

AD-A164 837

JOB PERFORMANCE MEASUREMENT CLASSIFICATION SCHEME FOR
VALIDATION RESEARCH. (U) MCFANN GRAY AND ASSOCIATES INC
MONTEREY CA M J KAVANAGH ET AL. FEB 86 AFHRL-TR-85-51
F41689-81-C-0022

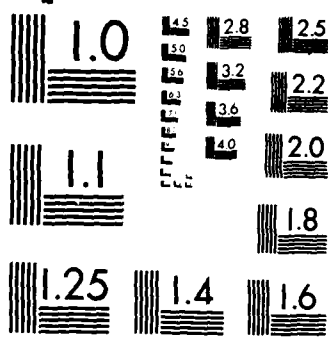
1/1

UNCLASSIFIED

F/G 5/9

NL





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

AIR FORCE



AD-A164 837

HUMAN RESOURCES

**JOB PERFORMANCE MEASUREMENT CLASSIFICATION SCHEME
FOR VALIDATION RESEARCH IN THE MILITARY**

Michael J. Kavanagh

School of Business
State University of New York at Albany
Albany, New York 12222

Walter C. Borman

Personnel Decisions Research Institute
43 Main Street, S.E.
Suite 405
Minneapolis, Minnesota 55414

Jerry W. Hedge
R. Bruce Gould

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

February 1986

Interim Paper for Period May 1982 - November 1982

Approved for public release; distribution unlimited.

LABORATORY

NTC FILE COPY

**AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5601**

SELECTED
MAR 05 1986

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Public Affairs Office has reviewed this paper, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This paper has been reviewed and is approved for publication.

NANCY GUINN VITOLA, Technical Director
Manpower and Personnel Division

RONALD L. KERCHNER, Colonel, USAF
Chief, Manpower and Personnel Division

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

-10-164, 37

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)			5. MONITORING ORGANIZATION REPORT NUMBER(S) AFHRL-TR-85-51		
6a. NAME OF PERFORMING ORGANIZATION McFann - Gray & Associates, Inc.		6b. OFFICE SYMBOL (if applicable)		7a. NAME OF MONITORING ORGANIZATION Manpower and Personnel Division	
6c. ADDRESS (City, State, and ZIP Code) 2100 Garden Road, Suite J Monterey, California 93940			7b. ADDRESS (City, State, and ZIP Code) Air Force Human Resources Laboratory Brooks Air Force Base, Texas 78235-5601		
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Air Force Human Resources Laboratory		8b. OFFICE SYMBOL (if applicable) HQ AFHRL		9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F41689-81-C-0022	
8c. ADDRESS (City, State, and ZIP Code) Brooks Air Force Base, Texas 78235-5601			10. SOURCE OF FUNDING NUMBERS		
			PROGRAM ELEMENT NO. 62703F	PROJECT NO. 7719	TASK NO. 18
			WORK UNIT ACCESSION NO. 21		
11. TITLE (Include Security Classification) Job Performance Measurement Classification Scheme for Validation Research in the Military					
12. PERSONAL AUTHOR(S) Kavanagh, Michael J.; Borman, Walter C.; Hedge, Jerry W.; Gould, R. Bruce					
13a. TYPE OF REPORT Interim		13b. TIME COVERED FROM May 82 TO Nov 82		14. DATE OF REPORT (Year, Month, Day) February 1986	
				15. PAGE COUNT 20	
16. SUPPLEMENTARY NOTATION					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)		
FIELD	GROUP	SUB-GROUP			
05	09		criterion development		
05	10		job performance assessment		
			performance measurement		
19. ABSTRACT (Continue on reverse if necessary and identify by block number)					
This paper outlines the development of a performance measurement classification scheme. Its major focus is on the quality of measurement when the purpose is for validation research in the military. It describes the background and purpose for development of the classification scheme and explains the information and procedures used with this approach. In order to develop a measurement methodology for job performance, a classification scheme of job performance measurement quality was needed (a) to summarize and organize research progress in terms of previous empirical work and (b) to identify future research needs. A more detailed presentation of this conceptually based model will be available in a forthcoming report, which will include a literature review and specific directions for future job performance measurement research.					
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION Unclassified		
22a. NAME OF RESPONSIBLE INDIVIDUAL Nancy A. Perrigo, Chief, STINFO Office			22b. TELEPHONE (Include Area Code) (512) 536-3877		22c. OFFICE SYMBOL AFHRL/TSR

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted
All other editions are obsolete.SECURITY CLASSIFICATION OF THIS PAGE
Unclassified

JOB PERFORMANCE MEASUREMENT CLASSIFICATION SCHEME
FOR VALIDATION RESEARCH IN THE MILITARY

Michael J. Kavanagh

School of Business
State University of New York at Albany
Albany, New York 12222

Walter C. Borman

Personnel Decisions Research Institute
43 Main Street, S.E.
Suite 405
Minneapolis, Minnesota 55414

Jerry W. Hedge
R. Bruce Gould

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

Reviewed by

Rodger Ballentine, Lt Colonel, USAF
Chief, Productivity and Performance Measurement Function

Submitted for publication by

R. Bruce Gould

Chief, Force Utilization Branch
Manpower and Personnel Division

This report is primarily a working paper. It is published solely to document work performed.

SUMMARY

This paper provides an overview of the background and development of a classification scheme of job performance measurement quality to be used for "research purposes only." It describes, in brief form, the development of a conceptual framework to be used in the planning and conduct of a long-term research and development effort to obtain job performance criterion measures in the military.



Accession For	
NTIS	☒
CRA&I	☐
DTIC	☐
TAB	☐
Unannounced	☐
Justification	
By _____	
Distribution / _____	
Availability Codes	
Dist	Avail and/or Special
A-1	

PREFACE

This paper describes the initiation of a long-term program of research and development focusing on job performance criterion development. As such, it represents the Air Force's initial contribution to a joint-Service effort to test the feasibility of linking enlistment standards and job performance. A much more detailed presentation of the Air Force approach to this research effort will be included in a forthcoming Technical Report.

The work was performed by McFann-Gray and Associates, Inc., under contract F41689-81-C-0022 from the Air Force Human Resources Laboratory (AFHRL) Manpower and Personnel Division. The work unit number for AFHRL was 77191821. Dr. R. Bruce Gould was the AFHRL Contract Monitor.

TABLE OF CONTENTS

	Page
I. INTRODUCTION.	1
II. CLASSIFICATION SCHEME	1
III. VALIDATION RESEARCH CLASSIFICATION SCHEME	7
IV. SUCCESSIVE APPROXIMATION: A RESEARCH STRATEGY.	10
V. TIME-TO-PROFICIENCY	11
VI. CONCLUSIONS AND RECOMMENDATIONS	11
REFERENCES.	12

LIST OF FIGURES

Figure	Page
1 A Job Performance Measurement Classification Scheme.	3
2 A Job Performance Measurement Classification Scheme for Validation Research.	8

LIST OF TABLES

Table	Page
1 Variables That Can Impact on Job Performance Measurement Quality	5
2 Quality of Job Performance Measurement Criteria.	9

JOB PERFORMANCE MEASUREMENT CLASSIFICATION SCHEME FOR VALIDATION RESEARCH IN THE MILITARY

I. INTRODUCTION

Over the past 2 years, reviews of the research and development (R&D) thrusts of the Manpower and Personnel Division of the Air Force Human Resources Laboratory (AFHRL) have strongly recommended that a major criterion development effort be initiated. Many individual R&D projects were involved in the development of job performance criteria to validate selection, classification, and training systems. This was thought to be inefficient, and a unified effort to develop a strategy for job performance criterion development was recommended for all future R&D projects.

A second driving force behind the development of a criterion measurement methodology was the recognition that the Armed Forces Vocational Aptitude Battery (ASVAB) must be validated against on-the-job performance. Historically, the ASVAB (and its predecessors) has been validated predominantly against training school success. The need to establish ASVAB relevance to job performance increased when Congress began asking why the ASVAB was not validated against this criterion.

In their 1982 recommendation regarding the research thrusts in the Manpower and Personnel Division, a Research Advisory Panel composed of scientist representatives from the Department of Defense (DOD), academia, and industry noted that it was not necessary for the Laboratory to "start from scratch" in terms of this criterion development effort. A number of measurement characteristics have been well-established in the empirical literature regarding criterion development. The panel recommended close examination of this literature as a basis for development of a structural model of job performance measurement quality to guide a research program to develop a criterion measurement methodology for the military. This paper describes the first step in the model-building process, namely a conceptual approach to the organization of the literature and the generation of research initiatives.

Implementation of these recommendations required extensive resources to completely research and develop a performance measurement system for use in validation research. After about 9 months of research planning, it appeared highly probable that resources necessary for this project were going to be made available starting in fiscal year 1983 for a long-term effort. Availability of these resources was expected to assure considerable reduction in magnitude of "the criterion problem," at least in terms of validation research in the military.

There are two reasons why the "criterion problem" has persisted. First, it has been a matter of convenience for some researchers, who use it to indirectly justify their use of easily accessible and often-used measures of job performance when they conduct validation research. Second, and more importantly, resolving the criterion problem involves extensive resources; in the past, no one has been willing to commit that needed level of resources to conduct the research. Thus, this planned AFHRL criterion research effort will serve two purposes: (a) It will develop a methodology for job performance measurement in the military, and (b) it will contribute to scientific research on performance measurement in the broader scientific community.

II. CLASSIFICATION SCHEME

The overall purpose of this program of R&D in the AFHRL is the development of a measurement methodology for job performance. To accomplish this purpose, it is necessary to develop a classification scheme for variables related to job performance measurement quality which (a)

summarizes and organizes previous empirical research and (b) identifies future research needs. The conceptual framework is derived from the theoretical and empirical literature in the field of psychology. Figure 1 is a graphic depiction of the variables in this conceptual framework. Note that this is a general performance measurement classification scheme. A performance measurement schemata for the purpose of validation research only will be discussed later. It is necessary to briefly cover the more general case in order to understand the model-building process.

The general classification scheme was developed using a combination of several approaches. By making performance measurement quality the outcome variable in Figure 1, the research focus centered on those variables that have, or can have, an impact on the quality of the measure. In evaluating the literature, these variables were identified and placed in the schemata. The input variables in the first box on the left side of Figure 1 are those characteristics of the individual or measurement system that can affect the quality of the measurement. The process variables in the center of the figure reflect the current thinking in the performance measurement literature that these variables play an important and pervasive role in the appraisal process. There has been a recent emphasis on cognitive variables (variables of importance in the decision-making process) (Feldman, 1981; Landy & Farr, 1980), as well as the acceptability/confidence users have in the system (Dipboye & dePontbriand, 1981; Kavanagh & Hedge, 1983; Landy, Barnes, & Murphy, 1978) and their hypothesized effects on measurement quality. In addition, the motivation the raters and ratees bring to the appraisal process (DeCotiis & Petit, 1978) and their trust in the appraisal system (Bernardin, Orban, & Carlyle, 1981) are considered important process variables. Although not much empirical evidence has been reported in the literature with respect to the role of these variables, the indications are that they act as intervening process variables. Thus, while the individual/system characteristics have been recognized as influencing measurement quality, these intervening process variables are hypothesized to be functionally related to both the independent and dependent variables; therefore, for purposes of classification, these variables have been separated from the other variables.

Note that the cognitive variables have been placed outside the main causal path. This indicates that these variables may not always play an important role in the appraisal process. When the measurement system relies heavily on human judgment, such as with ratings or trained observers, these variables would be expected to influence measurement quality. When human judgment is not as important, such as with productivity counts or absences, the impact of these cognitive variables on the dependent variable should be greatly reduced.

Note further in Figure 1 that the cognitive variables have been divided into two categories. The first category involves input and storage of information, which is primarily concerned with the observational heuristics that people use when gathering information about an individual's job performance. The other category involves judgment or decision heuristics that people use in assigning a quantitative index to the performance of a person on the job.

A brief example will illustrate the hypothesized relationships between these variables in the conceptual framework. A general hypothesis underlying this conceptualization is that the more complex (i.e., sophisticated, yet not necessarily cumbersome) the observational and/or decision heuristics used, the higher the quality of the performance measurement. However, individual/system characteristics could affect the complexity of these cognitive processes and thus lower or raise the quality of the measures. The best example for the military is in terms of the impact of organizational or unit norms in the current operational performance measurement systems. A strong norm exists to give enlisted military personnel high ratings (i.e., an "8" or a "9") on their performance evaluations. Regardless of the cause of this norm, its effect is to simplify the rater's cognitive processes, at least in terms of completing these annual evaluations. A rater, in this situation, may or may not use complex observational heuristics; however, the rater's decision heuristic is simple--"8" or "9."

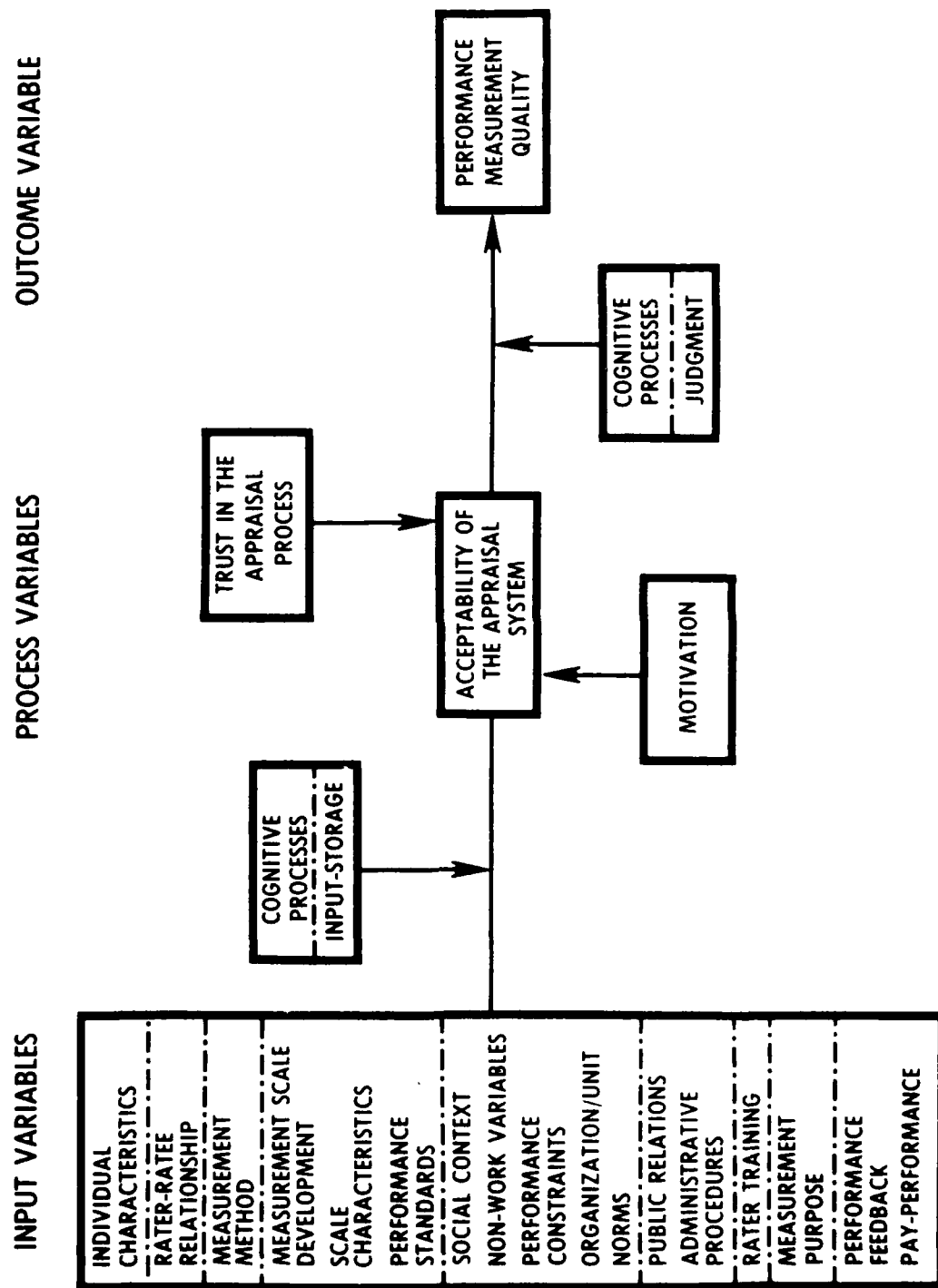


Figure 1. A Job Performance Measurement Classification Scheme.

Returning to Figure 1, two other approaches or perspectives were used to generate the classification scheme. The first of these perspectives was borrowed from test score theory. The model chosen was Spearman's classic test score model because of its simplicity and wide dissemination. The notion that an observed score, a performance measurement score, could be divided into true and error components suggested an approach to examine the impact of the variables that were identified on the quality of the measure. This approach also allowed an examination of the literature in terms of those variables that affect true variance and those that affect only error variance in the performance measurement situation. In approaching the literature in this way, it became obvious that one would want to minimize those factors that affect only error variance, while increasing the impact of those factors that influence primarily true variance. If the literature is examined in this manner, the resulting integration has clear implications for future research strategies.

The second perspective used to refine the classification scheme was to organize into categories the many variables that affect measurement quality. This allowed an identification of the linkages in the system, for use in classifying the empirical literature in an organized fashion. These linkages also provided a framework for classification of needed research in the AFHRL program. This allowed a careful examination and prioritization of the total research domain. The results of this classification of variables are contained in Table 1. This is a fairly exhaustive list of the variables that have been empirically demonstrated to affect performance measurement quality.

Returning to Figure 1, perhaps the most critical input variable, in terms of its impact on rating quality, is the measurement purpose. As Zedeck and Cascio (1982) have noted, if a measurement system is to be used for promotion or pay increases, it creates an entirely different situation in terms of the quality of the measurement than if the system is used for validation research purposes. In fact, if the conceptual framework in Figure 1 is constrained in terms of the purpose for which it is being developed and used, it changes the strength or importance of the individual/system characteristics with respect to their impact on measurement quality. For example, the pay-performance relationship with measurement quality is extremely important in a measurement system being used for administrative purposes, but has little effect if the system is used only for validation research.

According to Kavanagh (1982), there are four major purposes for performance measurement systems: (a) administrative decisions, (b) employee growth and development, (c) validation research, and (d) requirement to meet legal guidelines. The strength of the relationships between the individual/system characteristics and measurement quality changes as a function of changing the purpose. A good analogy is to consider the system characteristics as independent variables, and measurement quality as the dependent variable in a multiple regression equation. The intervening process variable could also be included, and thus the equation would have moderator terms. One would expect the beta weights for the system characteristics and the process variables to change as the measurement purpose changes. Since the initial thrust of the Air Force effort is for research validation only, it is necessary to impose this measurement use restriction, and the resulting changes to the general classification scheme in Figure 1.

Table 1. Variables That Can Impact on Job Performance Measurement Quality

1. Individual characteristics
 - a. Cognitive variables: Rater or ratee
 - b. Rater/ratee intelligence
 - c. Rater/ratee knowledge of the job being evaluated
 - d. Rater/ratee personal characteristics
 - e. Rater/ratee interpersonal trust
2. Relationship between ratee and rater/observer
 - a. Sex congruence
 - b. Race congruence
 - c. Job tenure together
 - d. Age congruence
 - e. Off-the-job relationship
 - f. History of conflict or cooperation
3. Method/Source of measurement
 - a. Supervisor ratings
 - b. Peer ratings
 - c. Self-ratings
 - d. Subordinate ratings
 - e. Assessment center (team) ratings
 - f. Work samples/simulations
 - g. Productivity records
4. Scale development
 - a. Critical incidents used
 - b. Based on job description/job requirements
 - c. Employee participation
 - d. Top management support during development
5. Rating scale characteristics
 - a. Content of the scale
 - b. Anchors versus no anchors
 - c. Behaviors versus traits
 - d. Format type
 - e. Number of anchors/scale points
 - f. Single versus multiple dimensions
 - g. Scaling metric/approach
6. Performance standards/goals
 - a. Present or not
 - b. Standards versus goals
 - c. Participatively set and communicated
 - d. Specificity of behavior or accomplishment expected
7. Social context
 - a. Performance level of others in work group
 - b. Existence of group norms
 - c. Rater's status in group
 - d. Ratee's status in group

Table 1. (Concluded)

-
8. Non-job variables
 - a. Marital status
 - b. Pre-school children at home
 - c. Dual career family
 - d. Participation in company activities off the job
 - e. Stressful life events in recent past
 9. Performance constraints
 - a. Poor information
 - b. Equipment deficiency
 - c. Supplies deficiency
 - d. Time limitations
 - e. Poor work environment
 10. Organizational/unit norms
 - a. Expectation of certain level of performance by upper management
 - b. Expectation by immediate supervisor regarding level of performance
 - c. Presence of a union
 - d. Pay/rewards tied to performance levels by contract
 - e. Pay/rewards tied to performance levels by informal norms
 11. Public relations/administrative procedures
 - a. Required or not
 - b. Mode of presentation
 - c. Content of procedure
 12. Rater training
 - a. Content of training
 - b. Format of training
 - c. Length of training
 13. Measurement purpose
 - a. Validation research only
 - b. Employee growth and development
 - c. Administrative purposes such as rewards
 - d. To meet legal guidelines
 14. Performance feedback
 - a. Required or not
 - b. Sources of feedback
 - c. Participative
 - d. Clarity of feedback
 - e. Frequency of feedback
 15. Pay-performance relationship
 - a. Are they related in the system?
 - b. Equity of the relationship
-

III. VALIDATION RESEARCH CLASSIFICATION SCHEME

The performance measurement classification scheme for validation research is depicted in Figure 2. The dependent variable, measurement quality, is something that has not been clearly defined in the literature. Different researchers (e.g., Borman, 1975, 1979; Kavanagh, MacKinney, & Wolins, 1971; Latham, Wexley, & Pursell, 1975) have used different criteria to assess measurement quality, and some of these criteria are listed in Table 2. Accuracy and construct validity are the crucial criteria to judge the quality of the measurement of job performance, and the literature that has relied only on the other criteria cannot be used to reach definitive scientific conclusions. The first four criteria are seen as important, but less critical, in that satisfying the requirements of the first four criteria does not guarantee the measure will meet accuracy/construct validity requirements. However, an accurate and construct valid measure will in all likelihood meet a number of these other criteria as well. In other words, if it can be demonstrated that a given performance measure is accurate, the evidence for the other quality indices will likely be positive. This logic is consistent with current theory in measurement (Nunnally, 1978) and performance ratings (Wherry & Bartlett, 1982).

Another aspect of Figure 2 that is critically important is the measurement method input variable. Like the variable of measurement purpose in Figure 1, constraining the orientation presented in Figure 2 by measurement method should affect the relationships within the model. If this is so, it may indicate that different measurement methods are capturing different pieces of the performance criterion space. That is, as noted by Borman (1974) and Schaefer (1977), supervisory ratings may well be assessing a different portion of the total job performance criterion space than are peer ratings, self-ratings, work sample tests, or objective indices of productivity. This is not meant to imply that there is no overlap among these methods in the part of the criterion space they are measuring; however, they are all measuring some unique aspects of the criterion space that have been treated frequently as error in research.

In the typical research to validate multiple measures of job performance, one or more methods have been eliminated because of low intercorrelations with the other methods. These low intercorrelations between different methods have been assumed to be the result of error in the measures. In a conceptualization of different measurement methods measuring different parts of the criterion space with differing degrees of fidelity, low correlations between measures may not indicate error. Thus, it can be argued that the typical validation approach may not be the best for assessing measurement quality.

These issues lead directly to the question of the content of the criterion space -- an issue that must first be addressed prior to explaining the importance of different measuring methods. The content issue addresses the question: "What is it, in terms of performance dimensions, that constitutes job performance?" In examining the literature, and from personal experience with developing performance measurement systems, approximately 12 to 15 performance dimensions have been identified that appear consistently. Furthermore, these dimensions seem to fit into two general categories or areas that define job performance: technical competence skills and job-relevant interpersonal skills. Although this may seem to be oversimplifying the criterion, it is supported by factor analytic studies in which two factors, roughly representing these two broad skill areas, have emerged (Borman, 1981; Borman, Mendel, Lammlein, & Rosse, 1981). Simplifying the criterion space in this manner allows scientists to better specify the needed research in the AFHRL criterion development project.

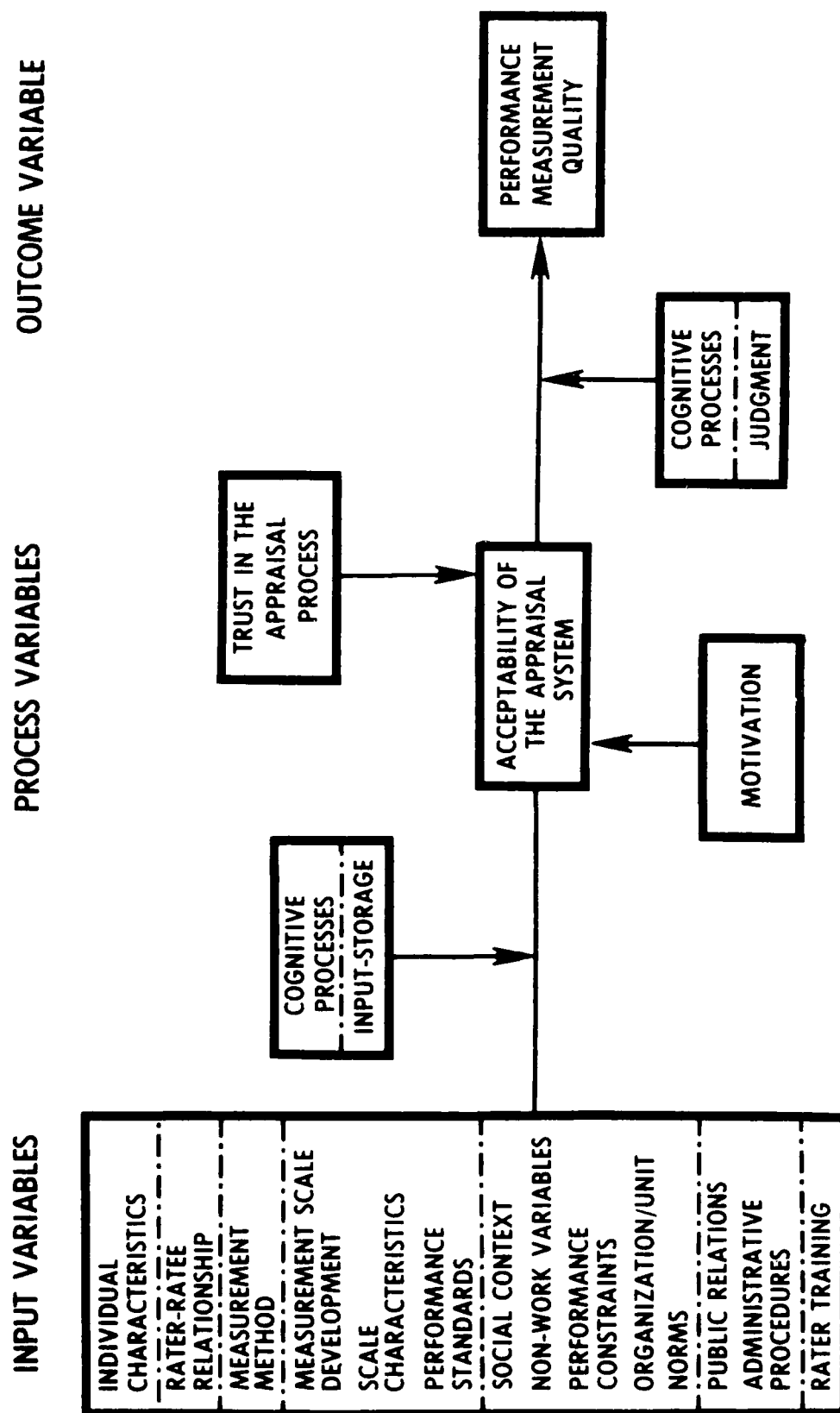


Figure 2. A Job Performance Measurement Classification Scheme for Validation Research.

Table 2. Quality of Job Performance Measurement Criteria

-
1. Psychometric errors: Halo, Leniency, Range Restriction
 2. Inter-rater reliability
 3. Content validity
 4. Discriminability: Separates individuals in terms of performance levels
 5. Construct validity
 6. Accuracy
-

To return to the idea of multiple methods research, the following methods of measuring job performance are the most frequently used: (a) supervisory ratings, (b) peer ratings, (c) self-ratings, (d) work samples, and (e) objective indices of productivity. The first three are widely used and will be applied to the present effort. However, rather than using a traditional work sample methodology, an alternative to this approach will be developed and tested. The new methodology is called Walk-Through Performance Testing (WTPT) and is being developed specifically for this project at the AFHRL. The WTPT methodology combines aspects of both observer interviewing and work sampling, but in addition, is designed to overcome certain limitations associated with the generic tasks used with work sampling. The method will be developed by accessing the Air Force database (see Christal, 1974) that contains information on the tasks performed in enlisted specialties. These tasks will form the basic content of the measurement scale. WTPT administrators will be trained to use these scales in terms of incumbent effective and ineffective performance on each of the tasks. The administrators will then examine the job incumbent by asking the person to perform certain tasks or to explain task procedures for that job. The administrators will record the person's behavior or answers on a rating checklist of tasks. The important characteristic of this method is that the job is being reduced to its smallest parts at the task level and will include not only a core set of tasks, but a series of unique tasks as well. Thus, this method will constitute an examination of job performance at the most "micro" level.

It is believed that the WTPT method will assess, with a high degree of fidelity, technical skills and competence -- one-half of the criterion space. In fact, walk-through testing may be one method that will allow removal of the interpersonal/social aspects of the job situation. However, as currently planned, it may be less accurate in assessing the job-relevant interpersonal skills side of the criterion space. On the other hand, supervisory ratings may be quite good at assessing interpersonal skill, but not very accurate in measuring technical skills, particularly if the job is one that has had significant changes in technology or is one in which the supervisor has never had direct work experience. The same logic can be applied for the other three methods. They are all measuring portions of the criterion space; however, they differ in their fidelity or accuracy of measurement across the different parts.

This does not mean that one or more of the methods could not be modified to assess both major parts of the job performance criterion space. The WTPT method could be modified, for example, to measure interpersonal skills. However, this modification may not be cost-effective if another method is available that can accurately assess the interpersonal skills without modification

(e.g., peer ratings). The central point of this argument is that the five methods, as they are currently used, assess different parts of the criterion space with differing degrees of accuracy. Any research in this program of criterion development must recognize this fact, and research must be designed with this point in mind.

This type of logic makes the typical multimethod validation study problematic. If job performance is measured with two or more methods, and zero-order correlations are calculated among the methods, the conclusions regarding the validity of the methods are based in part on significant non-zero correlations between methods for common traits. Methods showing non-significant values are rejected. However, if, as has been argued, the methods are not assessing the same portions of the criterion space with equal fidelity, then there is no reason to expect them to be correlated.

The extension of the logic that the different methods measure different portions of the criterion space with varying degrees of fidelity leads directly to the idea of specifying the construct space for job performance in terms of what Cronbach and Meehl (1955) have termed a nomological network (a network of relations that are tied to observables and, hence, are empirically testable). In this framework, the measures are the observables, and the construct is used to account for relationships among them. This suggests the use of one of the techniques for construct validation discussed by Nunnally (1978), namely, that of testing the a priori hypothesized relationships within a construct space with empirical data. In order to do this, the two major parts of the criterion space, technical competence and interpersonal skills, must be better specified in terms of the job performance dimensions that comprise these general categories.

With the delineation of a multidimension-multimethod matrix, the next step in this research strategy would be to hypothesize the expected level of relationship between each method-dimension and all others. Some might be high, others moderate, and some zero. In this manner, one would have specified, a priori, the hypothesized nomological net for these methods and the criterion space. After collecting data, the results would be examined to verify the expected correlations. In this strategy, a zero relationship would be as important as a non-zero one in establishing the construct validity of the methods of measurement.

For the Air Force research program, this empirical construct validation project cannot be done until at least the fifth or sixth year. First, the methods must be properly researched and refined so that this approach to construct validation is not subject to criticism for poorly designed measures. One implication of this line of thinking for the Air Force program is that there is a tremendous amount of research to be done during the first 4 or 5 years of the program before this type of study can be conducted.

IV. SUCCESSIVE APPROXIMATION: A RESEARCH STRATEGY

Another research strategy that will underlie the Air Force program is the notion of successive approximations to high fidelity measures. It is expected that, as currently envisioned, WTPT will accurately assess technical job skills of individuals. This method is quite time-consuming and expensive, particularly if it is going to be used for large-scale data collection across the Air Force. If, after the appropriate research, it is believed that WTPT does have high fidelity for technical skills, thorough research can be attempted to determine which one of the less expensive job performance data collection methods most closely approximates WTPT. It may be, for example, that some combination of peer and self-ratings are a close enough approximation to WTPT, for the technical job skills, to allow the walk-through testing method to be dropped in favor of these less expensive ones. Also, as earlier noted, WTPT may be modified to measure interpersonal skills. If it can be modified to measure accurately these different

parts of the job performance criterion space, the same research strategy of successive approximation with less costly and time-consuming methods can be done.

V. TIME-TO-PROFICIENCY

An additional variable, time-to-proficiency on job tasks, may need to be incorporated into the research program. It appears that it might be embedded within all of the methods of measurement. Research will be necessary to determine how best to adapt the five methods to measure this crucial part of job performance. Wide individual differences on this variable would be expected in newly assigned personnel, particularly in their first job in the military. Furthermore, it may be that some of the methods are able to measure this variable across task performance with varying degrees of fidelity/accuracy. In fact, there probably is a task/dimension by method effect in terms of accuracy of assessing time-to-proficiency. The important point is that this variable must be included in any research project. As a footnote, collecting this type of performance data may be most appropriate for the purpose of validating the learning/acquisition rate measures that are being collected in the Learning Abilities Research Function at AFHRL.

In terms of validating the ASVAB, this type of logic has certain implications. Validating the ASVAB against training school criteria has not been totally effective because training school success reflects only a part of the total job performance criterion space, namely, initial acquisition of technical competence and skills in a school environment. Since the ASVAB has been designed to measure and predict rate of learning on job skills, it seems reasonable that it should predict training school success. However, job performance in the military also includes job-relevant interpersonal skills, and thus, using the measurement of total job performance to validate the ASVAB may be less than ideal. In fact, measures of technical competence on the first military job, by whatever method is found to have the greatest accuracy, would probably be the most appropriate criterion on which to validate the ASVAB. An extension of this logic would indicate that other predictors are needed (e.g., vocational interests) for selecting persons who will succeed in terms of the job-relevant interpersonal skills.

VI. CONCLUSIONS AND RECOMMENDATIONS

The classification scheme pictured in Figure 2 will be used to summarize and organize the previous research as well as to specify needed future research. If a line is drawn between any of the variables (either individual/system characteristics or the process variables) and the dependent variable, a linkage in the system is defined. These linkages will be used to organize the literature search as well as to make recommendations for needed research in the Air Force program. The literature will be categorized in the appropriate linkage in the model, and we should be able to draw some conclusions regarding the known empirical "facts" within each linkage. Additional research needs for effective criterion development can then be identified. For example, in the linkage involving rater training, it is known that some rater training is necessary to ensure the accuracy and general quality of the ratings for this method (see Borman, 1980; Landy & Farr, 1980). This training could be quite minimal, e.g., an explanation of how to complete the form. It is clear from the literature that the absence of rater training has negative impact on the quality of the measurement. Note that this argument also applies to observer training in the walk-through testing method. It is not known what specific type of rater/observer training is best in terms of obtaining accurate measures.

Implementation of the procedures described in the previous paragraph will result in a technical report containing the following sections: (a) a background and description of the problem of criterion development, (b) a description of the approach to model development (mostly

detailed in this paper), (c) a review of the major and relevant literature organized by the linkages in the classification scheme, (d) a detailed specification of the "known" facts and needed research for successful criterion development, and (e) a discussion of the research priorities, both in terms of importance and time sequencing, for the Air Force program of research.

REFERENCES

- Bernardin, H. J., Orban, J. A., & Carlyle, J. J. (1981). Performance ratings as a function of trust in appraisal, purpose for appraisal, and rater individual differences. Proceedings of the 41st Annual Meeting of the Academy of Management, San Diego, California.
- Borman, W. C. (1974). The rating of individuals in organizations: An alternate approach. Organizational Behavior and Human Performance, 12, 105-124.
- Borman, W. C. (1975). Effects of instructions to avoid halo error on reliability and validity of performance evaluation ratings. Journal of Applied Psychology, 60, 556-560.
- Borman, W. C. (1979). Format and training effects on rating accuracy and rater errors. Journal of Applied Psychology, 64, 410-421.
- Borman, W. C. (1980, April). Performance judgements: The quest for accuracy in ratings of performance effectiveness. Paper presented at the First Annual Scientist Practitioner Conference in Industrial-Organizational Psychology, Old Dominion University.
- Borman, W. C. (1981). Evaluating the validity of the Bennett Mechanical Comprehension Test for predicting performance in operating and maintenance jobs at Florida Power and Light. Minneapolis: Personnel Decisions Research Institute.
- Borman, W. C., Mendel, R. M., Lammlein, S. E., & Rosse, R. L. (1981). Developing and evaluating the validity of a test battery to predict performance in transmission and distribution jobs at Florida Power and Light. Minneapolis: Personnel Decisions Research Institute.
- Christal, R. E. (1974, January). The United States Air Force occupational research project (AFHRL-TR-73-75, AD-774 574.). Lackland AFB, TX: Occupational Research Division, Air Force Human Resources Laboratory.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. Psychological Bulletin, 52, 281-302.
- DeCotlis, T., & Petit, A. (1978). The performance appraisal process: A model and some testable propositions. Academy of Management Review, 3, 635-646.
- Dipboye, R., & dePontbriand, R. (1981). Correlates of employee reactions to performance appraisals and appraisal systems. Journal of Applied Psychology, 66, 248-251.
- Feldman, J. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. Journal of Applied Psychology, 66, 127-148.
- Kavanagh, M. J. (1982). Evaluating performance. In K. Rowland & G. Ferris (Eds.), Personnel Management: New perspectives. Boston: Allyn and Bacon.

- Kavanagh, M. J., & Hedge, J. W. (1983). A closer look at the correlates of performance appraisal system acceptability. Paper presented at the Annual Meeting of the Eastern Academy of Management, Pittsburgh.
- Kavanagh, M. J., MacKinney, A. C., & Wolins, L. (1971). Issues in managerial performance: Multitrait-Multimethod analyses of ratings. Psychological Bulletin, 75, 34-49.
- Landy, F., Barnes, J., & Murphy, K. (1978). Correlates of perceived fairness and accuracy of performance evaluation. Journal of Applied Psychology, 63, 751-754.
- Landy, F., & Farr, J. L. (1980). Performance rating. Psychological Bulletin, 87, 72-107.
- Latham, G. P., Wexley, K. N., & Purcell, E. D. (1975). Training managers to minimize rating errors in the observation of behavior. Journal of Applied Psychology, 60, 550-555.
- Nunnally, J. (1978). Psychometric theory (2nd ed.). New York: McGraw-Hill.
- Schneier, C. (1977). Multiple rater groups and performance appraisal. Public Personnel Management, 6, 13-20.
- Wherry, R. J., & Bartlett, C. J. (1982). The control of bias in rating: A theory of ratings. Personnel Psychology, 35, 521-551.
- Zedeck, S., & Cascio, W. (1982). Performance appraisal decisions as a function of rater training and purpose of appraisal. Journal of Applied Psychology, 67, 752-758.

END

FILMED

3 - 86

DTIC