnover, improves the quality of work by training his employees, and serves as a source of stability during emergencies'. (15:3)

Broadwell offers a list of characteristics "common to all who have shown a high degree of success at supervision" (5:161).

1. WILLINGNESS TO WORK
2. WILLINGNESS TO TAKE RISKS
3. ENTHUSIASM
4. EMPATHETIC
5. ABILITY TO MOTIVATE
6. ABILITY TO COMMUNICATE (5:161-162)

Weger, in Motivating Supervisors, cites Edgar Schein's "complex model of man" which suggests:

1. The supervisor must be a good analyst in terms of how he manages the people under him, inasmuch as their motives are so complex and so variable.
2. The supervisor must have fine sensitivity to people to be able to appreciate the subtle differences in what motivates people, particularly the individuals in his work group.
3. The supervisor must have personal flexibility and the ability to meet the challenging needs and motives of his subordinates. (26:59)

The problem, therefore, is not in describing the desired characteristics of successful supervisors, nor even in the evaluation of supervisors' performance to determine their degree of success in organizational effectiveness. The problem is to develop criteria which will identify those employees possessing a high degree of potential for successful performance as first-line supervisors. The use, then, of such criteria in personnel selection procedures would increase the percentage of successful first-time, first-line supervisors.

Background

Historically, personnel placement was less than optimal.

> The job of supervising work and people has always been thought of in the past as a rather simple, direct

> operation. A workman elevated to the position of gang or office boss had only to exercise the authority vested in his new job to get people to work . . . . The blame for failure to follow orders or poor workmanship fell on the worker, not on the foreman. (13:3)

> When the pressure of war production . . . absorbed the labor surplus . . . [and] the supply of skilled mechanics, office personnel, and general workers [was] suddenly cut off . . . foremen and supervisors were told to make every effort to keep workers on the job, to improve the skills of these people, and to change their methods of dealing with subordinates. Most supervisors were incapable of making the required changes in methods, and as a result a chaotic period of readjustment followed. (13:4)

"The old-time, two-fisted, hard-boiled boss . . . [was] replaced by the modern leader" (13:7). Rapidly expanded educational opportunities resulted in "professionalization of the supervisor's job" (13:8). However, "as the job became more professionalized, the problem of selecting suitable candidates to fill these jobs became more difficult" (13:8-9). Thus, aptitude and achievement tests used in personnel selection processes developed extensively between 1920 and 1940 (24:1).

"Personnel testing has one specific objective: to contribute to the increasingly effective use of manpower within an organization" (10:3). During the past thirty years an abundance of material has been published detailing psychological tests and assessment methods with predictive validity in identifying personnel with high supervisory or managerial potential. So many tests are available that several comprehensive

references have been compiled providing descriptive summaries of assessment tests in psychology, education, and business (23).

Perhaps the oldest and most widely recognized measurement of supervisory potential is How Supervise?, developed in the 1940s by Q. W. File and H. H. Remmers. How Supervise? was "designed to measure 'knowledge and insight concerning human relations in industry'" (10:470). The targeted population for How Supervise? is supervisory personnel: two forms are intended for first-line supervisors, and a third form is designed for higher level supervisors (10:470). Subsequent studies, however, indicate a lack of data to support the predictive validity of How Supervise? scores in successful supervisory performance (7:406). Results of these studies suggest How Supervise? may, in fact, "measure 'supervisory knowledge'" or "verbal intelligence" (14:405).

Another early-developed measurement instrument, The Leadership Opinion Questionnaire, evidences an expected correlation between "Consideration" scores and future supervisory success in companies in which progressive personnel practices are supported. Actual predictive validity was biased by the extent to which supervisors conformed (or failed to conform) to the corporate attitude (2:345).

Spitzer and McNamara conducted a study in the early 1960s "to determine if tests could be selected which would be

positively related to managerial success" and, therefore, have predictive value in "selecting employees who have good managerial potential" (22:19). Findings showed the use of the Otis Test of Mental Ability, and the Background and *Contemporary Data Form* (a behavioral personality index)

> in the selection procedure would increase the number of managers capable of performing in an above-average manner . . . [by] as much as 36 per cent over the present situation as measured by the criterion of the study. (22:24,19)

More recent studies by Rowland and Scott reveal significantly positive relationships between work group performance (perceived to be a factor of successful supervisory performance) and the supervisor's intelligence, need for aggression, and influence with superiors (measured by *Edwards Personal Preference Schedule*) (17:375).

The current trend is toward the use of assessment centers to determine supervisory/managerial potential and predict candidates' successful performance (25:595).

> The combination of management's desire to select managers of high quality and the federal government's policy of demanding objective proof of nondiscriminatory selection has generated substantial interest in the assessment center method.
>
> The assessment center method is holistic; it engages almost totally the personalities of candidates for upgrading and provides an analytical and evaluative structure whereby a number of assessors can observe the candidates and can combine several predictors to form an overall rating of potential for each candidate. (15:91)

Considerable disparity exists, however, regarding the predictive validity of assessment center criteria. Dunnette and Borman, in the Annual Review of Psychology, 1979, found "reports have been favorable to the assessment center method" (8:510). Five years later, however, Turnage and Muchinsky identified a number of problems associated with the assessment center evaluations. Two fundamental problems are: (1) absence of reliable and objective supervisory performance measures, and (2) lack of comparative data for "alternative predictors" (25:595-596). In summary, Turnage and Muchinsky found

> that (a) both assessment center evaluations and traditional predictors were generally unrelated to job performance but that (b) assessment center evaluations were predictive of promotability. (25:595)

The California Psychological Inventory (CPI) is designed to measure

> poise, ascendancy, . . . self-assurance, . . . socialization, maturity, . . . responsibility . . . achievement potential and intellectual efficiency . . . and intellectual and interest modes. (10:316)

Findings reveal that use of only six of the 18 scales (dominance, good impression, communality, flexibility, femininity, and intellectual efficiency) give "a less ambiguous interpretation of the total profile" (10:315-317). Harrison G. Gough, Institute of Personality Assessment and Research, University of California, Berkeley, conducted a new analysis of the

findings of the CPI "good manager" scale developed in 1963 by Goodstein and Schrader. Both the original (1963) and revised (1983) versions of the CPI signify predictive validity for managerial success. Gough's conclusion recommended additional new research addressing "explicit measures of managerial performance for both sexes, . . . interactional issues, and . . . longitudinal criteria" (9:233).

Leaetta M. Hough, Personnel Decision Research Institute, Minneapolis, Minnesota, recently developed the "Accomplishment Record" (AR) method of identifying professionals for selection and promotion. The AR Inventory is based upon the hypothesis that "'the best indicator of future performance is past performance'" (11:135). Initial results suggest the AR Inventory "is equally predictive and fair" for men, women, minorities, and nonminorities. Further, "the AR Inventory appears to measure aspects of the individual not tapped by more traditional methods of testing," and connote significant predictive validity in job performance measures (11:142,143,146).

Scope of Research

The literature includes studies relative to the selection of supervisory and/or managerial personnel. Northrup and others (1978) in The Objective Selection of Supervisors, found

> Although few jobs have received as much attention as has the job of first-line supervisor, a search discloses that, at least until recently, the literature on supervisor selection has been relatively sparse . . . . Most writings . . . deal with three areas: the role of supervisors, their qualifications, or their training. Some of the literature . . . concerns itself with supervisor selection, but even that deals largely with different techniques . . . for improving the selection process. (15:11)

An aggregation of data resulting from findings of previous studies is available for input into a validity generalization model. The data are the correlation coefficients derived from the scoring of selected predictors administered to first-line supervisors rated by various degrees of successfulness. Validity generalization models have proven effective in identifying criteria with predictive validity for selecting first-line supervisors with high potential for successful performance.

Research Objectives

What predictors can be administered to first-time supervisory candidates which will provide a significant correlation with successful job performance? What criteria exist for predicting the probability of successful performance by first-time, first-line supervisors? How can such data be used in developing criteria for the personnel selection process to assist managers in selecting first-time, first-line supervisors who will be successful in their positions?

II. Methodology

Method of Approach

The method of approach to be used for this study is meta-analysis, utilizing the Schmidt-Hunter Validity Generalization Model. "Validity generalization is the degree to which inferences from scores on tests can be transported across different situations" (6:94). Meta-analysis is the term used to describe "the statistical analysis of the findings of many individual analyses" (21:352). It is a method of integrating the findings of research literature from numerous studies in order to make sense of them (21:352-355).

An existing data base developed by Shane (20) is used to provide the statistical inputs for the modified Schmidt-Hunter model. The existing data included findings from studies published through May 1977. A literature search failed to reveal any more recent statistical data which cited "job performance" as the measurement criterion. Therefore, the previously compiled data constitute the whole of the statistical inputs used in this study.

Statistical data were gleaned from studies which reported the correlation between selected test scores and successful

performance as first-line supervisors using job performance ratings as the measurement criterion. The selected tests include: a general mental ability measure, the Wonderlic Personnel Test, Otis Mental Ability Test, Bennett Mechanical Comprehension Test, How Supervise?, and two forms of the Leadership Opinion Questionnaire. The sample size and correlation coefficients from each study are input into the Schmidt-Hunter Validity Generalization Model, and an analysis of the output data identifies those tests having predictive validity for the successful performance of first-line supervisors. These findings may then be useful in developing selection criteria which will increase the probability of identifying and selecting candidates for promotion who will perform well as first-time, first-line supervisors.

Methodology Literature Review

The term "meta-analysis" was coined by Gene V. Glass in 1976 to describe the process of integrating research findings from many studies to ensure more meaningful and comprehensible results (21:352-354). Hunter, Schmidt, and Jackson reviewed "all the methods that have been proposed for cumulating knowledge across studies," referring "to the averaging methods as 'meta-analysis'" (12:11). According to Schmidt and others,

> Cumulation of results can be used whenever there are at least two studies with data bearing on the same relation. (12:28)

> Ideally, cumulation of results works best if it is based on a large number of studies acquired by exhaustive search procedures. However, cumulation is also valid for 'convenience' samples of studies that just happen to lie at hand. This is particularly true if the corrected standard deviation suggests that all the variation across studies is due to sampling error . . . . then the accuracy of the mean value in relation to the one true population value is determined by the total number of subjects across studies. Even a relatively small number of studies may have a large cumulative sample size. (12:29)

In meta-analysis correlations are corrected through mathematical formulas for sampling error, measurement error, and range variation across the studies. The correction formulas are resident in the Schmidt-Hunter Validity Generalization Model.

> Currently, six different procedures have been defined and tested for estimating the mean and variance of true validity coefficients. These procedures or equations are the noninteractive procedure (Pearlman et al., 1980), the interactive procedure (Schmidt, Gast-Rosenberg and Hunter, 1980), the independent and dependent multiplicative equations (Callender and Osborn, 1980), and the Taylor-Series Approximations 1 and 2 (Raju and Burke, 1983). The noninteractive procedure is an improved version of the Schmidt-Hunter (1977) procedure. (6:97)

Although several different formulas have been developed, "studies have found only trivial differences that would be expected from the minute size of the product terms involved" (12:92).

The derivation of different formulas has, however, resulted in the promulgation of a number of critiques of the Schmidt-Hunter model. Lauding Schmidt and Hunter as

> pioneers in introducing the basic notions of Bayesian statistics into the field of validity research, so that results of preceding studies can be used to develop cumulative knowledge, (1:208)

Algera and others, 1984, nonetheless

> criticized . . . the compilation of validity data, the use of criterion measures, and the test of the hypothesis of no situational specificity, [concluding] the Schmidt-Hunter approach to validity generalization shows fundamental shortcomings. (1:197)

Results of recently conducted computer simulation studies to assess the accuracy of the different validity generalization procedures

> tend to support the accuracy of the various procedures for estimating the mean and variance of true validity coefficients . . . . however, [they] do suggest that there is some room for improvement in procedures to determine validity generalization based on the correlation model. (6:106)

Justification of the Approach

Prior to 1976 personnel psychologists generally believed "that meaningful empirical validation studies are possible for most, if not all, jobs in most organizations" (18:529). Schmidt, Hunter, and Urry, as a result of a study in 1976, concluded "that empirical validity studies are 'technically feasible' much less frequently than the profession [had] assumed" (18:529).

A subsequent study by Schmidt and Hunter addressing "the belief that test validity is generally highly situation specific" (18:529) resulted in the development of a validity generalization model which, in many instances, permits "validity generalization to new settings without carrying out a validation study of any kind," or, in any case, "provides an improved method of data analysis and decision making for the necessary situational validity study" (18:529). Schmidt and Hunter found

> evidence suggesting that much of the variance in the outcomes of validity studies within job-test combinations may be due to statistical artifacts. (18:529)
>
> In a well-executed large sample series of studies [conducted by Brogden], it was found that when Army occupations were classified rationally into job families, tests showed essentially identical validities and regression weights for all jobs within a given family. (18:530)

Thus,

> Brogden . . . concluded that when methodological artifacts are controlled and large samples are used . . . obtained validities are in fact quite stable and similar across time and situations for similar jobs. (18:530)

The Schmidt-Hunter model, based upon the principles of Bayesian statistics

> directly relates methods of data analysis used in making inferences about validity in criterion-related validity studies to the concept of validity generalization. The generalizability of validity is seen to be a matter of degree and is quantified in the properties of the prior distribution . . . [providing] a direct answer to the question of whether validity

generalization is justified or not without a situation-specific empirical validation study. (18:531)

Some important features of the Schmidt-Hunter model which enhance its "credibility and acceptance" are:

> First, the Bayesian priors to be used . . . are empirically determined based on data from past studies.
>
> Second, the assumptions made about between-study variance in criterion reliability and range restriction are conservative.
>
> Third, certain sources of error variance in the obtained distribution [computational and typographical errors] are not corrected for, further ensuring conservatism.
>
> Fourth, corrections made to the mean of the prior for average range restriction effects . . . probably [tend] to underestimate the true mean of the corrected prior.
>
> Fifth, . . . this procedure provides a parsimonious, sophisticated, and technically sound solution for the overarching problem of validity generalization . . . [which] may lead to large dollar savings by eliminating the need for many criterion-related validity studies.
>
> Sixth, the model can be extended to provide an improved method of data analysis and decision making in criterion-related validity studies.
>
> Finally, the model . . . provides a tool that may lead to the establishment of general principles about trait-performance relationships in the world of work . . . . (18:538)

In a more recent article by Schmidt and Hunter (1980), they demonstrate that

> validity estimation and generalization by both empirical and rational methods will be widely possible without situation-specific empirical studies.

> Estimates of the dollar impact of selection instruments on productivity will be much more frequently computed, and many who are currently skeptical will become convinced of the critical importance of selection to organizational success. (19:42)

Decision Rules

According to Schmidt and Hunter, "a valid test will show a statistically significant validity in only about 50% of studies" (19:44). And, as previously stated, evidence suggests "that much of the variance in the outcomes of validity studies . . . may be due to statistical artifacts" (19:44). Further,

> if the variance in validity coefficients across situations for job-test combinations is due to statistical artifacts, then . . . the doctrine of situational specificity is false and validities are generalizable. (19:44)

Schmidt, Hunter and others "developed a method for testing this hypothesis" (19:44).

> [Starting] with a fairly large number of validity coefficients for a given test-job combination, . . . [compute] the variance of [the] distribution. From this variance, . . . [subtract] variance due to various sources of error. (19:44-45)

The seven commonly acknowledged sources of error variance include:

1. Differences between studies in criterion reliability.
2. Differences between studies in test reliability.
3. Differences between studies in range restriction.
4. Sampling error (i.e., variance due to $N < \infty$).
5. Differences between studies in amount and kind of criterion contamination and deficiency (Brogden and Taylor, 1950).

6. Computational and typographical errors (Wolins, 1962).
7. Slight differences in factor structure between tests of a given type (e.g., arithmetic reasoning tests). (19:45)

Separate studies have shown that artifactual variance may account for as much as 100% of the observed variance for certain distributions. "Thus there is now strong evidence that the observed variation in validities from study to study for similar test-job combinations is artifactual in nature" (19:45-46).

Therefore, after making the appropriate corrections to the mean and standard deviation of a validity distribution

> one may find that a large percentage, say 90%, of all values in the distribution lie above the minimum useful level of validity. In such a case, one can conclude with 90% confidence that the estimate of true validity would be at or above this minimum level . . . . (19:46)

Note, however, that Schmidt and Hunter continue by stating that

> the best estimate of test validity is the mean of the corrected validity distribution, not the value at the foot of the 90 percent confidence interval. (19:46)

Further, due to

> a number of [recent] significant developments in personnel psychology resulting . . . from rejection of the erroneous belief in the law of small numbers . . . it now appears likely that rational estimates [based on dollar utility gain formulas] of test validity may be found sufficiently accurate for operational use. (19:54)

Summary

Two-hundred-five data points will be input into the Schmidt-Hunter Validity Generalization Model. These data include the sample size and correlation coefficients resulting from job performance criterion studies of seven predictors. The cumulative sample sizes range from 951 for the Wonderlic Personnel Test to 5533 for the General Mental Ability Measures. Analysis of the output data will identify the predictors with the greatest predictive validity for the successful performance of first-line supervisors.

III. Findings

Explanation of Output Data

Seven predictors based upon job performance criterion are analyzed using the Schmidt-Hunter Validity Generalization Model. To facilitate analysis the predictors were coded as follows:

01 - Bennett Mechanical Comprehension Test

02 - General Mental Ability Measures

03 - Wonderlic Personnel Test

04 - Otis Mental Ability Test

05 - *How Supervise?*

06 - Leadership Opinion Questionnaire (Initiating Structure)

07 - Leadership Opinion Questionnaire (Consideration)

The measurement criterion, job performance, was the same for all predictors and was based, generally, upon the ratings of second-level supervisors. Since both manufacturing and service organizations were included in the studies, sample job titles were classified either as "foremen" or "supervisors."

The remainder of this section provides a brief description of the output data which was generated for each of the predictors.

Meta-analysis, as discussed in the preceding chapter of this study, is the cumulation of results from numerous studies. The validity generalization model, therefore, cumulates data from a large number of studies and identifies the cumulative sample size for each predictor as the "total N." The total number of data points entered into the model (the correlation coefficients from the data base) are designated "No. Rs" for each predictor. The actual observed standard deviation and the predicted standard deviation are generated for each predictor. It is noted that for all seven predictors used in this study, the predicted standard deviation is lower than the observed standard deviation. Pearlman and others addressed this situation, stating

> Within a given set of validity distributions representing a variety of job family-test type combinations, there are likely to be some distributions in which the three unassessed sources of variance are present to varying degrees . . . . In distributions of [this] type, we would expect the predicted standard deviation to fall below the observed standard deviation to varying degrees. (16:384)

The amount of variance for which correction for sampling error, measurement error (criterion and test reliability), and range variation across the studies has been made through the mathematical formulas inherent in the model is output as "% var acct for." "Residual SD", the residual standard deviation, is the square root of the variance remaining after

mathematical correction for the statistical artifacts (16:383-384). Both "mean R," "an estimate of the fully corrected mean validity" (16:402) coefficient corrected for test and criterion unreliability (measurement error) and range restriction, and "mean R unres," "an estimate of the unrestricted mean validity" (16:402), the mean observed validity coefficient corrected for range restriction but not for attenuation due to measurement error (16:402) are included in the output data. "True R" denotes the mean of the true validity distribution, the correlation coefficient of the predictor, corrected for the three types of artifacts (16:404-405). The "True R" and "True R Attenuated" are identical figures for all predictors used in *this study except* 02, General Mental Ability Measures. This is explained by the fact that the validity coefficient for the distribution of those predictors with a single predictor criterion is 1.0, while Predictor 02, consisting of several different tests, necessitated use of the model's assumed distribution for the predictor validity.

"The validity value at or above which 90% of all estimates of true validities lie" (16:387) is called the credibility value. "95 cred value" and "90 cred value" signify the lower bounds of the confidence intervals at 95% and 90% respectively. Finally, the amount of variance corrected for

due to sampling error only, and the amount of variance corrected for due to the other artifacts are designated as "Sample Size Var" and "Other Artif Var" in the output data.

Analysis

A data table of meaningful output has been constructed for ease of comparison of the significant statistics (see Table I). According to Schmidt and Hunter, "the best estimate of test validity is the mean of the corrected validity distribution" (18:46). Thus, the True R generated by the model is, perhaps, the most meaningful statistic in determining the predictive validity of the particular predictor. Comparing True Rs for the seven predictors, it is readily discernible from the Table that Predictors 05, How Supervise?, and 02, the General Mental Ability Measures, have the highest degree of predictive validity, with correlation coefficients of .692 and .655 respectively. Although the percent of variance accounted for in Predictor 05 is somewhat low (only 25.4%), this would suggest that some degree of situational specificity may be present, but does not alter the high degree of predictive validity evidenced by the correlation coefficient.

Predictors 01 and 04, the Bennett Mechanical Comprehension Test and the Otis Mental Ability Test, also demonstrate a high degree of predictive validity, with correlation

coefficients of .582 and .574 respectively. These predictors also exhibit high percentages of variance accounted for--46.4% for the Bennett Mechanical Comprehension Test and 83.1% for the Otis Mental Ability Test.

Further, it is easily recognized that Predictors 06 and 07, both forms of the Leadership Opinion Questionnaire, show no significant predictive validity for the successful performance of first-line supervisors. Predictors with True Rs near 0, and credibility values which are negative, evidence no predictive validity which can be generalized across studies.

The Wonderlic Personnel Test, Predictor 03, although not as highly predictive as 01, 02, 04, and 05, could prove useful in the absence of availability of any other predictive measures. It is also noteworthy that the percentage of variance accounted for in the Wonderlic is quite high at 67.6%.

TABLE I

Comparison of Meaningful Statistics for Predictors

Pred No.	Total N	Total Rs	True R	% Var Acct For	Res SD	90% c.v.*	95% c.v.*
01	1601	22	.582	46.4	.135	.376	.336
02	5533	80	.655	45.2	.142	.442	.402
03	951	18	.512	67.6	.098	.280	.236
04	1869	29	.574	83.1	.059	.357	.316
05	1441	15	.692	25.4	.210	.490	.451
06	1228	19	.060	85.4	.051	-.143	-.182
07	1477	22	.134	28.2	.188	-.060	-.097

*c.v. = credibility value

IV. Conclusion

Based upon the research conducted in this study it is concluded that at least two predictors can be administered to first-time supervisory candidates which provide significant correlation with successful job performance. Those predictors are the General Mental Ability Measures (including such tests as the Otis, Wonderlic, Scholastic Aptitude Test [SAT], Army General Classification Test, Purdue Adaptability Test, Thurstone Mental Abilities Test, and the General Aptitude Test Battery), and the How Supervise? questionnaire.

Additionally, two other predictors, the Bennett Mechanical Comprehension Test and the Otis Mental Ability Test, have a high degree of predictive validity for successful first-line supervisory performance. Another predictor, the Wonderlic Personnel Test, although not as highly predictive as those cited above, may prove valuable if none of the other predictors are available to the personnel selectors.

Findings from numerous studies provide a useful data base for the Schmidt-Hunter Validity Generalization Model. Analysis of the output from this model identifies existing criteria for predicting the probability of successful

performance by first-time, first-line supervisors. The use of predictors found to have significant correlation with successful job performance by managers when selecting first-time, first-line supervisors is expected to substantially increase the probability that the selected candidate will perform at a higher-than-average level in his/her new position.

Bibliography

1. Algera, Jen A. and others. "Validity Generalization: Some Critical Remarks on the Schmidt-Hunter Procedure," Journal of Occupational Psychology, 57:197-210 (1984).

2. Bass, Bernard M. "Leadership Opinions as Forecasts of Supervisory Success," Journal of Applied Psychology, 40 (5):345-346 (1956).

3. Beach, Dale S. Personnel: The Management of People at Work (Third Edition). New York: Macmillan Publishing Co., Inc., 1975.

4. Bittel, Lester R. Improving Supervisory Performance. New York: McGraw-Hill Book Company, 1976.

5. Broadwell, Martin M. Moving Up to Supervision. Boston: CBI Publishing Company, Inc., 1979.

6. Burke, Michael J. "Validity Generalization: A Review and Critique of the Correlation Model," Personnel Psychology, 37:93-115 (1984).

7. Decker, Robert L. "An Item Analysis of How Supervise? Using Both Internal and External Criteria," Journal of Applied Psychology, 40 (6):406-411 (1956).

8. Dunnette, Marvin D. and Walter C. Borman. "Personnel Selection and Classification Systems," Annual Review of Psychology, 30:477-525 (1979).

9. Gough, Harrison G. "A Managerial Potential Scale for the California Psychological Inventory," Journal of Applied Psychology, 69 (2):233-240 (1984).

10. Guion, Robert M. Personnel Testing. New York: McGraw-Hill Book Company, 1965.

11. Hough, Leaetta M. "Development and Evaluation of the 'Accomplishment Record' Method of Selecting and Promoting Professionals," Journal of Applied Psychology, 69 (1):135-146 (1984).

12. Hunter, John E., Frank L. Schmidt, and Gregg B. Jackson. Meta-Analysis. Beverly Hills: Sage Publications, Inc., 1982.

13. Loken, Robert D. and Earl P. Strong. Supervision in Business and Industry. New York: Magazines of Industry, Inc., 1949.

14. Mowry, Harley W. "A Measure of Supervisory Quality," Journal of Applied Psychology, 41 (6):405-408 (1957).

15. Northrup, Herbert R. and others. The Objective Selection of Supervisors. Philadelphia: University of Pennsylvania, 1978.

16. Pearlman, Kenneth, Frank L. Schmidt, and John E. Hunter. "Validity Generalization Results for Tests Used to Predict Job Proficiency and Training Success in Clerical Occupations," Journal of Applied Psychology, 65 (4):373-406 (1980).

17. Rowland, Kendrith M. and William E. Scott, Jr. "Psychological Attributes of Effective Leadership in a Formal Organization," Personnel Psychology, 21:365-377 (1968).

18. Schmidt, Frank L. and John E. Hunter. "Development of a General Solution to the Problem of Validity Generalization," Journal of Applied Psychology, 62 (5):529-540 (1977).

19. -----. "The Future of Criterion-Related Validity," Personnel Psychology, 33:41-60 (1980).

20. Shane, Guy Stanley. The Schmidt-Hunter Approach to Validity Generalization: An Application to Supervisory Selection. PhD dissertation. The Graduate School of Arts and Sciences, George Washington University, Washington DC, May 1978.

21. Shulman, L. S. Review of Research in Education, 5. Itasca IL: American Educational Research Association, 1977.

22. Spitzer, Morton Edward and Walter J. McNamara. "A Managerial Selection Study," Personnel Psychology, 17:19-40 (1964).

23. Sweetland, Richard C. and Daniel J. Keyser. Tests. Kansas City: SKS Associates, 1983.

24. Thorndike, Robert L. Personnel Selection. New York: John Wiley and Sons, Inc., 1949.

25. Turnage, Janet J. and Paul M. Muchinsky. "A Comparison of the Predictive Validity of Assessment Center Evaluations Versus Traditional Measures in Forecasting Supervisory Job Performance: Interpretive Implications of Criterion Distortion for the Assessment Paradigm," Journal of Applied Psychology, 69 (4):595-602 (1984).

26. Weger, John J. Motivating Supervisors. USA: American Management Association, Inc., 1971.

VITA

Frances A. Burke was born 14 August 1938 in Enid, Oklahoma. She graduated from high school in Wichita, Kansas, in 1955 and began her career in federal service at McConnell AFB, Kansas, in May 1956. In July 1962 she transferred to the Defense Electronics Supply Center, Dayton, Ohio, as a Procurement Agent. She transferred to Wright-Patterson AFB, Ohio, in September 1965 as a Contract Negotiator in the Aeronautical Systems Division. In December 1967 she returned to the Defense Electronics Supply Center, continuing in the Directorate of Contracting and Production as a Contract Negotiator, and later as a Contract Specialist, and Supervisory Contract Negotiator/Contracting Officer. She was selected for the position of Chief, Inventory Accounting Branch/Accountable Officer in the Directorate of Supply Operations in September 1982 where she served until entering the School of Systems and Logistics, Air Force Institute of Technology, in June 1984.

Permanent address: 207 Beverly Place
Dayton, Ohio 45419

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

AD-A160 87c

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) AFIT/GLM/LSM/85S-9	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION School of Systems and Logistics 6b. OFFICE SYMBOL (If applicable) AFIT/LSM	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State and ZIP Code) Air Force Institute of Technology Wright-Patterson AFB, Ohio 45433	7b. ADDRESS (City, State and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State and ZIP Code)	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT NO.
11. TITLE (Include Security Classification) See Block 19	
12. PERSONAL AUTHOR(S) Frances A. Burke, B.A., GS-12	
13a. TYPE OF REPORT MS Thesis 13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Yr., Mo., Day) 1985 September 15. PAGE COUNT 38
16. SUPPLEMENTARY NOTATION	

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	
05	09	n/a	Civilian Personnel, Personnel Selection, Supervisory Personnel, Promotion Evaluation

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Title: CRITERIA FOR SELECTION OF FIRST-LINE SUPERVISORS

Thesis Advisor: Dr. John A. Muller
Associate Professor of English

Approved for public release IAW AFR 190-1.
LYNN E. WOLAVER 11 Sept 85
Dean for Research and Professional Development
Air Force Institute of Technology (AFSC)
Wright-Patterson AFB OH 45433

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION	
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	UNCLASSIFIED	
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. John A. Muller	22b. TELEPHONE NUMBER (Include Area Code) 513-255-6761	22c. OFFICE SYMBOL AFIT/LSM

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

This ~~study~~ identified criteria with predictive validity for the selection of successful first-time, first-line supervisors. Through meta-analysis it is possible to generalize the validity of predictors across cumulative studies. A validity generalization model, which corrects for artifactual variance, was used to process data from the findings of many previous studies based upon the job performance measurement criterion of first-line supervisors.

Analysis revealed two predictors, How Supervise? and the General Mental Ability Measures, with substantial validity for predicting successful performance of first-line supervisors. The Bennett Mechanical Comprehension Test and the Otis Mental Ability Test also showed relatively high predictive validity; however, neither form of the Leadership Opinion Questionnaire evidenced predictive validity for successful job performance by first-line supervisors. The Wonderlic Personnel Test, although not highly predictive, may be useful in the absence of the other predictors.

These findings may be of value to middle managers during the selection process in identifying promotion candidates with potential for successful performance as first-time, first-line supervisors.

END

FILMED

12-85

DTIC