



Kerner, Susan E. & O'Mara, Francis E., US Army Research Institute for the Behavioral and Social Sciences, Alexandria, Virginia. (Wed. A.M.)

Commanders' Assessment of Unit Effectiveness Measures

The purpose of this research was to investigate how the important area of unit effectiveness is assessed in the Army. Data was collected from a sample of senior Army commanders regarding their perceptions of existing standard Army measures of battalion effectiveness. These measures naturally classify into three groups: (1) command indicators (e.g., AWOL rates, Articles 15); (2) readiness measures (e.g., equipment rated ready, annual general inspections); and (3) the personal judgments of subordinate Army leaders. Senior Army leaders chose those measures from all of these groups which provided for them the most accurate picture of a battalion's effectiveness. It was found that military leaders not only have predetermined attitudes toward all existing effectiveness measures, but that even when this rater bias is controlled, there exists a definite preference for specific groups of measures. The command indicators were found to have the least perceived validity and utility for Army leaders, while personal judgments and readiness measures were rated significantly higher for their credibility in assessing battalion effectiveness.

AD P 001339



PREVIOUS PAGE IS BLANK

## COMMANDERS' ASSESSMENT OF UNIT EFFECTIVENESS MEASURES

Susan E. Kerner-Hoeg and Francis E. O'Mara

### INTRODUCTION

A critical facet of successful operations by any organization is the continual monitoring of organizational performance. The information gathered through such activity is of importance for the development of realistic goals, for planning optimal strategies for achieving these goals, and for the identification and remediation of organizational deficiencies. Nowhere is organizational effectiveness measurement more vital than in the military, given the potentially disastrous consequences of misjudging national defense capabilities. Further, the estimate of aggregate military potential has broader national implications inasmuch as such estimates influence decisions in other areas of national concern, such as development of Federal budget priorities and the formulation of U.S. foreign policy. Thus, the development of means to accurately assess the strengths and deficiencies of military units is a vital concern both for the military as well as for our larger society.

Given the importance of measuring unit effectiveness, it is not surprising that the Army has traditionally monitored quantified measures of many facets of unit operations at all echelons. These measures have encompassed such disparate areas as the compilation of the maintenance status of mission-essential equipment to the tallying of chapel attendance by unit personnel. The manifest importance of unit effectiveness assessment is evidenced by the command attention paid to it and the diversity of measures employed in this assessment. Despite such attention, however, there has been a growing body of criticism regarding the accuracy and adequacy of current methods of monitoring the effectiveness of Army units.

Much of this criticism has centered around reported deficiencies in systems of unit readiness reporting (Barzily, Catalogne, and Marlow, 1980; Bowser, 1976; Robinson, 1980; Sorley, 1980; U.S. Army War College, 1976). This degree of attention is appropriate since this system constitutes the major means by which higher echelons monitor the effectiveness of Army battalions and separate companies. Even though the data from this reporting system provide major input to the development of Army contingency plans and guide high level resource allocation decisions, the consensus of opinion of those who have examined this system is that it is seriously faulted. As an example, in the Army War College study (U.S. Army War College, 1976), questionnaires measuring perceptions regarding the Unit Status Report (USR) were administered to approximately 2100 Army personnel. A full 70 percent of this sample reported that the USR does not reflect the true readiness condition of a unit. This opinion was likewise voiced in the course of interviews with over 1200 personnel conducted as another component of this same study. In addition to some technical problems in the actual computation of indices contained on the USR, this study found two major factors undermining the accuracy and credibility of this reporting system. The first factor concerned the substantial degree of latitude for subjective interpretation of unit conditions that was permitted in filling out the Unit Status Report. As an example, the Training Readiness Condition index, which constitutes a major component

of the unit's readiness, is based totally upon the unit commander's subjective estimate of the number of weeks of training the unit would need to be fully ready for combat. It was the conclusion of this study that "The training portion of the USR was too subjective to be anything more than a wishful-thinking guess. The training REDCONs being reported are therefore regarded as both inflated and invalid by a sizable majority of those interviewed, particularly at company level." This opinion has also been advanced by others who have examined the validity of current unit readiness reporting procedures (Robinson, 1980; Ross, Murphy, March, and Robinson, 1979; U.S. Army Concepts Analysis Agency, 1975).

The second major problem in the USR uncovered in the Army War College study concerned a conviction by those surveyed and interviewed that there was pressure on unit commanders to portray the unit's capabilities in the best possible light even to the extent of masking genuine unit deficiencies. This problem area is one which bodes ill not only for the validity of the Unit Status Report itself, but also for the validity of any systematized quantification of unit readiness indices. A key conclusion of this study therefore was that the current Unit Status Report reflected Army units not as they actually were but rather the units as all would wish them to be.

Since the publication of that study, efforts have been made to revise and improve unit readiness reporting by increasing the reliance on objective measurements of unit conditions rather than on the subjective interpretation of the unit's capabilities. As of 1980, however, Sorley held that the Unit Status Reporting system continued to suffer major deficiencies. The most central of these deficiencies continued to be the need to separate the process of evaluating and monitoring unit effectiveness from the process by which the performance of Army officers is evaluated. Sorley believes that only by removing the responsibility for unit readiness reporting from the chain of command, which likewise evaluates the performance of the individuals who provide unit effectiveness data, can the real or perceived pressure to inflate estimates of unit effectiveness be removed.

Sorley further holds that the Unit Status Report has excluded variables which are essential to combat readiness and therefore the USR can only partially reflect the total capability of the unit. Factors such as unit cohesion and the turnover and competence of key unit personnel are those which he feels are important contributors to total unit capability but which are not now employed in estimating the unit's effectiveness. Sorley likewise suggests that the information contained on the USR be complemented with the professional judgment of individuals familiar with the unit. This, of course, would only be feasible where this judgment could be rendered frankly and openly.

The evaluation of unit effectiveness in the Army is not restricted to the USR. The Army has had a long tradition of monitoring an extensive series of variables which purportedly reflect the state of morale and discipline in the unit. Known collectively as "command indicators" or "traditional indicators," this set of unit measures typically includes such variables as

reenlistment rates, crime rates, and indices associated with the administration of military justice. Unlike the USR measures, these indices are not systematically reported at the unit level to the higher echelons of the Army command structure. However, unit measures on these variables are used frequently at the local level as indicants of unit conditions and problems. Sorley (1979) has been critical of the use of such measures inasmuch as he sees them leading to a "management by statistics" in which those factors which are more readily quantifiable are given greater command emphasis than those which are more difficult to quantify, but which more substantively support and reflect unit effectiveness. Too often, he feels, command attention is expended on "getting the numbers right" in such areas of questionable military value as motor vehicle accident rates or the number of letters of indebtedness among unit personnel. This occurs at the expense of diverting command attention from such areas as unit training and equipment maintenance, which are more directly supportive of the unit's mission. The position underlying his assertions is that statistical indices of unit operations, particularly those relevant to the personnel area, are of questionable utility in assessing areas pertinent to unit effectiveness. This is somewhat inconsistent with his position (Sorley, 1980) that the USR be supplemented with measures in such areas as drug abuse, race relations, and the alienation and commitment of the unit personnel. Clearly, some statistical indices are more germane to unit effectiveness assessment than others. The question remains unresolved as to the identity of these measures. The proliferation of statistical indices used to monitor unit functioning has been fed by the variation in opinion as to which of the wide variety of possible measures are the most accurate indicants of unit capability. This proliferation has in turn led to many of the abuses and problems which have been identified in the literature.

The purpose of this research is to contribute to the resolution of these problems by examining the value of the most typically employed statistical indices in reflecting unit effectiveness. To date there has been no systematic examination across the broad spectrum of unit effectiveness measures which would permit a determination of the relative value of these measures. The absence of such information leaves unchallenged the possible reliance on inaccurate or incomplete assessment of unit effectiveness and thus the development of priorities based on apparent rather than real problems.

#### METHOD

Forty-eight battalion commanders, twenty-eight brigade commanders, and eight general officers located at six CONUS installations were interviewed on the topic of battalion effectiveness. During the approximately one-hour-long interviews, each subject was asked to discuss the most pressing management problems confronting him in maintaining readiness, to operationally define battalion effectiveness, and to evaluate the performance of his subordinate battalions. Each subject was also asked to assess various given measures of battalion effectiveness. These measures can be classified into three groups: Readiness Measures, Command Indicators, and Personal Judgments.

The first of these groups, Readiness Measures, is a relatively direct assessment of a unit's capability to perform its mission. This group of measures includes the REDCON ratings from the USR, the percentage of unit equipment that is operational (as gleaned from USR data) as well as ARTEP and AGI results. A listing and definition of each of these measures can be found in Figure 1.

Command Indicators are those measures which are traditionally held to reflect a unit's state of morale and discipline. The specific Command Indicators used in this study are listed and defined in Figure 2.

Personal Judgments are those opinions of battalion effectiveness held by individuals at various echelons. The ascending hierarchy of authority and the six specific levels used in this study are as follows: (1) Service members in the battalion (SM); (2) Noncommissioned officers in the battalion (NCOs); (3) Company grade officers in the battalion; (4) Brigade Commanders; (5) Assistant Division Commanders; and (6) Division Commanders.

Subjects were asked to indicate "how accurate an assessment of battalion effectiveness would be if it were based on any single piece of information from the list provided." A measure providing complete accuracy would be rated 100%, while a measure providing no information on unit effectiveness would be rated 0%.

Subjects were further asked to choose from the given list of measures the five which, in combination, would provide "the most complete picture of a battalion's overall effectiveness."

## RESULTS

Analyses of these data began with an examination of the degree to which there was a difference in the perceived validity of the effectiveness measures across positions (i.e., battalion commander vs. brigade commander vs. general officer). A three-level one-way ANOVA was therefore performed on the validity ratings given to each of the unit effectiveness measures. Of the twenty-two measures tested, on only one was there a significant position difference (Drug Arrest Rate). It was concluded that there existed no consistent position differences in the perceived validity of the unit effectiveness measures, since such a proportion of significant results is essentially what would be expected from chance alone. Accordingly, the data from the three groups were combined in all further analyses.

The mean accuracy ratings assigned to each measure are rank ordered and presented in Table 1. As shown, a wide range of mean ratings was obtained, varying from 72.5% accuracy attributed to ARTEP results to an accuracy rating of only 29.2% for desertion rates. In general, the Readiness Measures and the Personal Judgments were given the highest validity as measures which individually render an accurate assessment of a battalion's effectiveness.

Table 2 presents a rank ordering of the frequency with which each measure was included in the group of five providing the most complete picture of a battalion's effectiveness. The sharp drop in the frequency of selection after the fourth measure indicates that these first four are important measures in providing an overall effectiveness assessment. Measures of readiness, specifically the ARTEP and AGI, and the personal judgments of those in the unit, specifically the company grade officers and NCOs, were seen to provide the most information about battalion effectiveness. Further, the top four measures in Table 2 are the same as the top four in Table 1, implying that these measures are the most valid, whether they are considered individually or in combination. For the most part, the group of command indicators are located at the bottom of the continuum on both Tables 1 and 2.

As seen in Table 1, all measures within each group (Command Indicators, Personal Judgments, Readiness Measures) tended to be assigned similar accuracy ratings. Thus, there appears to have been a consistent rating applied to all measures within each of the three groups of measures. That is, while the judgments of unit NCOs may have been accorded higher validity by subjects than the judgment of division commanders, the fact that both measures entail personal judgments tended to produce very similar validity ratings for both measures. To test this, coefficient alphas were computed for each of the three groups to determine the internal consistency of this grouping. These reliability measures are presented in Table 3A. All three coefficient alphas are above .89, revealing a high degree of internal consistency within each group of effectiveness measures.

In order to detect whether there was a consistent style of rating (i.e., preference of one group of measures to the exclusion of others), the relationships among the three groups of measures was examined. A mean rating was computed for each subject for each of three groups of measures. Pearson correlations were in turn computed among these mean ratings. These intergroup correlations are presented in Table 4A. A substantial positive correlation exists among the three groups, suggesting that even in the presence of a wide variation of mean ratings (as shown in Table 1), there was a tendency for subjects to display a rater bias reflecting a global impression of the validity of any formal effectiveness measure. Thus, a subject who gave high ratings to one group of measures likewise gave high ratings to the other two groups. As an extreme example, one subject's accuracy ratings of the individual measures ranged from 90-100, while another's ranged from 0-16.

Thus, to correct for this rater bias, a set of corrected ratings was computed for each subject on each measure. This was accomplished by first computing the average rating on all measures given by each subject. Each subject's average rating was then subtracted from his original rating on each measure to establish a set of corrected ratings for each subject.

The rank ordering for these corrected ratings agrees totally with the rank ordering of the uncorrected ratings (Table 1), indicating that these earlier results were not artifacts of subjects' biases toward effectiveness measures in general. Thus, the ordering of the ratings for those with high confidence in unit effectiveness measures is the same as that for those with low confidence in unit effectiveness measures. The existence of this rater bias, however, does raise the possibility that the high internal consistencies in each of three groups of measures was not due to these existing as natural groupings in the minds of the raters, but rather was an artifact of the rater bias. It is possible that the rater bias produced high intercorrelations among the ratings given to all of the unit performance measures. This would in turn produce high coefficient alphas for the ratings given to the measures within each of the groups. This interpretation is made plausible by the substantial positive correlations among the mean ratings given to each group of measures (Table 4A). To test this interpretation, coefficient alphas were recomputed for each of the three groups, based on the corrected ratings. These reliability measures are presented in Table 3B. Though the alphas drop slightly from those based on the uncorrected ratings, the internal consistency of the groups remains acceptable, indicating that while these coefficient alphas were inflated by the rater bias, they were not totally attributable to them.

Intercorrelations among the three groups of measures were next computed, based on the corrected ratings. These intercorrelations are presented in Table 4B, showing correlations which are negative, in contrast to those based on uncorrected ratings (Table 4A). The use of difference scores to correct for the rater bias reveals a clearer picture of Army leaders' preferences for different types of measures. That is, a definite tendency to favor one type of measure over the other two is now displayed.

#### DISCUSSION

The results of this research indicate that the evaluation of the existing Army battalion effectiveness measures has two facets. First, there is the attitude military leaders have towards the general category of formal battalion effectiveness measures. There was seen wide variation from commander to commander in this attitude, with some commanders attributing very little credibility to any of the battalion effectiveness measures. Second, there appears to exist in the perceptions of senior commanders a distinct typology of battalion effectiveness measures. The wide spectrum of measures studied in this research broke down into only three types, or groups, of measures in the eyes of the interviewed commanders--Readiness Measures, including formal evaluations of unit capability; Personal Judgments, the estimation of battalion effectiveness by individuals within the unit and above the battalion in the chain of command; and Command Indicators, which include traditional measures of unit morale and discipline. These groups were seen to have differing degrees of utility as indicants of battalion effectiveness. Readiness Measures and Personal Judgments were seen to be similar in value whereas Command Indicators

were seen to have substantially less merit as indices of unit effectiveness. These findings are very much in line with those of other studies both in military (Bowser, 1976) and in civilian organizations (Mahoney & Weitzel, 1969; Weitzel, Mahoney & Crandall, 1971) which show that managers display a clear preference for operationally oriented measures as indices of organizational performance, especially in contrast to measures related to the personnel and human relations areas.

From these results, at least two points are salient in their operational implications: the low value ascribed to Command Indicators as measures of unit effectiveness and the high value ascribed to Personal Judgments.

The first of these two results calls into question the practice of using Command Indicators as measures of unit effectiveness. This finding is surprising in light of the long history of utilization which such measures have had in the military. Despite this tradition, the interviewed commanders saw each of these measures as being ambiguous in their implications regarding unit effectiveness since a high score on any of these measures could stem from a multitude of possible causes. Thus, while each of these measures may truly reflect the status of various personnel areas in the unit, this information by itself is held to be too ambiguous to judge the effectiveness of the command. The use of such measures should therefore be restricted to the monitoring of personnel trends and issues at levels above the unit. If such measures are used at the unit level to assess overall unit effectiveness, the present results suggest that these measures should be used only in combination with other indicants of unit operation. In using Command Indicators to assess the morale and the state of discipline of a given unit, especially high or low scores on these measures should not be taken at face value but rather should precipitate a fuller investigation into the reasons behind the scores. Only with this fuller base of information can sound conclusions be reached regarding the state of the unit.

The relatively high validity ascribed to the Personal Judgment measures supports Sorley's contention that an ideal Unit Status Reporting system would include the reporting of the professional judgment of battalion effectiveness by key personnel. In light of the already well-documented pressure on unit personnel to have USR data appear maximally positive, it is doubtful that judgments rendered within the context of the present USR system would substantially contribute to a fuller assessment of unit readiness at higher command levels. At the unit level, however, the high ratings given to NCO and junior officer judgments of battalion effectiveness support the notion that unit commanders can best assess the day-to-day status of their units by relying on the input of their subordinates. It is these individuals who have the most detailed and direct knowledge of unit strengths and weaknesses inasmuch as it is they who directly address these areas in the course of their daily duties. For such a rich source of information to be used to its fullest potential, however, the commander must have the skill to solicit this information in a frank and unbiased form. The high value attributed to these estimates of unit performance, therefore, underlines the importance of leadership development for the Army. Without the requisite leadership skills to involve his or her subordinate leaders in the development of the



total unit potential and to solicit and evaluate their frank judgment of unit capabilities, unit commanders would be deprived of one of the major instruments of monitoring and improving the operation of their units.

In the concept of prior work in this area (Concepts Analysis Agency, 1975; Robinson, 1980; Ross, et. al., 1979; Sorley, 1979; US Army War College, 1976), the present results call for a careful re-examination of the structure and process currently employed to evaluate unit effectiveness. There continues to be a pressing need to amend and improve current unit effectiveness assessment systems used in the military. The process is ongoing; it needs to continue and accelerate.

## REFERENCES

- Barzily, Z., Catalogne, P. R., & Marlow, W. H. Assessing Marine Corps Readiness (T-430). Washington, D.C.: The George Washington University, September 1980.
- Bowser, S. E. Determination of Criteria of Operational Unit Effectiveness in the US Navy (NPRDC TR 76TQ-41). San Diego, California: Navy Personnel Research and Development Center, August 1976.
- Mahoney, T. A. & Weitzel, W. Managerial Models of Organizational Effectiveness. Administrative Science Quarterly, 1969, 14, 357-365.
- Robinson, R. M. Objective Measurement of Training Readiness. Carlisle Barracks, Pennsylvania: US Army War College, May 1980.
- Ross, G., Murphy, J., March, M., Robinson, R., & Tullington, B. Military Organizational Effectiveness/Readiness and Sustainability. Final report to Headquarters, Department of the Army, under Contract DAAK21-79-C-0015, McLean, Virginia: Science Applications, Inc., September 1979.
- Sorley, L. Professional Evaluation and Combat Readiness. Military Review, October 1979, 41-53.
- Sorley, L. Prevailing Criteria: A Critique. In Sarkesian, S. C. Combat Effectiveness--Cohesion, Stress, and the Volunteer Military. Beverly Hills: Sage Publications, 1980.
- US Army Concepts Analysis Agency, Readiness System Study, Phase I Analysis. AD-A029 387, August 1976.
- US Army War College, US Army Unit Readiness Reporting. ACN 75025, June 1976.
- Weitzel, W., Mahoney, T., & Crandall, N. A Supervisory View of Unit Effectiveness. Final report to Office of Naval Research, under Contract ONR N00014-68-A-0141-0003, Minneapolis, Minnesota: Industrial Relations Center, University of Minnesota, 1971.

Figure 1

READINESS MEASURES

**OVERALL READINESS**

A battalion's overall readiness status as reported in the monthly Unit Status Report.

**PERSONNEL READINESS**

A battalion's personnel readiness status as reported in the monthly Unit Status Report.

**EQUIPMENT ON HAND**

An index of the degree to which a battalion possesses all authorized equipment, a reflection of the battalion's supply system.

**EQUIPMENT SERVICEABILITY**

The maintenance status of a battalion's equipment, a reflection of the battalion's maintenance system.

**EQUIPMENT ON HAND RATED READY**

The proportion of equipment a battalion actually has on hand that is operational.

**ARTEP**

The percentage of the missions/tasks rated "satisfactory" during a battalion's most recent field training exercise.

**AGI**

The percentage of the areas rated "satisfactory" during a battalion's most recent annual general inspection.

Figure 2

COMMAND INDICATORS

ARTICLES 15	The percentage of enlisted personnel administered non-judicial punishment (e.g., fines, reductions in grade) during a given month.
COURTS MARTIAL	The percentage of enlisted personnel receiving a court martial during a given month.
AWOL	The percentage of enlisted personnel who were involved in unexcused absences during a given month.
DESERTIONS	The percentage of enlisted personnel who deserted during a given month.
FIRST TERM RE-UP	The percentage of a battalion's first-term reenlistment objective that was achieved in a given month.
CAREER RE-UP	The percentage of a battalion's reenlistment objective for career personnel that was achieved in a given month.
CRIMES OF VIOLENCE	The percentage of a battalion's enlisted strength involved in crimes of violence in a given month.
PROPERTY CRIMES	The percentage of a battalion's enlisted strength involved in crimes against property in a given month.
DRUG ARRESTS	The percentage of a battalion's enlisted strength arrested for drug and marijuana violations in a given month.

Table 1

Accuracy Assessments of The Effectiveness Measures

<u>Measure</u>	<u>Mean Accuracy Rating</u>
ARTEP	72.5 %
Company grade officer's judgment	71.2 %
NCO's judgment	71.0 %
AGI	66.7 %
Equipment on hand rated ready	63.3 %
Brigade commander's judgment	62.5 %
Service member's judgment	62.1 %
Equipment status	59.9 %
Overall readiness	57.2 %
Assistant Division Commander's judgment	56.9 %
Division Commander's judgment	53.7 %
First-term reenlistment rate	51.3 %
Personnel readiness	50.7 %
Equipment on hand	46.6 %
AWOL rate	45.9 %
Career reenlistment rate	43.8 %
Crimes against property	38.1 %
Article 15s	37.4 %
Crimes of violence	36.7 %
Courts-martial rate	32.5 %
Drug/marijuana convictions	30.8 %
Desertion rate	29.2 %

Table 2

Group of Five Measures  
Providing Most Complete Picture of Effectiveness

<u>Measure</u>	<u>Number Choosing Measure</u>	
ARTEP	63	(75 %)
NCO's judgment	54	(64%)
AGI	48	(57%)
Company grade officer's judgment	44	(52%)
Brigade commander's judgment	28	(33%)
First-term reenlistment	27	(32%)
Service member's judgment	25	(29%)
Overall readiness rating	25	(29%)
Equipment on hand rated ready	21	(25%)
AWOL rate	15	(18%)
Personnel readiness	13	(15%)
Equipment status	12	(14%)
Article 15s	9	(11%)
Career reenlistment rate	7	(08%)
Division commander's judgment	7	(08%)
Assistant division commander's judgment	5	(06%)
Crimes against property	4	(05%)
Drug/marijuana convictions	3	(04%)
Crimes of violence	3	(04%)
Equipment on hand	3	(04%)
Courts-martial rate	2	(02%)
Desertion rate	1	(01%)

Table 3

Reliability Coefficients of Effectiveness Measure Groups

	A Uncorrected Ratings	B Corrected Ratings
Readiness Measures	.914	.656
Command Indicators	.950	.704
Personal Judgments	.894	.644

Table 4

Effectiveness Measure Intergroup Correlations

	A Uncorrected Ratings			B Corrected Ratings		
	Readiness Measures	Command Indicators	Personal Judgments	Readiness Measures	Command Indicators	Personal Judgments
Readiness Measures	1.000			1.000		
Command Indicators	.6537	1.000		-.6641	1.000	
Personal Judgments	.6297	.6802	1.000	-.3391	-.4781	1.000

