

UNITED STATES AIR FORCE RESEARCH LABORATORY

APPLICATION OF A DISTRIBUTION-BASED ASSESSMENT OF MISSION READINESS SYSTEMS FOR THE EVALUATION OF TECHNICAL TRAINING

David J. Woehr
Michael J. Miller

Department of Psychology
Texas A&M University
College Station, TX 77843-4235

WINSTON BENNETT, JR.

HUMAN EFFECTIVENESS DIRECTORATE
MISSION CRITICAL SKILLS DIVISION
7909 Lindbergh Drive
Brooks AFB, TX 78235-5352

19991004 069

December 1998

Approved for public release; distribution is unlimited.

AIR FORCE MATERIEL COMMAND
AIR FORCE RESEARCH LABORATORY
HUMAN EFFECTIVENESS DIRECTORATE
7909 Lindbergh Drive
Brooks Air Force Base, TX 78235-5352

DTIC QUALITY INSPECTED 4

NOTICES

This report is published in the interest of scientific and technical information exchange and does not constitute approval or disapproval of its ideas or findings.

Using Government drawings, specifications, or other data included in this document for any purpose other than Government-related procurement does not in any way obligate the US Government. The fact that the Government formulated or supplied the drawings, specifications, or other data, does not license the holder or any other person or corporation, or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

The Office of Public Affairs 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 report has been reviewed and is approved for publication.

WINSTON BENNETT, JR
Project Scientist

R. BRUCE GOULD
Acting Chief
Mission Critical Skills Division

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

AGENCY USE ONLY (Leave blank)		2. REPORT DATE December 1998	3. REPORT TYPE AND DATES COVERED Interim Report – May-August 1995	
4. TITLE AND SUBTITLE Application of a Distribution-Based Assessment of Mission Readiness Systems for the Evaluation of Technical Training			5. FUNDING NUMBERS PE – 62202F PR – 1123 TA – A2 WU – 19	
6. AUTHOR(S) David J. Woehr Michael J. Miller Winston Bennett, Jr.				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Department of Psychology Texas A&M University College Station, TX 77843-4235			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Research Laboratory Human Effectiveness Directorate Mission Critical Skills Division 7909 Lindbergh Drive Brooks AFB, TX 78235-5352			10. SPONSORING/MONITORING AGENCY REPORT NUMBER AFRL-HE-BR-TR-1998-0044	
11. SUPPLEMENTARY NOTES Air Force Research Laboratory Technical Monitor: Dr. Winston Bennett (480) 988-6561; DSN 474-6561				
a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release: distribution unlimited.			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) The present study summarizes research focusing on ways to improve the usefulness of organizational level measures of unit readiness/effectiveness through the evaluation of numerous aircraft maintenance related measures of performance. In addition, a measurement approach using unit level outcome measures is presented, which adapts and extends the performance distribution assessment approach proposed by Kane (1986; 1992). It is demonstrated that, while originally used with subjective performance judgements, the system is readily adapted to regularly collected unit level outcomes. An important characteristic of the measurement system presented a focus on the range of performance observed, which considers the fluctuation or variability in performance as well as the level of performance. In addition, the system incorporates a relativistic scaling of performance information. That is, performance is expressed as a ratio of measured performance to some 'benchmark distribution'. This benchmark distribution may represent established standards, expected, or previously attained levels of performance. This scaling process serves to express actual performance in terms relative to some previously established range of performance. The representation of performance in distributional form along with relativistic scaling has several important advantages to traditional measurement approaches. It allows for an assessment of the consistency of performance and the extent to which negatively valued outcomes are avoided, which serves to facilitate the comparison and combination of data across diverse performance measures. The current study presents a demonstration of the proposed measurement system with aircraft maintenance data. Preliminary results indicate that this approach does in fact have the potential to improve the utility of organization level criterion measures.				
14. SUBJECT TERMS Training Evaluation Distributional Assessment			15. NUMBER OF PAGES 39	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL	

Table of Contents

NOTICES.....iv

PREFACE.....v

SUMMARY.....vi

INTRODUCTION.....1

IDENTIFICATION OF AIRCRAFT MAINTENANCE RELATED MEASURES OF
PERFORMANCE.....3

A DISTRIBUTIONAL APPROACH TO CRITERION MEASUREMENT.....6

USING DEAMR FOR TRAINING EVALUATION..... 11

REFERENCES..... 13

APPENDIX A: List of Maintenance Related Outcome Measures Identified..... 14

APPENDIX B: Sample Survey Measure.....27

PREFACE

The views and opinions expressed in this paper are those of the authors and do not reflect the official policies or opinions of their respective organizations.

The authors wish to thank the United States Air Force personnel that participated in this project, for these scientific advances would not be possible without their support and cooperation. The authors would also like to thank Kathleen Sheehan for preparing the final format of this report.

SUMMARY

The present study summarizes research focusing on ways to improve the usefulness of organization level outcome measures of unit readiness/effectiveness through the evaluation of numerous aircraft maintenance related measures of performance. In addition, a measurement approach using unit level outcome measures is presented, which adapts and extends the performance distribution assessment approach proposed by Kane (1986; 1992). It is demonstrated that, while originally used with subjective performance judgments, the system is readily adapted to regularly collected unit level outcomes.

An important characteristic of the measurement system presented is a focus on the range of performance observed, which considers the fluctuation or variability in performance as well as the level of performance. In addition, the system incorporates a relativistic scaling of performance information. That is, performance is expressed as a ratio of measured performance to some 'benchmark distribution'. This benchmark distribution may represent established standards, expected, or previously attained levels of performance. This scaling process serves to express actual performance in terms relative to some previously established range of performance.

The representation of performance in distributional form along with relativistic scaling has several important advantages to traditional measurement approaches. It allows for an assessment of the consistency of performance and the extent to which negatively valued outcomes are avoided, which serves to facilitate the comparison and combination of data across diverse performance measures.

The current study presents a demonstration of the proposed measurement system with aircraft maintenance data. Preliminary results indicate that this approach does in fact have the potential to improve the utility of organization level criterion measures.

APPLICATION OF A DISTRIBUTION-BASED ASSESSMENT OF MISSION READINESS SYSTEMS FOR THE EVALUATION OF TECHNICAL TRAINING

INTRODUCTION

Background

A vital concern for the Air Force is the maintenance of mission capability and readiness. A crucial mechanism for the maintenance of mission readiness is personnel training. There is little, if any, dispute that effective personnel training serves to enhance the effectiveness and capability of the Air Force in general. This fact is reflected in the overwhelming scope of training conducted throughout the Air Force and the tremendous amount of time and resources committed to the training endeavor.

Of tremendous importance to the design, implementation, and revision of training throughout the Air Force, as with any organization, is the ability to evaluate the effectiveness of training interventions. Goldstein (1991) defines training evaluation as: "the systematic collection of descriptive and judgmental information necessary to make effective training decisions related to selection, adoption, value, and modification of various instructional activities." (p. 557) More specifically, the effective evaluation of any training intervention is crucial to informed decision making regarding the intervention. Central to effective training evaluation is the standard or criteria against which the training is evaluated. In addition, the comprehensive evaluation of training interventions mandates the use of multiple criterion measures. The impact of training interventions must be assessed at different levels (e.g., person, work group, organization). Unfortunately, organization level outcome measures are often dismissed as criterion measures due to contamination by extraneous aspects of the work environment. Despite this limitation, the use of these measures is extremely important for demonstrating the utility of training interventions.

Organization Level Criterion Measures

Organization level outcome measures represent global indices of effectiveness. While many commonly used criterion measures focus on the assessment of individual effectiveness, organization level measures often provide more aggregate measures of effectiveness. They typically include results-oriented measures such as quality control indices, productivity or maintenance indices, promotion rate, salary progression or level, and turnover rates. The value of these measures as criterion measures is somewhat controversial. Two schools of thought can be found in the literature with respect to ways of conceptualizing the criterion construct. One school of thought emphasizes a conceptualization of performance as reflected in overt individual behaviors (e.g., Campbell, et al. 1970; Borman, 1983). This view focuses on the identification of behavioral regularities important to organizational functioning. The other school of

thought focuses on outcomes. This view emphasizes the importance of outcomes and results to organizational functioning. Recent theories of the criterion construct, however, have begun to recognize the inextricable relationship between job behaviors and outcomes. Along these lines Binning and Barrett (1989) argue: "... optimal description of the performance domain for a given job requires careful and complete delineation of valued outcomes and the accompanying requisite behaviors" (p. 486).

Problems with Outcome-Based Criterion Measures

The detailed conceptual delineation of the relationship between job performance and outcomes is especially relevant to training evaluation. An important direction for future research is a focus on behavior/outcome linkages and generating empirical support for these linkages. Unfortunately, the operationalization of specific outcome measures generates somewhat of a dilemma for training evaluation. On the one hand, the ultimate value of training lies in its ability to impact outcomes of value to the organization. Outcome measures (e.g., productivity levels, turnover rates, error rates, etc.) at both individual and aggregate levels would appear to be the ultimate criterion of interest for evaluating training interventions. On the other hand, these measures suffer from a number of problems that limit their usefulness as a standard against which to judge the impact of training.

First and foremost among these problems is the fact that these measures are typically contaminated to an undetermined extent by sources of variance over which the individual has no control. Specifically, the measured outcome is to some extent determined by factors other than individual performance. A second problem with outcome measures is that they are not based on a common metric. Outcome measures are often unique to particular units within an organization and thus are difficult to interpret and compare across organizational work groups or divisions. Additionally, the lack of a common metric typically precludes the meaningful aggregation of performance information across organizational units. A third problem is that these measures only provide an indication of outcome as opposed to the process underlying the outcome. Thus these measures provide little, if any, information about the nature of performance. Finally, the traditional use of outcome measures offers little, if any, means of assessing measurement quality (i.e., how good are the measurements obtained with these measures).

Another major limiting factor with respect to the use of organizational level outcome measures is the lack of conceptual and/or empirical formulations specifying the potential linkages between personnel action and specific outcome measures. For example, if the goal is to evaluate the impact of a particular training program with respect to organizational outcomes, it is important to match the nature and content of the training with specific outcome measures likely to be influenced. If the training program focuses on improving maintenance skills then measures most directly related to maintenance

outcomes should be identified and examined. While there may be numerous outcome measures available, little if any information exists pertaining to the performance relevance of these measures.

Thus, although regularly collected and typically readily available, as a criterion against which to judge the impact of various training interventions in organizations, outcome measures have not proven as useful as criteria which are defined in terms of individual behavior. Despite this, however, the use of these measures is extremely important for demonstrating the ultimate utility of training interventions. Consequently, an important goal with respect to training evaluation is the development of ways to improve the utility of organization level criterion measures.

In summary, while organizational level outcome measures are a potentially valuable criterion against which to evaluate training effectiveness, several factors have limited the utility of these measures. These factors include: a) contamination by non-performance related factors; b) lack of a common measurement metric; c) a focus on overall level rather than the performance process; d) lack of any indication of measurement quality; and, e) no conceptual/empirical formulations of the linkage between specific actions and outcomes. Thus, any system that uses outcome measures must address these issues.

IDENTIFICATION OF AIRCRAFT MAINTENANCE RELATED MEASURES OF PERFORMANCE

One of the primary objectives of the present study was to identify and examine the utility of aircraft maintenance related measures of performance typically collected and used by the Air Force. Measures of performance (MOPs) are qualitative or quantitative measures of system capabilities or characteristics (USAF/TEP, 1994) Toward this end, several sources of data were identified through interviews with supervisory level maintenance personnel. One source of such data was a combination of CAMS- based maintenance data and unit mission characteristic data. This data is routinely collected and reported by aircraft maintenance units as an index of mission effectiveness. This data takes into consideration both equipment and unit mission and manpower characteristics. Example measures include fully mission capable rate, man hours per flying hours, air and ground abort rates, etc. Table 1 presents specific examples of these measures. A complete listing of the actual measures identified is presented in Appendix A.

One question with respect to these measures is the degree to which these measures are routinely collected and reported. Quality assurance and summary aircraft performance reports were examined for three different fighter wings. Appendix B indicates which of the MOP's presented in Appendix A are currently reported in the summary reports. Appendix B indicates considerable overlap across the three wings. It

is also likely that the measures not currently reported are collected and available for analysis.

TABLE 1.
Sample Maintenance Related Measures of Performance

MEASURE	DEFINITION	FORMULA
Awaiting Maintenance Rate	AWM is a deferred discrepancy that is a repair that cannot be accomplished within 5 days of the original write-up.	$\frac{\# \text{ of AWM}}{\# \text{ of poss acft}} \times 100$
Chargeable Deviations	Number of inspection discrepancies	based on actual count
Fix Rate	# of aircraft that return with inoperable systems & must be returned to MC status within a specified amount of time	$\frac{\# \text{ 4/8/12 hour fixes}}{\text{total \# of code 3 breaks}}$
Fully Mission Capable Rate	% of aircraft possessed hrs that were fully mission capable for a unit over a specified period of time	$\frac{\text{FMC}}{\text{avg. possessed hours}} \times 100$
Man Hour Per Fly Hour	all flying hour categories totaled	$\frac{\text{man-hours}}{\text{total flying hours}}$
Repeat Rate	Repeat = the same system malfunctioning on the next flight.	$\frac{\# \text{ repeats}}{\text{local sorties flown}} \times 100$

Another concern with respect to these measures is that any information derived may be severely limited in that the data might reflect expected levels and/or standards rather than actual performance. If this were the case very little variability in these measures would be expected across units and time frames and their usefulness as criterion measures would be minimal. However, examination of actual data pertaining to these measures indicates that there is in fact sufficient variability to warrant further investigation into the differences across units. Table 2 provides a small sample of a much larger set of actual data collected in the present study for the measures presented in Table 1. These data represent summary information for one fighter wing across 2 fiscal years.

TABLE 2.
Data Sample for Maintenance Related Operational Measures

MEASURE	FY '94				FY '95			
	MEAN	SD	MIN	MAX	MEAN	SD	MIN	MAX
Awaiting Maintenance Rate	3.57	5.80	1.08	42.85				
Chargeable Deviations	19.70	9.20	3.00	48.00	18.70	18.70	1.00	67.00
Fix Rate	88.35	6.60	73.07	100.00	91.24	6.50	75.00	100.00
Fully Mission Capable Rate	88.46	3.9	78.02	96.87	84.61	6.90	63.32	96.11
Man Hour Per Fly Hour	5.82	3.8	0.30	28.70	5.97	2.40	0.70	21.50

Although a valuable source of information with respect to mission capability, these measures illustrate many of the disadvantages associated with operational measures. For example, the metric of each measure is unique to the characteristic being measured. Thus the data are difficult to combine and summarize across measures. Further, the measures are cumbersome to summarize. That is, while the measures lend themselves to typical overall summations such as mean performance level, such measures of central tendency only provide part of the overall picture. Other important information including the amount of fluctuation and the percent of time at or above some preset standard is typically not presented in any summary metric.

Despite these limitations, the indices presented in Appendix A have many desirable characteristics with respect to training evaluation. These characteristics include:

1. The measures are regularly and systematically collected.
2. It appears that these indices are both required by and reported to Major AF Commands. Thus it is likely that these measures are available Air Force wide.
3. The mission capable/readiness indices reflect both equipment, mission, and manpower characteristics.
4. The indices are easily aggregated from the individual unit level to higher levels of the organization (wing, command, etc.).
5. The indices reflect multiple measures of performance within a specified time span (iterated job function) and thus are readily amenable to the distribution-based measurement system (presented below).

In summary, numerous maintenance related measures of performance were identified. These measures represent organizational level outcomes that provide an indication of system performance. Further, these measures are routinely collected and reported and thus are a potentially valuable source of information for the evaluation maintenance related training programs.

A DISTRIBUTIONAL APPROACH TO CRITERION MEASUREMENT

A second objective of the present study was to develop and evaluate a measurement system that increases the utility of regularly collected operational measures of performance. Toward this end a specific measurement system is presented. The measurement approach presented here extends the system for assessing individual performance developed by Kane (1986) to outcome level criteria measurement. It is believed that this approach may offer a partial solution to the problems associated with outcome measures. The original system presented by Kane (1986), labeled Performance Distribution Assessment (PDA), is based on the distributional measurement model postulated by Kane and Lawler (1979). An important characteristic of this model is a focus on the range of performance observed. Specifically, the model stipulates that not only is the level of performance important, but the fluctuation or variance in performance must also be considered. For example, two individuals may both be appropriately characterized as "average performers"; however, if one is consistently average and the other alternates between very poor and very good, very different pictures emerge with respect to the individuals' performance. Thus performance measurement must assess the range of performance over time. Specifically, performance is defined in terms of the outcomes of job functions that are carried out on multiple occasions within a specified time span (i.e., iterated job functions). It is expected that, due to varying levels of individual ability and motivation as well as varying levels of external constraints, these

outcomes will reflect different levels of effectiveness. Performance can subsequently be represented in terms of the frequency at which various outcome levels occurred within a given time span.

Another important characteristic of the PDA approach is that it incorporates a relativistic scaling of performance information. More specifically, performance is expressed as a ratio of actual performance (as reflected in the performance distribution generated) to a maximum feasible performance distribution. This maximum feasible distribution reflects the highest level of performance attainable given the constraints under which the work occurs. This scaling process serves to express performance in terms of a relative range of potential performance. Thus, the method allows for quantifiably excluding from consideration in the evaluation of performance the range of performance that is attributable to circumstances beyond the performer's control.

The representation of performance in distributional form along with relativistic scaling has several important advantages. First, it allows for a consideration of performance variability as well as average levels of performance. Thus it allows for an assessment of the consistency of performance and the extent to which negatively valued outcomes are avoided. In this way more information is provided regarding the idiosyncratic nature of performance. Second, the relativistic scaling process advocated by the PDA process produces measures of the effectiveness of performance on relativized 0-100% scales with common zero and common upper limits of 100%. Thus any given percentage level remains constant in its meaning regardless of the job, division, or even the organizational level in which it occurs. At the same time, the particular outcome measures used to assess performance may be individualized to meet situational demands and organizational constraints. Specifically, if positions have appreciably different content and extraneous-constraint conditions, measures can be scaled to account for these differences.

The PDA approach was originally advocated as method for enhancing individual performance ratings. Specifically, it was formulated to incorporate subjective estimates of individual performance outcome frequencies (i.e., supervisory ratings of the frequency at which individuals performed at a particular level). However, its focus on the frequency of particular performance outcomes make it particularly amenable to use with more objective outcome measures. Thus, the application of this methodology to the measurement of organizational outcomes using iterative operational measures appears to be a fruitful avenue for research and may serve to increase the utility of these measures in the training evaluation process.

TABLE 3.

Sample Performance Level Frequencies and Distributional Characteristics for the Man Hour per Fly Hour Measure

Performance Frequencies						Distribution Characteristics		
Perf. Level	Perf. Range	Utility Weights	Perf. Level Freq.	Perf. Level %	Comparison Level %		Utility Wt. Scale	Perf. Level Scale
1	28.70	-100	0	0	0	Mean =	38.46	3.77
2	21.60	-50	1	8	1	SD =	34.83	0.70
3	14.50	0	2	15	10	Skewness=	-1.02	-1.02
4	7.40	50	9	69	80	Kurtosis=	1.19	1.19
5	.30	100	1	8	9	Negative Range Score =	-3.85	
		Total Obs. =	13			Total Perf. Effectiveness	85.09	

Adaptation of the PDA Approach for Outcome Level Measures

As noted above, the PDA system appears to be well suited for the measurement and scaling of operational criterion measures. For purposes of illustration, Table 3 presents hypothetical evaluation data, presented in PDA format, for the man hour per fly hour MOP. In Table 3 the performance range represents 5 equidistant steps between the highest (best) possible performance outcome (listed as .30 in the Table) and the lowest (worst) performance level (listed as 28.70 in the Table) for the man hour per fly hour measure. These levels represent the lowest and highest (respectively) number of man hours required per fly hour for the wing across the 2 fiscal years. The performance level frequencies are based on an actual count of man hour per fly hour outcomes each month over the course of 1 fiscal year. The utility weights represent the utility or value to the organization of performance at each of the 5 levels. In the present example, these are hypothetical values. In actuality these weights would be based on SME estimates. The comparison level values represent a "benchmark" distribution. This "benchmark" distribution may represent either an estimated ideal distribution of performance or the actual performance distribution of a comparison unit (i.e., an earlier time frame or another work unit).

FIGURE 1.
Graphical representation of the actual and comparison performance distributions.

Performance Outcome Distribution

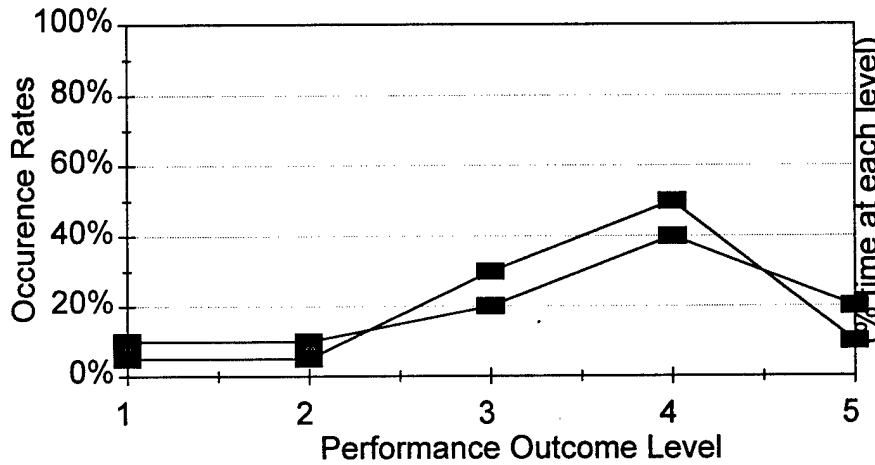


Figure 1 shows the relationship between the two performance distributions represented by the actual performance values and the comparison level. Based on this information distributional characteristics for the actual performance values are presented (the “Distributional Characteristics” in Table 3). These characteristics may be expressed in either the utility weight metric or the performance level scale.

An important component of the distributional assessment system presented here is that it provides an overall index of performance that takes into account the level, variability, and utility of performance over time. The total performance effectiveness score (TPE) represents a quantitative index, in a percentage metric, of the proximity of the actual distribution to the comparison distribution. The TPE index is calculated as:

$$TPE = \frac{\left[\sum_{i=N}^2 [A_i - (C_i + I_i)]^2 W_i \right]^{\frac{1}{2}}}{\left[\sum_{i=N}^2 [C_i + I_i]^2 W_i \right]^{\frac{1}{2}}}$$

where:

- N = the number of steps or levels in the performance continuum with the Nth level representing the highest level of performance
- A_i = the actual occurrence rate observed for the i th level of the performance continuum
- I_i = the occurrence rate for the i th level of the comparison performance distribution
- C_i = the difference between the sums of the actual and comparison occurrence rates for all levels above the
- W_i = the utility weight for the i th level of the performance continuum.

The TPE index specifies one minus the ratio of the observed distance between the actual and comparison distributions and the maximum distance possible. It encompasses all variation present in the actual distribution of performance across all levels of performance. Thus it represents a suitable summary measure of performance. Scores on the index range from 0 to 100 and are comparable across measures and units.

This revised approach to the performance distribution model is labeled here as Distribution-Based Evaluation and Assessment of Mission Readiness (DEAMR). This approach extends the beneficial characteristics of relative distribution based performance assessment to organization level outcome measures. More specifically, characteristics of the DEAMR process include:

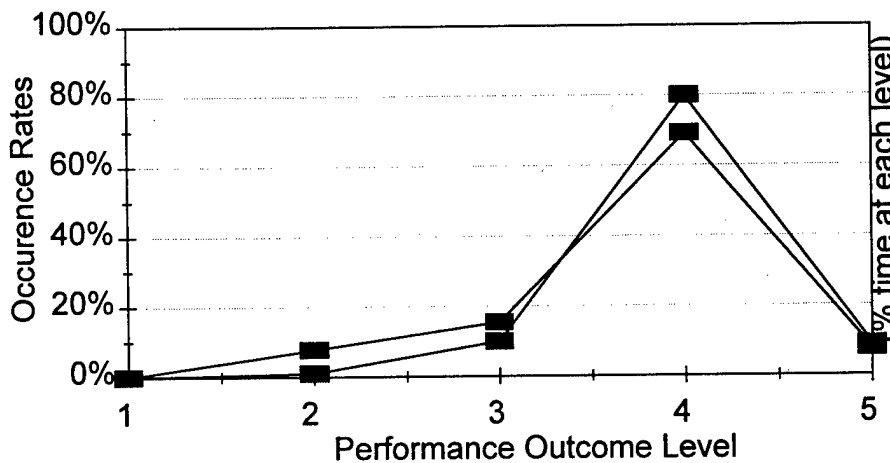
1. Performance measurement is relativistic. Outcome measures are scaled relative to maximum possible and minimum acceptable performance levels. Performance distributions are relative to some "benchmark" distribution. Thus, measurement considers the extraneous factors that may influence outcome measures.
2. Performance measurement is based on common metric. All measures are expressed in terms of percentages and thus have minimum and maximum points.
3. Multiple measures of performance are provided; performance is described in terms of mean level, consistency, and negative range avoidance. These multiple measures provide more information about the nature of performance and performance problems.

Another important characteristic of the DEAMR system is that it is easily automated. Relatively little data is required in order to calculate the distributional parameters. These data required include the highest possible and lowest acceptable performance level, an estimate of the utility weight associated with the lowest acceptable performance level, and the actual frequency of performance outcomes at each of the

performance levels. Figure 2 represents the output of a spreadsheet based program specifically designed to provide distributional performance information. The highlighted boxes indicate where data must be input into the program. Performance distribution information is then automatically calculated and displayed both numerically and graphically.

FIGURE 2.
Spreadsheet-based program designed to input MOP data and calculate distributional characteristics

Performance Outcome Distribution



USING DEAMR FOR TRAINING EVALUATION

The mission capability/readiness indices discussed above represent viable potential criterion measures for the evaluation of training effectiveness at the outcome level. Further these data meet the requirements for use with the DEAMR system. Thus it is possible to rescale the data in distributional form. The DEAMR format could then be used to evaluate the effectiveness of specific training interventions.

While the measures identified generally provide a potentially valuable source of information with respect to training evaluation, it is important to consider the further

refinement of the data. More specifically, it would be beneficial to establish a pool of potential indices most relevant to specific training interventions to be evaluated (e.g., maintenance indices for maintenance technician training). Here it is important to identify and evaluate key measures from the larger pool of potential measures. The focus of this measure evaluation would be to identify indices that are: a) important to unit effectiveness, b) frequently and reliably collected, c) sensitive to individual performance, d) relatively insensitive to system variables, and e) relevant to training intervention. Information about outcome indices may be obtained through either SME workshops or through structured questionnaires. SME's would be used to provide information about each potential measure (e.g., the relative importance of each measure, sensitivity to individual performance) as well as information relevant to the DEAMR process (e.g., item utility weights, optimal possible and minimal acceptable levels, etc.). The listing of measures presented in Appendix B may be modified to provide a preliminary instrument that might be refined and used to identify those measures that are most likely to be affected by better trained personnel. It is only through such systematic examination of the measures available that detailed conceptualizations of the linkages between individual performance and organizational outcome can be established. Ultimately, the effective use of outcome measures for the evaluation of personnel training depends on the delineation of these linkages.

REFERENCES

- Binning, J.F., & Barrett, G.V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. Journal of Applied Psychology, 74, 478-494.
- Borman, W.C. (1983). Implications of personality theory and research for the rating of work performance in organizations. In F. Landy, S. Zedeck, & J. Cleveland (Eds.) Performance measurement and theory (pp. 127-172). Hillsdale, NJ: Erlbaum.
- Campbell, J.P., Dunnette, M.D., Lawler, E.E., & Weick, K.E. (1970). Managerial Behavior, Performance, and Effectiveness. New York: McGraw-Hill.
- Goldstein, I.L. (1991). Training in work organizations. In M.D. Dunnette & L.M. Hough (Eds.) Handbook of Industrial and Organizational Psychology, vol. 2, (pp. 507-620). Palo Alto: Consulting Psychologists Press.
- Kane, J.S. (1986). Performance distribution assessment. In R. Berk (Ed.), Performance Assessment: Methods and Applications, (pp. 237-273). Baltimore, MD: Johns Hopkins University Press.
- Kane, J.S., & Kane, K.F. (1992). The analytic framework: The most promising approach for the advancement of performance appraisal. Human Resource Management Review, 2(1), 37-70.
- Kane, J.S., & Lawler, E.E. (1979). Performance appraisal effectiveness: Its assessment and determinants. In B. Staw (Ed.) Research in Organizational Behavior, Vol 1. Greenwich, CT.: JAI Press.
- USAF (USAF/TEP). (1994). AFI 99-103, Test and evaluation Process, 25 July 94. Washington, DC.: Author.

Appendix A:

List of Maintenance Related Outcome Measures Identified

MEASURE	DEFINITION	FORMULA
Abort Rate	% of scheduled sorties which must be canceled due to system malfunction	$\frac{\# \text{ of air sorties}}{\# \text{ of departures or sorties}} \times 100$
Air Aborts	Number of air aborts.	
Air Abort Rate		$\frac{\text{air aborts}}{(\text{LCL sorties flown} + \text{ground abort rates})} \times 100$
Actual Fly Hours		
Actual UTR Rate	Average number of sorties. Deigned to measure how well maintenance community is supporting contracted flying commitment	$\frac{\text{total sorties flown}}{\text{APA aircraft}}$
Adjusted Scheduling		
Adjusted Sortie Schedule		local sorties scheduled + weather adds + ferry/FCF adds + other adds - weather deletes - sympathy deletes - other deletes
Aircraft Battle Damage Repair Time	self explanatory	sortie generation rate
Aircraft Rearmed Time		sustain forces & operations
Aircraft Refueling	unit's ability to provide air refueling services to users	
Aircraft Regeneration Timing	# of aircraft regeneration within X amount of time	
Aircraft Scheduling Effectiveness Rate	Deals with the flying schedule and deviations to it.	$\frac{\text{adjusted sortie sched.} - \text{chargeable deviation}}{\text{adj. sortie sched.}} \times 100$

MEASURE	DEFINITION	FORMULA
Authorized Aircraft	The number of aircraft for the wing as authorized by MAJCOM.	
Availability	the probability that a system is operable & ready to perform its intended mission at any given time	
Aircraft Possessed Hours	total # of aircraft availability over the past 12 months	
Average Possessed Aircraft	average # of aircraft availability, to include depot NMC time for aircraft possessed by depot above and beyond back-up aircraft inventory (BAI), over the past 12 months	$\frac{\text{total Acft possessed hrs}}{\text{total days in month} \times 24}$
Average Utilization Per Aircraft Per Month	average life units that pass per system during a month	
Awaiting Maintenance		
Awaiting Maintenance Rate		$\frac{\# \text{ of AWM}}{\# \text{ of poss acft}} \times 100$
Awaiting Parts		
Awaiting Parts Rate		$\frac{\# \text{ of AWP}}{\# \text{ of poss acft}} \times 100$

MEASURE	DEFINITION	FORMULA
Break Rate	the % of sorties from which an aircraft returns with an inoperable mission-essential system that was previously operable. System malfunction occurring in-flight that renders aircraft not mission capable after landing	$\frac{\# \text{ of aircraft breaks}}{\# \text{ sorties flown}} \times 100$
Cancellation Rate	% of all scheduled sorties or departures that were canceled	$\frac{\# \text{ of cancellation}}{\text{scheduled departures \& sorties}} \times 100$
Cannibalization Rate	maintenance efforts to compensate for supply problems or for maintenance convenience to launch aircraft on time	
Cannibalizations Per Average Possessed Aircraft	avg # of cannibalization per avg possessed aircraft	$\frac{\text{total \# of cannibalizations}}{\text{avg possessed aircraft}}$
Cannibalizations (removals only) Per Departure Per Sortie	avg # of cannibalization removals per departure or sortie	$\frac{\text{total \# of Cannibalizations}}{\# \text{ of departures \& sorties}}$
Cannibalization Rate		$\frac{\# \text{ of Canns}}{\text{total sorties flwn}} \times 100$
Cannot Duplicate Rate		
Chargeable Deviations		
Code 3 Breaks		
Code 3 Break Rate		$\frac{\# \text{ code 3 breaks}}{\text{total sorties flown}} \times 100$

MEASURE	DEFINITION	FORMULA
Combat Rate	average # of consecutively scheduled missions flown before aircraft experience critical failures	$\frac{\# \text{ of successful sorties flown}}{(\# \text{ of scheduled missions} - \# \text{ of ground aborts} - \# \text{ of air aborts})}$
Deferred Discrepancy Rate (repair which can't be done within 5 days)	A repair which cannot be accomplished within five days of the original write-up.	$\frac{\# \text{ of AWM/AWP}}{\# \text{ of possessed aircraft}} \times 100$
Delay Discrepancies	total # of non-grounding discrepancies that have been delayed or deferred & will not be worked on within 24 hrs from time discrepancy was found	
Delayed Discrepancy Average	avg number of delayed discrepancies per possessed aircraft	$\frac{\text{total delayed discrepancies}}{\text{adjusted avg possessed aircraft}}$
Delay Discrepancy Average, Awaiting Maintenance	avg # of delayed discrepancies per aircraft awaiting maintenance	$\frac{\text{total discrepancies delayed for maintenance}}{\text{adjusted avg possessed aircraft}}$
Delayed Discrepancy Average, Awaiting Parts	avg # of delayed discrepancies per aircraft awaiting parts	$\frac{\text{total discrepancies delayed for parts}}{\text{adjusted avg possessed aircraft}}$
Deployability	whether the system can be efficiently deployed to the theater of operations within the constraints of the user defined requirements	
Dropped Object Rate	rate of dropped object per 100 sorties. Dropped objects may be a manifestation of material, personnel, or design deficiencies	$\frac{\# \text{ of aircraft breaks during measured period}}{\# \text{ of sorties flown during period}} \times 100$

MEASURE	DEFINITION	FORMULA
Engine Foreign Object Damage Rate	rate of engine FOD's per 1,000 departures	$\frac{\# \text{ of FOD incidents}}{(\# \text{ of departures \& sorties}) \times (\# \text{ of engines on aircraft}) \times 1,000}$
Essential System Repair Time Per Flight Hour	avg clock time needed to repair mission-essential equipment per operational flight hour	elapsed corrective maintenance + elapsed preventive maintenance flight hours
Fault Detection Rate		
Fault Isolation Rate		
Fleet Availability	a total # of aircraft availability, to include depot NMC time for aircraft possessed by depot above and beyond back-up aircraft inventory, over the past 12 months	
Fix Rate	# of aircraft that return with inoperable systems & must be returned to MC status within a specified amount of time	$\frac{\# \text{ 4/8/12 hour fixes}}{\text{total \# of code 3 breaks}}$
Fully Mission Capable Hours		
Fully Mission Capable Rate	% of aircraft possessed hrs that were fully mission capable for a unit over a specified period of time	$\frac{\text{FMC}}{\text{avg. possessed hours}} \times 100$
Ground aborts	Number of ground aborts.	
Ground Abort Rate	% of sorties or departures that aborted of the total attempted departures or sorties	$\frac{\text{ground aborts}}{\text{LCL sorties flown} + \text{ground aborts}} \times 100$
Hangar Queen Days		

MEASURE	DEFINITION	FORMULA
Hangar Queen Rate		
Hours Flown		
Local Sorties Flown		
Local Sorties Scheduled		
In Flight Emergencies		
In Flights Emergency Rate		
Maintainability	ability of an item to be retained in, or restored to, a specified condition within a given time period when maintenance is performed by personnel having specified skills using prescribed procedures & resources at each prescribed level of maintenance & repair	
Maintenance Completes		
Maintenance Delivery Reliability	% of times aircraft is mission capable at scheduled or actual crew show time & aircraft is capable of flight & will be accepted by aircrew	$\frac{(\text{total departures or sorties}) - (\# \text{ of aircraft broke at scheduled or actual crew show time}) \times 100}{\text{total departures or sorties}}$
Maintenance Man Hour Per Fly Hour - Corrective	for inherent malfunctions, induced malfunctions, no-defect actions, or total events	
Maintenance Man Hour Per Fly Hour - Improvement	product improvement	

MEASURE	DEFINITION	FORMULA
Maintenance Man Hour Per Fly Hour - Preventive	preventive maintenance	
Maintenance Man Hour Per Fly Hour - Support	direct maintenance man hours required to support a system	
Maintenance Man Hour Per Life Unit	MAJCOMs estimate maintenance man hours per flying hour on their specific needs	
Maintenance Personnel Per Operational Unit	total # of direct maintenance personnel needed for each specified operational unit to perform direct on-equipment maintenance	
Maintenance Plan Rate		$\frac{\text{total pts earned}}{\text{total pts scheduled}} \times 100$
Maintenance Starts		
Maintenance Turn Time	time required to prepare a returning mission-capable aircraft for another sortie	
Man Hour Per Fly Hour	all flying hour categories totaled	$\frac{\text{man-hours}}{\text{total flying hours}}$
Man Hour Per Sortie		
Max Schedule Points Earned		
Max Schedule Points Possible		
MDC Man Hours		
Mean Down Time	avg elapsed time between losing mission capable status & restoring the system to MC status	sortie generation rate

MEASURE	DEFINITION	FORMULA
Mean Repair Time	avg corrective maintenance time required to return a system or part to operational status	
Mean Time Between Critical Failure	avg time between failure of mission-essential system functions	$\frac{\# \text{ of operating hours}}{\# \text{ or critical failures}}$
Mean Time Between Maintenance Actions	avg flying hours between maintenance events, including scheduled & unscheduled events	sortie generation rate
Million Ton Miles Per Day	aggregate, unconstrained measure of airlift capacity used as a top-level comparative metric	$\frac{(\text{objective utilization rate}) \times (\text{blockspeed}) \times (\text{payload}) \times \text{productivity factor}}{1,000,000 \text{ nautical miles}}$
Mission Capable Hours		
Mission Capable Rate	% of aircraft possessed hours that were fully and partially mission capable for a unit over a specified period	$\frac{\text{PMCM} + \text{PMCS} + \text{PMCB} + \text{FMC}}{\text{APH}}$
Non Mission Capable Both Hours		
Non Mission Capable Hours		
Non Mission Capable Both Rate		$\frac{\text{NMCB}}{\text{avg possessed hours}} \times 100$
Non Mission Capable Rate		$\frac{\text{NMCS} + \text{NMCM} + \text{NMCB}}{\text{avg possessed hrs}} \times 100$
Not Operationally Ready -Maintenance	% to total systems not operationally available due to unperformed required maintenance	

MEASURE	DEFINITION	FORMULA
Number of Aircraft Necessary to Perform Mission		
Number of Aircraft Successfully Employed	scheduled aircraft arriving at employment base	
Object Utilization Rate	avg # of hours per day the primary aircraft inventory fly, & is measured over two periods: "surge" & "sustained"	surge = the first 45 days of a contingency sustained = time after first 45 days
O&M Days		
O&M Days Not Flown		
Partially Mission Capable-Both Hours		
Partially Mission Capable Hours		
Partially Mission Capable Rate	can perform at least one but not all of its assigned missions	
Partially Mission Capable-Both Rate		partially mission capable <u>both</u> _____ x 100 avg possessed hours
Possessed Availability Rate	a % of aircraft availability over the past 12 months	
Pproductivity factor	a factor to account for the aircraft returning empty from the theater & positioning legs to onload locations. The productivity factor is constant at 47%	

MEASURE	DEFINITION	FORMULA
Payload	based on avg payload observed in the Mobility Readiness Study modeling process using a critical leg distance of 3,200 NM	
Possessed Availability Rate	a % of aircraft availability over the past 12 months	
Pproductivity factor	a factor to account for the aircraft returning empty from the theater & positioning legs to onload locations. The productivity factor is constant at 47%	
Program Hours		
Program Fly hours		
Program Hour UTE Rate		
Program Sorties		
Program UTE rate		
Recurs	The same sytem malfunctioning within 3 flights of the original writeup.	
Recur Rate		$\frac{\# \text{ recurs}}{\text{local sorties flown}} \times 100$
Refueling Time		sortie generation rate
Regeneration After Deployment	the deployed unit's ability to attain a combat ready posture for the in-theater commander as soon as possible after arriving at a deployment base	

MEASURE	DEFINITION	FORMULA
Reliability	probability that an available system/mission will perform its required function at a specified mission time, in a specified environment, or during a scenario over the duration of a specified mission or over a specified # of sorties	$\frac{\text{\# of missions completed}}{\text{\# of missions attempted}}$
Repair Turn-Around Time	measured from time item is removed from aircraft until it is repaired & ready for reissue	Air Force standard for repair turn around times for avionics items is 8 days
Repeats	The same system malfunctioning on the next flight.	
Repeat Rate		$\frac{\text{\# repeats}}{\text{local sorties flown}} \times 100$
Retest Okay Rate		
Sorties Flown		
Sorties Scheduled		
Sustainability	system's ability to maintain the necessary level & duration of operations to achieve military objectives. Often measured in # of days	
Total Abort Rate		ground abort rate + air abort rate
Total Aircraft Possession Hours		

MEASURE	DEFINITION	FORMULA
Total Non Mission Capable Maintenance Hours		
Total Non Mission Capable Supply Hours		
Total Non Mission Capable Maintenance Rate		$\frac{\text{NMCM} + \text{NMCB}}{\text{avg possessed hrs}} \times 100$
Total Non Mission Capable Supply Rate		$\frac{\text{NMCS} + \text{NMCSB}}{\text{avg possessed hrs}} \times 100$
Total Partially Mission Capable Maintenance Hours		
Total Partially Mission Capable Hours		
Total Partially Mission Capable Maintenance Rate		$\frac{\text{PMCB} + \text{PMCM}}{\text{PAH}} \times 100$
Total Partially Mission Capable Supply Rate		$\frac{\text{PMCB} + \text{PMCM}}{\text{avg possessed hrs}} \times 100$
Utilization Rate	avg life units that pass per system during a specific period, expressed in flight hours or sorties per aircraft per relevant period of time, such as a day or month	$\frac{\text{avg flight hours}}{\# \text{ of possessed aircraft or authorized aircraft}}$

Appendix B:
Sample Survey Measure

MEASURE	56th F Wing	31st F Wing	52nd F Wing
Abort Rate		X	X
Air Aborts	X	X	
Air Abort Rate	X		
Actual Fly Hours	X		
Actual UTR Rate	X		
Adjusted Scheduling	XX		
Adjusted Sortie Schedule			
Aircraft Battle Damage Repair Time			
Aircraft Rearmed Time			
Aircraft Refueling			
Aircraft Regeneration Timing			
Aircraft Scheduling Effectiveness Rate	XX	X	X
Authorized Aircraft	X		
Availability			
Aircraft Possessed Hours			
Average Possessed Aircraft	XX		
Average Utilization Per Aircraft Per Month			
Awaiting Maintenance	X		
Awaiting Maintenance Rate		X	
Awaiting Parts	X		
Awaiting Parts Rate		X	
Break Rate	X		X
Cancellation Rate			

MEASURE	56th F Wing	31st F Wing	52nd F Wing
Cannibalization Rate	X		X
Cannibalizations Per Average Possessed Aircraft			
Cannibalizations (removals only) Per Departure Per Sortie			
Cannibalization Rate			
Cannot Duplicate Rate			
Chargeable Deviations	XX		X
Code 3 Breaks	X		
Code 3 Break Rate			
Combat Rate			
Deferred Discrepancy Rate (repair which can't be done within 5 days)			
Delay Discrepancies	X		X
Delayed Discrepancy Average			
Delay Discrepancy Average, Aawaiting Maintenance			
Delayed Discrepancy Average, Awaiting Parts			
Deployability			
Dropped Object Rate			
Engine Foreign Object Damage Rate			
Essential System Repair Time Per Flight Hour			
Fault Detection Rate			

MEASURE	56th F Wing	31st F Wing	52nd F Wing
Fault Isolation Rate			
Fleet Availability			
Fix Rate	X	X	X
Fully Mission Capable Hours	XX		
Fully Mission Capable Rate			
Ground aborts	X	X	
Ground Abort Rate			
Hangar Queen Days	X		
Hangar Queen Rate	X		
Hours Flown	X		X
Local Sorties Flown	X		
Local Sorties Scheduled	X		
In Flight Emergencies	XX	X	
In Flights Emergency Rate	X		
Maintainability			
Maintenance Completes	X		
Maintenance Delivery Reliability			
Maintenance Man Hour Per Fly Hour - Corrective			
Maintenance Man Hour Per Fly Hour - Improvement			
Maintenance Man Hour Per Fly Hour - Preventive			
Maintenance Man Hour Per Fly Hour - Support			

MEASURE	56th F Wing	31st F Wing	52nd F Wing
Maintenance Man Hour Per Life Unit			
Maintenance Personnel Per Operational Unit			
Maintenance Plan Rate			
Maintenance Starts	X		
Maintenance Turn Time			
Man Hour Per Fly Hour	X		
Man Hour Per Sortie	X		
Max Schedule Points Earned			
Max Schedule Points Possible			
MDC Man Hours	X		
Mean Down Time			
Mean Repair Time			
Mean Time Between Critical Failure			
Mean Time Between Maintenance Actions			
Million Ton Miles Per Day			
Mission Capable Hours	XX		
Mission Capable Rate		X	X
Non Mission Capable Both Hours	XX		
Non Mission Capable Hours	X		
Non Mission Capable Both Rate			
Non Mission Capable Rate			

MEASURE	56th F Wing	31st F Wing	52nd F Wing
Not Operationally Ready - Maintenance			
Number of Aircraft Necessary to Perform Mission			
Number of Aircraft Successfully Employed			
Object Utilization Rate			
O&M Days	X		
O&M Days Not Flown	X		
Partially Mission Capable-Both Hours	XX		
Partially Mission Capable Hours	X		
Partially Mission Capable Rate			
Partially Mission Capable-Both Rate			
Possessed Availability Rate			
Pproductivity factor			
Payload			
Possessed Availability Rate	XX		
Pproductivity factor			
Program Hours	X		
Program Fly hours	X		
Program Hour UTE Rate	X		
Program Sorties	XX		
Program UTE rate	XX		

MEASURE	56th F Wing	31st F Wing	52nd F Wing
Rekurs	XX		
Recur Rate	XX	X	X
Refueling Time			
Regeneration After Deployment			
Reliability			
Repair Turn-Around Time			
Repeats	XX		
Repeat Rate	XX	X	X
Retest Okay Rate			
Sorties Flown	X		X
Sorties Scheduled	X		X
Sustainability			
Total Abort Rate	XX		X
Total Aircraft Possession Hours			
Total Non Mission Capable Maintenance Hours	XX		
Total Non Mission Capable Supply Hours	XX		
Total Non Mission Capable Maintenance Rate		X	X
Total Non Mission Capable Supply Rate		X	X
Total Partially Mission Capable Maintenance Hours	XX		
Total Partially Mission Capable Hours	X		

MEASURE	56th F Wing	31st F Wing	52nd F Wing
Total Partially Mission Capable Maintenance Rate			
Total Partially Mission Capable Supply Rate			
Utilization Rate			