2 of 3

AD 603 303

VOLUME I

ASSESSING HUMAN FACTORS REQUIREMENTS IN THE TEST AND EVALUATION STAGE OF SYSTEMS DEVELOPMENT

TECHNICAL REPORT

June 1964

Report No. ND 64-68

Prepared Under
Contract NOnr 4203(00)
for
New Developments Research Branch
Personnel Research Division
Bureau of Naval Personnel

61 p

Prepared by
James M. McKendry, Ph.D.
Paul C. Harrison, Ph.D.
HRB-Singer, Inc.
Science Park
State College, Pennsylvania

ACKNOWLEDGMENTS

This research was conducted under the technical direction of the New Developments Research Branch, Personnel Research Division, Bureau of Naval Personnel. Considerable assistance and guidance was provided by Captain R. M. Stuart, USN, Division Director, Mr. A. A. Sjoholm, Branch Head, Mr. W. L. Hopkins, Scientific Officer, and Mr. N. T. Burwell, Contract Monitor. Enthusiastic cooperation was also received from numerous personnel of the Staff of the Commander, Operational Test and Evaluation Force, Norfolk, Virginia. Finally, the work of other company personnel not listed on the title page is gratefully acknowledged, especially Mr. Robert Carter and the technical consultant, Commander James Odom, USN (Ret.)

ABSTRACT

This is the first volume of a two-volume final report. Volume I outlines in detail the methodology employed in the development of a model for human factors evaluation incorporated in system testing. Volume II translates this model into an integrated approach to human factors testing by the OPTEVFOR project officer.

This report outlines an initial attack upon the problem of determining when, where, and how human factors inputs should be provided during the planning and development cycle of Naval Systems, with particular emphasis being given to measuring the consequences of such inputs in terms of systems effectiveness. All work was directed toward the Test and Evaluation (T and E) phase of system development in order to develop measures of effectiveness capable of being applied to a specific system undergoing T and E by the Operational Test and Evaluation Force (OPTEVFOR).

The procedure developed is one capable of being employed under present conditions as well as handling more complete data likely to be available in the future. By using this technique two advantages result: first, it is possible for human factors specialists to give specific estimates of the change in systems effectiveness induced by attending to or overlooking a number of human considerations; second, the procedure allows systematic accumulation of reference data by which the cost-effectiveness of human factors programs can be contrasted with that of other "software" or even hardware changes. As part of this same effort a separately published preliminary format of a reference guide was provided to OPTEVFOR project officers.

SUMMARY AND CONCLUSIONS

PERSPECTIVE

This report summarizes initial work on a question of considerable importance and magnitude. what demonstrable, beneficial or deleterious effects on systems performance occur as a result of attending to or ignoring human factors specifications in the design and development of Naval systems. One means by which such a question could be answered--albeit a time consuming, tedious one--would be to select a number of representative systems for study, each of which would have numerous design variations representing values along quantitative scales of operability and maintainability. These differently constructed variations of the same system could then be presented to groups of subjects having varied aptitudes for their work, and varied types and amounts of training. System performance could then be measured as a consequence of those variations. In short, the study would be empirical with a high degree of experimenter control along each of four independently manipulated complex dimensions. (1) the human factors design dimension, (2) the operator and maintenance technician aptitude dimension, (3) the training dimension, and (4) the systems dimension. Results of such work could be summarized in terms of a series of prediction equations relating systems performance to a host of "human factors."

Needless to say, the scope of the problem and countless practical contingencies preclude use of the approach described above. The most common substitute for this type of evidence involves the use of two appeals. First, fear is struck into the designer by citing horrible examples of practical cases when things went wrong. Second, an extrapolation is made from the data of countless studies relating human factors variables to some aspects of system performance. A difficulty arises in using this combination of approaches; however, since available functional relationships relate human factors effects to intermediate--not final--criteria, such as the mean downtime per trouble, chances of making an error, etc. What results is a serious question concerning how both maintenance technician and operator effects (training and aptitude) work in combination and how these effects interact with the equipment features and system influences.

APPROACH

The purpose of this project was to examine the question of how human factors effects can be related to systems performance through the use of existing field data and to demonstrate the feasibility of the proposed technique on an actual Naval system. Particular emphasis was placed upon data gathered by the Operational Test and Evaluation Force and the role that such a group could play as a data source. For a variety of reasons, the improved TARTAR system was selected as the example and research was conducted in cooperation with the staff of the Commander, Operational Test and Evaluation Force, Norfolk, Virginia (COMOPTEVFOR).

The problem was attacked by devising a model relating human factors effects to actual TARTAR systems performance. Primary difficulties involved relating intermediate criteria to "ultimate" systems performance and allowing for the combined effect of both operator and technician effects on systems performance. Since there was only one equipment configuration with which to work, equipment design features were not investigated, nor was any attempt made to relate the TARTAR system to some rigorously defined classification scheme for Naval systems in general. This choice, while being forced essentially by constrictions dictated by using field data, was not felt to preclude consideration of such factors since appropriate terms could be added to the model at a later time.

RESULTS

Conclusions of the study may be summarized as follows:

(1) It is possible to construct models which relate intermediate criteria to ultimate systems criteria--at least for the TARTAR--and to take account of the combined technician and operator effects within a single model. This can be done in a manner which simplifies considerably the task of relating existing data to their implications in systems performance terms.

(2) Any demonstrated improvement gained by attending to human factors effects can be easily translated into systems improvement terms provided some basic data are available, viz., equipment availability figures and operating time data.

(3) OPTEVFOR test conditions are such that sample sizes are very restricted and the amount of possible human factors data per test is comparatively small. Still, it is possible to obtain valuable estimates of the importance of human factors by the use of such data, especially when results of several tests are pooled.

(4) When OPTEVFOR data of the type described within are added to fleet performance data and aptitude and training data available from the Bureau of Naval Personnel, more definitive relationships can be established and prediction equations can be generated. To complete the necessary data pool, however, simulator studies would be of considerable use.

(5) From all present indications, prediction equations can be generated as early in the development and design sequence at the prototype development stage and in less detail at earlier stages using estimations.

These general conclusions are warranted by what was found to be the case during the one year's effort. However, certain restrictions should be kept in mind. First, data of the type needed from and by OPTEVFOR are *not* presently available, although they could be made so by following the approach suggested in Volume II in this same series. Second, data now available from scattered sources must be collated and integrated *systematically* before precise prediction equations can be devised. Third, the ability to use such equations during earlier system design and development stages, while appearing readily feasible, remains to be demonstrated. Finally, attention should be paid to the problem of insuring that the plans suggested in Volume II are indeed workable in detail at OPTEVFOR. These restrictions provide the basis for recommending that further efforts be conducted in the same field of activity.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

1.0 OBJECTIVES AND METHOD

This chapter states project objectives and stresses the methodological orientation taken. It includes a technical description of the plan of attack and a discussion of how this plan was adapted to the particular system selected for study. It closes with a description of the extent to which the method, and the basic question which caused the study to be undertaken, were tested within this first year's effort.

1.1 PROJECT DESCRIPTION

1.1.1 Nature of the Work.

This investigation stresses methodology, i.e., the development of tools or techniques by which a problem can be approached. The problem under consideration is how to assess (test and evaluate) the contribution of human factors to output effectiveness of Naval systems. Although this problem can be attacked at a series of points in the research and development cycle, an obvious point of departure is the operational test and evaluation stage.

The reasons for this choice are many. First, by taking advantage of operations research techniques and accumulated knowledge available at OPTEVFOR, considerable time can be saved in reaching the crux of the problem. Second, if similar efforts at earlier system development stages are to be successful, they must by necessity be integrated with an approach that includes operational test and evaluation. Third, at this stage of system development, subproblems become more discernible and more demonstrable, although their solutions may not.

1.1.2 Goals.

Purposes of the research were four in number: First, to develop the methodology previously described. Second, to demonstrate its feasibility on a complex system (the improved TARTAR) which recently underwent operational tests and evaluation. Third, to derive characteristics and a format for future test plans in which human factors tests will parallel existing systems tests. Fourth, to uncover and clarify areas requiring further research.

In view of these goals, project personnel had to work closely with operational personnel from the office of the Commander, Operational Test and Evaluation Forces, Norfolk, Virginia. By virtue of a close association with OPTEVFOR personnel, the project staff was able to increase the chances of deriving measures which are of practical importance and at the same time compatible with techniques currently employed by OPTEVFOR. In return, the OPTEVFOR Staff obtained a separately published plan of attack to human factors problems which should simplify their task of deciding whether or not System X will operate satisfactorily in the fleet using normal personnel complements.

1.2 TECHNICAL ORIENTATION

1.2.1 Clarification of Terms.

By virtue of its emphasis on methodology, the project attempted to treat human factors problems in integrated form rather than on a piecemeal basis. For this reason before discussing the approach taken, it is necessary to clarify what is meant by the term "human factors" as well as stipulating the role such effects play in system performance.

By "human factors" we mean the abilities, skills, and other important characteristics determined by training of individuals serving as equipment operators, decision makers, and maintenance technicians. *In addition,* we also consider *equipment* characteristics which influence how effectively a given person can accomplish his assigned task. Thus, interest includes (1) the persons themselves, and (2) equipment features which determine working effectiveness through their effect on task difficulty. For example, lack of attention to maintainability in design means that technicians of higher skill levels are needed to keep equipment availability at some fixed level.

The type of tradeoff problem cited above is only one of many. One of a more basic nature is the operator-technician tradeoff in systems design. For example, when building a system such as the TARTAR, designers generally attempt to reduce (or minimize) the time required to fire a missile. In accomplishing this goal many portions of the firing cycle are automated or placed under machine control. However, it is worthwhile to note that such automation requires substantial increases in the *number* of electronic components involved as well as increases in the complexity of various circuits within the system.

As has been amply demonstrated in reliability literature increases in equipment complexity serve, in general, to reduce reliability (the mean time between failures). Thus, to a marked extent, reducing the number of human operations through increasing automation of operator functions is accompanied by increases in the work *load* (number of troubles) placed upon the maintenance force. At the same time, the difficulty of finding and correcting each trouble is complicated by the increased equipment complexity. Finally, systems checkout procedures become more complicated. In short, human factors problems are not eliminated by automation; rather, the nature of these problems changes. It might be noted here that the introduction of micro-miniaturization, modularization and automatic test points have reduced but not eliminated these difficulties.

1.2.2 System Performance Criteria.

Because of the existence of the operator-maintenance technician tradeoff problem, system evaluators at OPTEVFOR, in addition to their other duties, are faced with the problem of determining if the operator and/or technician demands of a particular system are compatible with existing personnel skills in the fleet. In arriving at such a judgment it is advantageous to have data available concerning the effects of each demand on system effectiveness. In other words, both demands should be related to the same criteria of system effectiveness.

In the case of the TARTAR system for example, an appropriate measure is the mean rate of sustained fire. The ways in which the various human factors influence this criterion are shown in Figure 1.1. The reasons for choosing these particular criteria in the TARTAR system are given below.

The criterion question, the question "How well does the system do the job it is supposed to do?", is not an easy one to answer. As a matter of fact, there often exists no single unique answer, but rather a series of answers depending on the aspect viewed and the perspective of the viewer. In the investigation and evaluation of the human contribution to performance, this

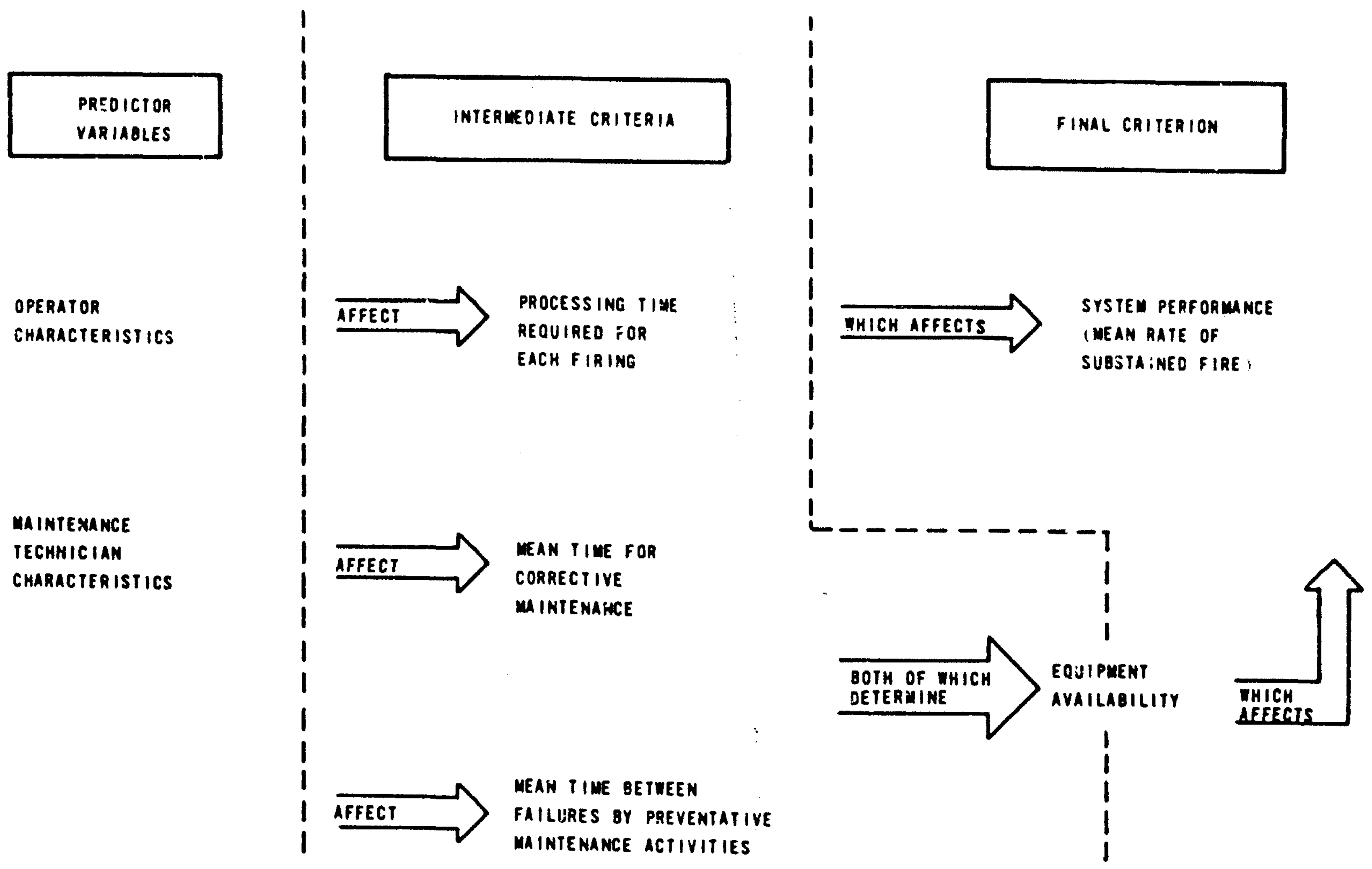

FIG. 1.1 - SIMPLIFIED ILLUSTRATION OF THE WAY HUMAN FACTORS AFFECT SYSTEM PERFORMANCE IN A SAM SYSTEM

is an intolerable ambiguity. It is necessary to define carefully an accurate, realistic, and valid index of systems effectiveness. In the case of TARTAR the possibilities for a criterion of performance are numerous. (For example, one might consider kill probability, number of targets required to saturate the system, ship survival probability or the rate of effective fire.)

Ultimately, criteria of effectiveness of any weapons system must be reduced to some quantifiable index that takes into account the values attached to the various aspects of the system's assigned mission. This may take the form of a kind of cost effectiveness index or net cost optimization. To do this for many of the possible criteria would involve consideration of variables which cannot be given specified values or held constant in any meaningful fashion (i.e., variables related to attack force tactics and characteristics). Thus it is required that the selected criterion be as free, or independent, of these effects as possible. This can be done by introducing certain assumptions in accord with the various restraints under which we are operating. The assumptions are as follows:

1. System hardware is regarded as fixed.

2. The operating environment of the DDG will be dictated by the current doctrine for a mission of interdiction and support (recognizing its inherent self defense characteristics).

3. In the posture of interdiction and support, deployment of the DDG will be such as to optimize its role in the total air defense capability.

4. When under attack, the TARTAR system will be operating under a condition of saturation or near saturation and therefore will be required to fire at a near maximum rate.[1]

---

[1] This assumption does not mean that the entire magazine needs to be exhausted. The maximal firing rate can be established reasonably accurately by considering 8 to 10 consecutive firings.

These assumptions are considered to be definitive of a realistic situation and have the added advantage of permitting us to ignore questions associated with various alternative spacings of elements of the enemy air attack. For example, by postulating the integration of the DDG into an overall air defense scheme, it is reasonable to consider that the raid will have been "thinned out" by longer range SAM's and CAP before reaching the engagability envelope of TARTAR--thereby equalizing the load placed on the system during the progress of the attack rather than permitting intervals of idleness followed by periods of complete saturation. Thus, the only question of concern is the average maintainable rate of effective fire by the TARTAR battery.

This permits an evaluation of the role of the human factors subsystem with respect to the single criterion of "average rate of effective fire." It is anticipated that under conditions expected to exist, this criterion will bear a close relationship to expected kills per minute and hence to the effectiveness of the air defense system under high-load conditions.

The use of the "firing rate" criterion will not permit us to answer directly the ultimate cost/effectiveness questions; e.g., is an increase of X-dollars in training costs for maintenance technicians justified by the expected reduction in fleet losses during subsequent air attacks? It will, however, yield *comparison* data, e.g., between the expected "firing rate" gain from an X-dollar increase in maintenance training as opposed to an X-dollar increase in operator training. Absolute questions, such as whether the X-dollars *should* be spent in the first place, are beyond the scope of this task.

1.2.3 Complicating Factors.

The conceptualization offered in Figure 1.1 is obviously simplified. In reality, the study of human factors problems is complicated by a number of subtle interactions. For example, if preventive maintenance (PM) checkouts fail to achieve proper alignment throughout the system, causing decrements less than those requiring corrective maintenance (CM), the task of the operator is made more difficult. In the TARTAR system for example, failure to "peak" the 51B radar will reduce the range where video outputs appear, thereby requiring operators to rely on audio outputs produced by doppler effects for tracking at greater ranges. This type of interaction, shown in Figure 1.2, can be illustrated by many other examples in the TARTAR system.

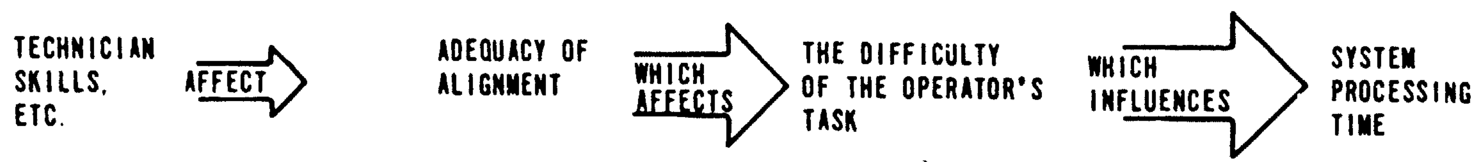

FIG. 1.2 - ILLUSTRATION OF AN INTERACTION IN SAM SYSTEMS

The important point to note from such interactions is that there is not only an operator effect and a technician effect but also an effect stemming from their particular interaction as well. This latter effect can be beneficial as well as deleterious as in the case where the operator serves to reduce trouble detection and localization times by diagnosing what has happened to his display. Incorporating these interactions into the basic picture, Figure 1.1 becomes revised as shown in Figure 1.3.

1.3 TECHNICAL APPROACH

1.3.1 Method.

Throughout the preceding discussion emphasis was placed upon the desire to express both operator and maintenance technician performance in terms of the same system effectiveness measure. Unless such a step is taken, any investigation terminates in the embarrassing position of attempting to compare "apples and oranges." For example, operator performance is traditionally measured in terms of the probability of committing an error-- or as the amount of time taken to perform a task correctly. Maintenance technician performance, on the other hand, is generally scored in terms of the amount of time taken to isolate and correct troubles, time taken to perform preventive maintenance tasks correctly, or, relatedly, as the probability of completing an assigned task at a time $\underline{t}$.

To systems designers and analysts, such as OPTEVFOR Staff, such relationships, while being useful, fall considerably short of what is needed. For these persons, interest is centered about final system performance and not "chances of errors" and "times to correct troubles." Without explicit consideration of a large number of related questions, such measures have an unknown relationship to the criterion of interest. Thus, systems analysts are given the task of interpreting what relationships exist *as well as* the task of making a judgment based upon a knowledge of such relationships.

For these reasons, an approach covering the *entire* sequence of operations shown in Figure 1.3 was implemented in the following manner.

(1) A time sequence analysis of the various operator functions was performed to determine:

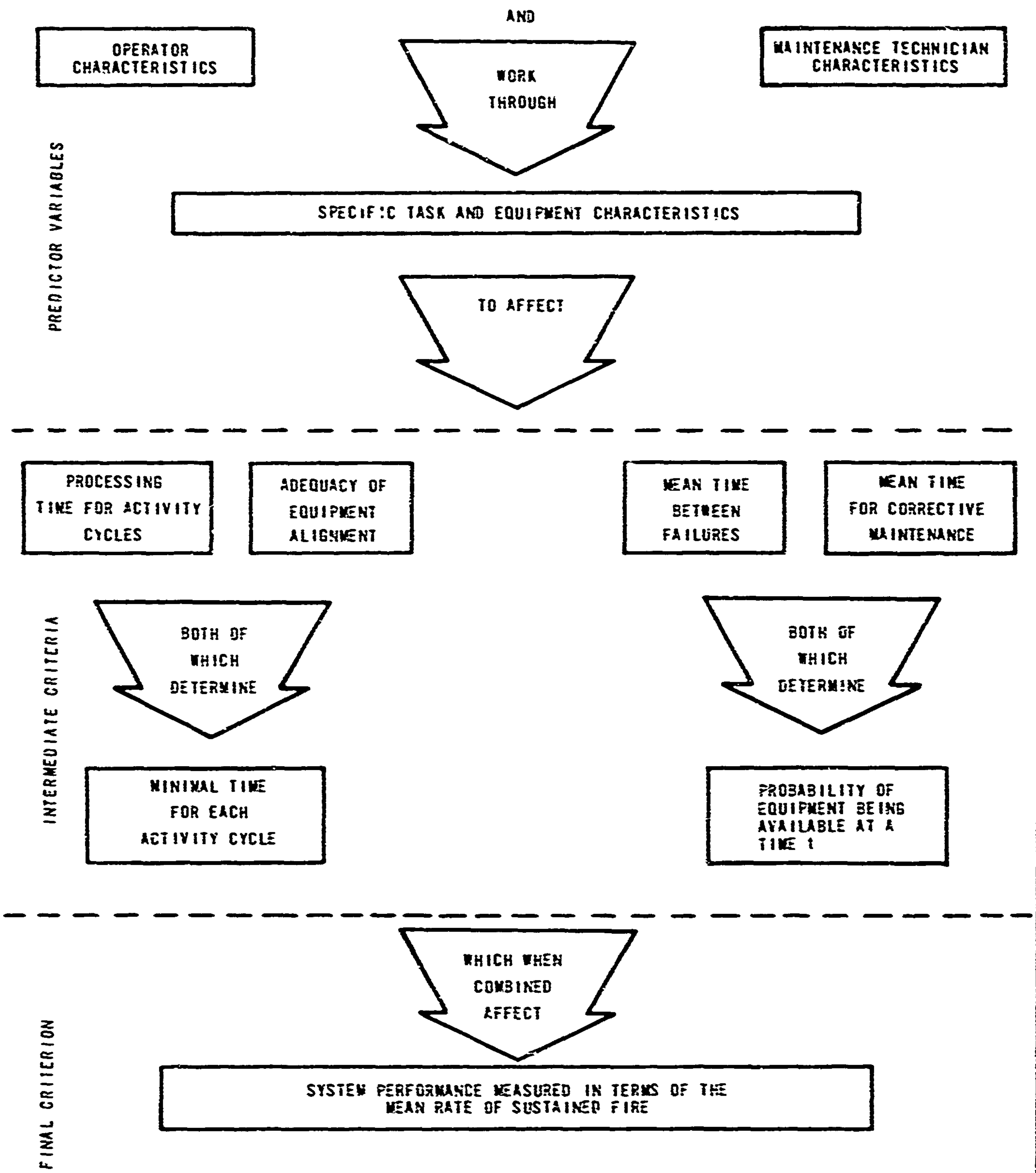

FIG. 1.3 - A MORE ACCURATE DEPICTION OF THE PROBLEMS FOR SAM'S

(a) how large a portion of the firing cycle time is associated with human operator activities, and

(b) the nature of the relationship between various operator errors and specific operator processing times.

(2) An activity analysis of the maintenance technician task was performed to determine:

(a) the ways and degree to which preventive maintenance can influence equipment reliability (mean time between failures),

(b) the influence that alignment procedures have upon the difficulty of an operator's task, and

(c) the mean time for corrective maintenance in each subsystem.

(3) A list of *potentially* important predictor variables was developed and refined on the basis of past empirical knowledge.

(4) The *potentially* important predictor variables are related to intermediate criteria by use of:

(a) simulation runs,

(b) field data gathering, and

(c) failure report analysis.

(5) The intermediate criteria were related to systems performance criteria by analytical expressions.

(6) Data gathering needs associated with the above step were catalogued (with alternatives included) for use in deriving human factors test plan suggestions, published separately as part of this same project.

1.3.2 Problems and Restrictions.

The method described above is admittedly ambitious, but, as the reasons for its selection show, problems are sufficiently important to demand solutions. In regard to providing a complete demonstration of the method within the first year's effort, it was recognized that any *immediate* solution developed would bear a lack of mathematical elegance because of its provisional character. Similarly, since the proposed solution was demonstrated on a particular system in specific terms, its generality across a wide range of systems has to be determined in subsequent studies. Finally, an additional constraint was added by existing practicalities; viz., that the technique suggested could be followed by OPEVAL project officers at COMOPTEVFOR individuals working at far less than optimal testing conditions and who do not have advanced training in the social sciences. Thus, in many ways the project resembled an engineering feasibility test: experimental controls had to be sacrificed to the practicalities of field testing; questions concerning cause-effect relationships could only be attacked partially, thereby necessitating a restricted use of the term causality;[1] and the inevitable data losses stemming from operational problems had to be recognized and accepted.

All work was based upon the assumption that a human factors requirement could be defined as any possible mixture of human factors inputs which results in a stated level of system performance. Thus, in most cases an implicit assumption is made that a single dimension of systems' effectiveness exists (the firing rate measure evolved earlier) and that the desired level

---

[1] The criteria for using the term causality or cause, such as A *causes* B, was not coincident with the one proposed by Schlict (1931) which states, in essence, that if we can analytically express B as a function of A, it is proper to say that A causes B. The trouble with this argument is that A can similarly be written as a function of B; therefore, most investigators strive for additional criteria. Chief among those demanded by the writers are that A occurs before B in time and that the plausibility of A causing B is enhanced by the sum of available empirical evidence. As to the exact level of "enhancement" required, the decision rule of Bayes theoretically solves the problem; but, unless the sum of evidence available can be converted to an *a posteriori* probability figure, it is impossible to express a standard of acceptance even in terms of prior "belief" ratio necessary to accept one hypothesis over another. For these reasons the use of the term causality was admittedly based upon subjective estimates of the writers based upon their knowledge of applicable literature.

of performance upon this dimension could be stated by Navy system planners. The way that the numerous possible mixtures mentioned above come into play is that this single dimension of system performance is influenced by numerous intermediate criteria as shown in Figure 3 (e.g., system down time, weapon accuracy and burst radius, etc.).

At the beginning of the project it was tentatively assumed that a complete solution of the analytical problem was possible, i.e., that *all* relationships could be established in continuous fashion and combined within an empirical weighting scheme such as that obtained in multiple regression work. In accordance with this assumption the model shown in technical Appendix A was developed. However, it soon became apparent that at least three practical contingencies precluded such an approach from being successful during the time period of this contract: First, practical constraints within normal OPEVALs made it impossible in most cases to manipulate independently all possible human factors considerations that might effect performance, largely because the project officer has limited time in which to conduct the test, a large number of demands to satisfy, and a small sample of humans with which to work (a single test ship crew for example), as well as a single equipment configuration. Second, the accessibility of other sources of operational performance data was limited to maintenance records since only meager amounts of operator performance data were available. Finally, even when the maintenance records themselves were combined with personnel data available from the Bureau of Naval Personnel, many gaps were noted (see Appendix B for a complete discussion of the data synthesis problem).

When confronted with such problems the writers tried to follow the approach described by John W. Tukey.

> The most important maxim for data analysis to heed, and one which many statisticians seem to have shunned, is this: "Far better an approximate answer to the *right* question, which is often vague, than an *exact* answer to the wrong question, which can always be made precise." Data analysis must progress by approximate answers, at best, since its knowledge of what the problem really is will at best be approximate. It would be a mistake not to face up to this fact, for by denying it, we would deny ourselves the use of a great body of approximate knowledge, as well as failing to maintain alertness to the possible importance in each particular instance of particular ways in which our knowledge is incomplete (Tukey, 1962). [2]

2 This same quotation was cited for similar reasons by Banks and Textor, (1963), p. 7.

In view of this orientation it was decided that whatever proposed solution was offered, it would have had to demonstrate its capacity to work with *real* data in a *real* situation. This point, it appears, is sufficiently important to overshadow problems accompanying losses of elegance and generality. In addition, the technique developed during the year of study is one which provides at least a tentative answer to the type of questions likely to be posed by systems planners, regarding the importance of considering human factors suggestions and plans. Finally, the proposed solution is one which focuses upon *actual* factors found to be present (the available sample space) instead of factors likely to be present in operational situations (the hypothetical sample space). In this manner it is less likely that large amounts of time will be spent studying the effects of variables which cannot be manipulated under operational conditions or a range of manipulation along single variables which would be precluded in actual practice.

1.3.3 Amended Method.

Specific problems encountered in demonstrating the method described earlier stemmed from a combination of its breadth of scope and difficulties and limitations in field data gathering. To begin, it was recognized that most OPEVAL's simply cannot gather - within the span of a single test - the amount of data required. Thus, instead of relating equipment performance to human factors variables in precise fashion by relating each pair of variables independently and then determining the thousands of possible variable combinations available, it was decided that an appropriate first step would be to define the upper and lower limits of system performance variation which could be accurately attributed to human factors considerations. If a series of OPEVAL's took this step, their data could be combined by Bureau of Naval Personnel specialists in a manner which would overcome most of the problems of sampling met in a single test.

With this modification in mind, it became necessary to do the following: (1) stipulate a means of accomplishing this first step (which was solved by publishing a separate guidebook for OPEVAL's); (2) supplement the data base by the use of other field performance data available in failure reports; and (3) modify the basic analytic model developed so that it would be used under present as well as future conditions, when functional relationships were more precisely established.

In using the field data available, it was necessary to accept what was there and use it to best advantage. Therefore, only some of the human factors considerations could be studied, and, because of numerous inadequacies in the amount and form of the data (see Appendix B), it was necessary to change the original goals of obtaining empirical multiple regression equations for the human factors that could be studied, to a demonstration of range of human effect similar to that advocated for OPTEVFOR OPEVAL's. This does not mean that such work cannot be done in the future; only a few simple modifications in the present procedure are needed before the goal becomes attainable. In fact, the reason that the regression equations were omitted is not that they could not be computed, but because of present data inadequacies, such computations would have been a fruitless endeavor open to valid criticisms concerning whether or not anything of practical concern was demonstrated.

2.0 DATA ANALYSIS TECHNIQUES

In the preceding chapter the proposed method was discussed at a fundamental level, without giving concrete details concerning how much of the method was tested during the one-year duration of this contract. This chapter explains by means of illustrations and nonsymbolic treatments how the basic model works and what uses stem from its present form. Individuals interested in the symbolic development, again coupled with illustrative data, are referred to Appendix A. After the discussion of the model, a segment is included which covers, in general, results of the analyses performed on available data. Again, details concerning the data format and methods of analysis have been separated in technical Appendix B.

2.1 THE BASIC MODEL

The analytical model, which is discussed at greater length in Appendix A, is aimed at representing both operator and maintenance technician effects in terms of a single set of systems performance units, viz., the mean rate (or the change in mean rate) of sustained fire for a surface-to-air missile (SAM) system. The operator effect can be taken into account by relating numerous predictor variables to mean times taken to perform a task. If the predictor variables relate in such a way as to produce an extension of the mean time for task performance, the task performance rate in consecutive firing cycles decreases. By taking account of the different times required to perform each task in a chain or sequence of consecutive tasks, a rationale develops for weighting the potential importance or susceptibility of such a task to improvement by human factors changes.

Consider, for example, that three tasks have to be performed in sequence in order to produce the output of a hypothetical system. Presume that the entire sequence of the tasks takes 80 seconds on the average with the mean of tasks A, B, and C being 40, 30, and 10 seconds respectively. If our hypothetical system has no or only nominal further time delays due to equipment effects, the entire 80 seconds is "human" time.

As for the mean time values taken to perform each task, two facts are evident. first, variance around the mean can be expected from person to person, equipment configuration to equipment configuration, etc., second, if percentages of improvement and decrement are taken so that they are equal in absolute value and symmetrical around the mean, it is obvious that each percentage increment will be greatest in absolute time value for task A and least for task C. Thus, if change factors from 10% to 60% are proposed, which can be expressed as the proportions ± .1 through ± .6, a table of values can be readily computed such as the one in Table 2.1.

In looking at the table it is worthwhile to note that if the above and below average performances are compared in terms of an improvement ratio, e.g., in task A for a change factor of ± .6 if only extreme cases are considered, the potential improvement ratio from using the "best" as opposed to the "worst" operators is 64/16 or 4/1. Similarly, ± .5 shows a 3/1 ratio, etc. If, however, the improvement ratio is computed in terms of the best performance (the shortest time) as compared to the *average* performance for task A (40 sec.), the change factor ± .5 shows an improvement ratio of 40/20 = 2/1.

Ratios such as those computed above are often cited in OPEVAL data; e.g., when an experienced operator was used, performance improved by a factor of 5/1.* Keeping in mind the decision to use a notion of human variance limits in OPEVAL, the value of such an approach begins to be seen.

The change factors shown in Table 2.1 are task specific. What is needed is to change these into system performance terms. If we consider using only above average operators at each task position in turn, the percentage improvements in the simplified table 2.2 result. If comparisons are made between better than average and worse than average operators, the percentages increase in value.

So far this explanation has been a simple arithmetic exercise showing some of the implications of improvement factors or human "tolerance limit" notion. Before the picture gets complicated by having to account for numerous practicalities, it is best to indicate the practical value of the steps taken so far. The primary advantage stems from a rather clear illustration of the definitions of a human requirement which was offered in Chapter 1.

---

*Such improvement was noted in WEXVAL tests. What is particularly striking is that these improvement ratios were in *system* performance terms *not* task performance terms.

PERFORMANCE RATINGS

Symmetrical Change Factor	TASK A		TASK B		TASK C	
	Above Aver.	Below Aver.	Above Aver.	Below Aver.	Above Aver.	Below Aver.
$\pm .1$	36	44	27	33	9	11
$\pm .2$	32	48	24	36	8	12
$\pm .3$	28	52	21	39	7	13
$\pm .4$	24	56	18	42	6	14
$\pm .5$	20	60	15	45	5	15
$\pm .6$	16	64	12	48	4	16

TABLE 2.1

Hypothetical Performance Scores for Each of Three Sequentially Executed Tasks. (Table entries are in terms of performance time in seconds.)

Change Factor	Task A	Task B	Task C
± .1	5.00%	3.75%	1.25%
± .4	20.00%	15.00%	5.00%
± .6	30.00%	22.50%	7.50%

TABLE 2.2

Illustration of Percentage of Improvement in Total Task Time by Using Better Than Average Operators Instead of Average Operators

Assume, for example, that a systems planner states flatly that tactical conditions necessitate a 70 sec. mean response time for the system (a 12.5% improvement). Such a requirement can be met in a variety of ways: by improving performance on task A slightly more than 30%; by improving performance on task B by slightly less than 40%; and by any number of possible combinations of improvement on all three tasks simultaneously.

At this point in the development the approach is extremely simple to understand. Complications begin to arise rapidly, however, beginning with the need to take into account the maintenance effect. Briefly, a connection is possible because the operator effect appears only when the system is operative and has no chance to appear when the system is inoperative. In other words, we must weigh effects such as those shown in Tables 2.1 and 2.2 by the proportion of time a system is "up." The remaining time, of course, is when the system is "down."

In order to express both the influence of the maintenance technician who causes the system to be "up" more often if he is better than average, and the influence of the better than average operator (remember that task changes can make him better than average, as well as training, experience, and aptitude) in the same terms, we must abandon the total service time criteria used in Table 2.2. One way of circumventing the problem is to consider the .80 sec. time as a mean time to perform a "service" or produce an output. Thus, if a fixed interval of time is taken, say 1000 minutes, and it is assumed that average operators are used and the equipment is always "up," the service rate is 1 output every 1.33 minutes. If, however, portions of the system are "down" for maintenance, the service rate must decrease. If the "down time" is 20% for the system instead of producing 1000/1.33 outputs, only 800/1.33 outputs can be produced leading to a diminished servicing rate *for that fixed interval of time*. Thus, an obvious "tradeoff" possibility begins to emerge when we can use better than average technicians or better than average operators in various combinations at different points in the system to reach a desired output rate.

The maintenance technician problem can be made more analogous to operational conditions if we consider that the hypothetical system can produce an output when only two of the three subsystems are "up." This condition can be met in the following ways by letting a "1" indicate an "up" state for subsystems A through C and a "0" indicate a "down" state.

Subsystems State

State	A	B	C	System State
j = 1	1	1	0	UP
j = 2	1	0	1	UP
j = 3	0	1	1	UP
j = 4	1	1	1	UP

The system is "down" whenever any of the following occur:

Subsystems State

State	A	B	C	System State
j = 5	1	0	0	DOWN
j = 6	0	1	0	DOWN
j = 7	0	0	1	DOWN
j = 8	0	0	0	DOWN

If it is assumed that failures are independent of each other in each of the three subsystems--which is quite plausible--it is possible to compute the probability of equipment being "up," that is, in any of the first four states (j = 1 through j = 4), or "down" (j = 5 through j = 8). This result is accomplished by inserting the appropriate probability of a failure where each 0 appears, and the complementary probability of a subsystem being "up" when a 1 appears. By summing the cross-products for states j = 1 through j = 4, the total "up" time probability can be computed. Also, an indication can be gained of how much opportunity arises for operator effects in each equipment state.

At this point in the discussion the following points have been made: (1) both operator and technician effects can be expressed in terms of the same system performance criterion, (2) a system requirement can therefore be restated in terms of numerous mixes of human improvements that will yield such performance;[1] (3) the size of the net effect on system performance by operator improvements varies as a function of the proportion of total task time taken by each task, while in the case of the maintenance technician, a similar function results from differences in equipment failure rates (and probabilities); and (4) all requirements at this stage can be stated in terms of percentage of improvement needed at various points in the subsystem by either operators and/or maintenance technicians. Thus, whenever a human factors specialist is given a system performance requirement in precise terms, he can immediately translate this into human performance changes of fixed magnitude.

Even at this stage of development, something of value has been gained because a great deal of psychological literature relates variations in task structure, training, etc., to intermediate criterion such as time to perform a task. Thus, given a few simple data from systems people, the human factors specialist can translate the problems into terms where available literature is of immediate value. Admittedly, from this point, the specialist must guess as to what system performance change is likely to be obtained in each practical situation if such procedures are followed, but he can make his estimates conservative and thereby have a reasonably high degree of assurance that the stated requirements can be met.

What remains, of course, is to relate *specific* human factors variables to *specific* improvement factors for *particular* systems. Then, the specialist can show that an x% improvement in system performance can result from following plan of action y. The plan of action would be to select on the basis of a weighted prediction scheme which takes into account all of the factors that can be manipulated and their interactions as well. Since, as was stated before

[1] This simply means that if y = f(x), it is possible to write x = f(y).

(and is substantiated in detail in Appendix B), this more complete solution was not possible using available data, an alternative was used. This alternative was to identify *specific* variables which were found on the basis of an empirical analysis to yield *specific* improvement factors and to then project on the basis of knowledge of TARTAR failure rates, etc., how much these variables influence system performance. Since only maintenance technician data were available, only that effect was studied.

2.2 FAILURE REPORT ANALYSES

The aim of the failure report data analysis was twofold: first, to determine which members of a large set of potential human variables gave evidence of being actually related to the intermediate criterion of total system down time; and second, to convert these down time effects into estimates of effects on total system performance. At the same time this data gathering effort was initiated, an attempt was made to gather data aboard a test ship about the human operator effect. Unfortunately, the time spent aboard ship did not yield the data needed because the test runs necessary for data collection were cancelled. The only tangible gain from this shipboard experience was that it proved to be possible to collect such data using automatic recorders (provided the test runs are completed, of course) and without interfering with the ongoing OPEVAL. Therefore, all data obtained concerned the maintenance technician's performance.

By circumventing a number of problems, it was possible to compute correlations between the personnel variables capable of investigation and a number of performance scores. Since data were not available on an individual score basis, that is, there was no way of telling who worked on what trouble, the number of samples available was based upon the number of maintenance teams working on the same type of equipment. Since there is only one team working on each type of equipment per ship, and the available data were beset with omissions, the total available N was very small ($N \leq 15$). For this reason computation of regression equations was omitted and only Spearman rank order r's were used.

Because of the necessity to work with team data, human variable scores were computed in four different ways: first, by taking the mean score of the group on that variable (e.g., GCT); by taking the size of the variance of scores on each variable for each group; by taking the score of the lead man of each group; and by taking the difference between scores of the lead man and the mean of scores for the rest of the team. The reason for these latter three scores being used was an attempt to tease out any team effects as such.

Correlations were computed for the fire control radar group only, since there were not sufficient data to merit analyses of other equipment groups ($N \leq 6$). Performance scores were obtained in terms of (1) the mean number of clock hours of down time per trouble; (2) the average clock hours taken to diagnose a trouble; (3) the average clock hours taken to repair a trouble; (4) the average man hours taken to diagnose a trouble; and (5) the average man hours taken to repair a trouble.

From the above description it can be seen that many correlations were computed on a sample sufficiently small to exclude cross-validation. For this reason the data were taken to be suggestive only. In addition, it was decided to take the additional precaution of noting which variables seemed promising when looking at correlations with average *man* hours for diagnosis and repair, and to test these variables against the mean down time per trouble as measured in *clock* hours. By this technique at least some of the possibility of criterion contamination was eliminated.

Results of the analysis which are presented in detail in Appendix B and an accompanying classified supplement were as follows:

(1) There appears to be little correlational value in using any other than mean scores for each group on the prediction variables.

(2) Only four of the ten predictor variables had significant correlations ($P < .05$) in both screening analyses (mean man hours to diagnose and repair a trouble). These were as follows:

(x_1) Mechanical Aptitude,

(x_2) Clerical Aptitude,

(x_3) The number of months since the most relevant training school, and

(x_4) a composite "relevant training school score" explained in Appendix B.

These four variables were then examined in terms of their effect on the mean number of clock hours per trouble in the following way. The top three teams with respect to each predictor variable and the bottom three teams had their mean clock hours per trouble performance scores averaged and expressed as an "advantage ratio; " i. e., if one selected persons high on that predictor score instead of low on that score. In three of the four cases computed, these ratios were sizeable; i. e., 4.8/1 for x_3, 3.7/1 for x_1, 2.7/1 for x_2, but, in one case, only a small ratio was noted, 1.3/1, for x_4.

In order to establish what these ratios meant in terms of systems performance, these ratios were combined with failure frequency probabilities for that subsystem, gathered from OPTEVFOR classified data, to predict what each of these meant in terms of system performance improvement if individuals high on these scores were used instead of individuals low on these scores. The exact results obtained are contained in a classified supplement, but it suffices to say that their order of magnitude, in terms of percent efficiency gained, was sufficiently great to leave little doubt as to the potential importance system planners could gain by exploiting these relationships.

APPENDIX A: THE PREDICTION MODEL

The general purpose of this project is to quantitatively assess the contributions of human factors (variations in human performance) to the performance of naval systems. This quantification must be effected in terms of units which are common to those of other contributing factors; e.g., firepower, kill probabilities, survival probabilities, overall dollar cost, etc. A necessary condition for this is that the different types of human factors contributions be expressible in the same units. As a feasibility exercise, therefore, we have defined two types of human factors contributions to a naval system, namely, operation and maintenance, and attempted to derive common units for their assessment.

The following example involves a hypothetical weapons defense system whose firing rate, R, is a partial function of operator and maintenance variables as follows:

A. OPERATOR CONTRIBUTIONS

Consider the following regression equation.

$$(1) \quad R_{jk} = \bar{R}_j + b_{1j}\, x_{1jk} + b_{2j}\, x_{2jk} + \ldots + b_{ij}\, x_{ijk} + \ldots + b_{mj}\, x_{mjk},$$

where

R_{jk} = sustained rate of fire for personnel team k when system hardware is in state j

$\bar{R}_j$ = sustained rate of fire when system hardware is in state j and $\Sigma b_{ij}\, x_{ijk} = 0$ (i.e., when the operator team contribution is at the average level of those studied).

b_{ij} = regression coefficient for the i^{th} personnel position with system at state j.

x_{ijk} performance index obtained for the k^{th} team at the i^{th} position with system hardware in state j. This index will be given in time units, sometimes from direct time measurement and, in cases of other positions, in derived time measurements (from time required to correct position errors, etc.).

A separate regression analysis will be performed with respect to each value of j (state of system hardware). The different regression weights yielded from these analyses for any given position will be averaged according to the probabilities of the system states involved, i.e.,

(2) $\overline{b}_i = \Sigma p_j b_{ij}$

where $\overline{b}_i$ represents the "weighted average regression weight" for position i. This weight will be given in final criterion units (rate of sustained fire).

B. MAINTENANCE PERSONNEL CONTRIBUTIONS

Consider the following equation:

(1) $\overline{R}_{..} = \Sigma p_{jk} R_{jk}$,

where

$\overline{R}_{..}$ = average sustained rate of fire by system

R_{jk} = sustained rate of fire for personnel team k when system hardware is in state j

P_{jk} = joint probability that team k is operating and hardware is in state j

Assuming that choice of team is independent of the state of hardware, Equation (2) now becomes

(2) $\overline{R}_{..} = \Sigma p_j P_k R_{jk}$. Now let $\overline{R}_{j.} = \Sigma P_k R_{jk}$

Thus, $\overline{R}_j$, represents the average rate of fire, over all operator conditions, of the system when its hardware is in state j. Equation (2) now becomes

(3) $\overline{R}_{..} = \Sigma p_j \overline{R}_j$.

Now, let us define the hardware state j. The following matrix assumes a system of seven components, each of which may be "up" or "down" (1 or 0) at any instant. Let us assume that component A is absolutely essential for the system to fire, so that any combination of states in which A = 0 yields a firing rate of zero. Component B is not essential to the operation of the system, but its failure will reduce the firing rate by 40 percent. This is a "series effects" in that it is a constant percentage regardless of the state of the rest of the system.

Component E also exerts a "serious effect," but of a different nature, in that operability of E is necessary if the system is to achieve more than 25% of its maximum firing rate. It has no necessary function when the system is so degraded in other respects that it cannot operate at greater than 25% of its maximum rate, but its failure absolutely limits the system to the 25% figure.

Components C and D are a parallel set of identical elements, as are components F and G. The system cannot function with both C and D down; it cannot function with both F and G down. With one member of the set, C and D, operable, the system can achieve rates up to 50% of maximum, subject to the series effects of A, B, and E. With one member of the set, E and F, operable, the system can achieve rates up to 50% of maximum, subject to the above series effects. It is necessary to have both members of both these pairs operable to achieve anything more than 50%. If both members of both pairs are operable, the maximum rate is 100% subject to the series effects of A, B, and E.

In the foregoing discussion, note that "maximum rate," $\overline{R}_{127}$, refers to that rate achieved by the system when all components are "up" but when operator contributions are averaged. Thus, it is a maximum only with respect to the hardware variables. All other values of $\overline{R}_{j\cdot}$ are expressed as proportions of this maximum value, $\overline{R}_{127}$.

TABLE A.1

Payoff Matrix

J	A	B	C	D	E	F	G	$\overline{R}_{j\cdot}$
0	0	0	0	0	0	0	0	0.00
1	0	0	0	0	0	0	1	0.00
2	0	0	0	0	0	1	0	0.00
3	0	0	0	0	0	1	1	0.00
4	0	0	0	0	1	0	0	0.00
5	0	0	0	0	1	0	1	0.00
.								0.00
.								0.00
.								0.00
63	0	1	1	1	1	1	1	0.00
64	1	0	0	0	0	0	0	0.00
65	1	0	0	0	0	0	1	0.00
66	1	0	0	0	0	1	0	0.00
67	1	0	0	0	0	1	1	0.00
68	1	0	0	0	1	0	0	0.00
69	1	0	0	0	1	0	1	0.00
70	1	0	0	0	1	1	0	0.00
71	1	0	0	0	1	1	1	0.00
72	1	0	0	1	0	0	0	0.00
73	1	0	0	1	0	0	1	0.15
74	1	0	0	1	0	1	0	0.15
75	1	0	0	1	0	1	1	0.15
76	1	0	0	1	1	0	0	0.00
77	1	0	0	1	1	0	1	0.30
78	1	0	0	1	1	1	0	0.30
79	1	0	0	1	1	1	1	0.30
80	1	0	1	0	0	0	0	0.00
81	1	0	1	0	0	0	1	0.15
82	1	0	1	0	0	1	0	0.15
83	1	0	1	0	0	1	1	0.15
84	1	0	1	0	1	0	0	0.00
85	1	0	1	0	1	0	1	0.30

	A	B	C	D	E	F	G	$\overline{R}_{j\cdot}$
86	1	0	1	0	1	1	0	0.30
87	1	0	1	0	1	1	1	0.30
88	1	0	1	1	0	0	0	0.00
89	1	0	1	1	0	0	1	0.15
90	1	0	1	1	0	1	0	0.15
91	1	0	1	1	0	1	1	0.15
92	1	0	1	1	1	0	0	0.00
93	1	0	1	1	1	0	1	0.30
94	1	0	1	1	1	1	0	0.30
95	1	0	1	1	1	1	1	0.60
96	1	1	0	0	0	0	0	0.00
97	1	1	0	0	0	0	1	0.00
98	1	1	0	0	0	1	0	0.00
99	1	1	0	0	0	1	1	0.00
100	1	1	0	0	1	0	0	0.00
101	1	1	0	0	1	0	1	0.00
102	1	1	0	0	1	1	0	0.00
103	1	1	0	0	1	1	1	0.00
104	1	1	0	1	0	0	0	0.00
105	1	1	0	1	0	0	1	0.25
106	1	1	0	1	0	1	0	0.25
107	1	1	0	1	0	1	1	0.25
108	1	1	0	1	1	0	0	0.00
109	1	1	0	1	1	0	1	0.50
110	1	1	0	1	1	1	0	0.50
111	1	1	0	1	1	1	1	0.50
112	1	1	1	0	0	0	0	0.00
113	1	1	1	0	0	0	1	0.25
114	1	1	1	0	0	1	0	0.25
115	1	1	1	0	0	1	1	0.25
116	1	1	1	0	1	0	0	0.00
117	1	1	1	0	1	0	1	0.50
118	1	1	1	0	1	1	0	0.50
119	1	1	1	0	1	1	1	0.50
120	1	1	1	1	0	0	0	0.00

	A	B	C	D	E	F	G	$\bar{R}_j$
121	1	1	1	1	0	0	1	0.25
122	1	1	1	1	0	1	0	0.25
123	1	1	1	1	0	1	1	0.25
124	1	1	1	1	1	0	0	0.00
125	1	1	1	1	1	0	1	0.50
126	1	1	1	1	1	1	0	0.50
127	1	1	1	1	1	1	1	1.00

In the present example, it was assumed that the interactions among the various system components were orderly and completely known. This state of affairs does not, of course, always exist. In order for the present scheme to be applicable with high generality, it will be necessary in many cases to resort to empirical analyses, such as multiple regression studies of fire-power vs. operability of the various components.

In order to evaluate the role of differences in maintenance efficiency at the various positions, it is necessary to supply the following additional data.

1. The expected (average) proportions of "down time" of various equipment across all maintenance conditions.

2. The variation in each of these average values that would be expected on the basis of differences in experience on the part of maintenance personnel.

The average proportions of "down time" depicted in Table 2 do not correspond to any particular military system, although they are perhaps not too unrealistic. By use of the simplifying assumption that the "down time" is completely accounted for by corrective maintenance (CM) time,* it is possible to compute adjusted values based on times for experienced (rated) personnel and inexperienced (nonrated) personnel to perform a sample of CM activities, as obtained by McKendry, Corso, and Grant (1960). For the present illustration, this 13/22 ratio was applied to the data in Table 2 to produce symmetrical upward and downward deviations about the postulated averages. Thus, average time for CM by the more experienced technician was estimated to be about 26% lower than the average value; average time by a less experienced technician as 26% higher. This is the only adjustment made on these CM values in the current example, although subsequent real data efforts would doubtlessly incorporate other predictors such as aptitude or job sample scores, amount of experience with particular items of equipment, etc.

---

* Frequencies of failure are treated as constant, regardless of CM procedures. Eventually they can be treated in relation to preventive maintenance skills.

TABLE A.2

Expected Proportion of Time Position X is Operable When Maintained as Indicated

Component	Rated	Nonrated	Average
A	.837	.723	.780
B	.911	.849	.880
C	.555	.345	.400
D	.555	.345	.400
E	.992	.987	.990
F	.882	.798	.840
G	.882	.798	.840

$$P(X = 0) = 1 - P(X = 1)$$

Utilizing the Table 2 data, it is possible to calculate the effect of interest: How closely will $\overline{R}_{..}$ approach $\overline{R}_{127.}$ as a function of varying the experience of maintenance personnel at each system position? Or, alternatively, how much difference does it make when maintenance is performed by a rated, as opposed to a nonrated, man at any given position? The rational equation for this effect is

$$(4) \quad \Delta R_x = R/X_r - R/X_n$$

where X_r, X_n refer to maintenace (given) by average rated and nonrated men, respectively, and R/X_r, R/X_n refer to the corresponding expected averages, $\overline{R}_{..}$ for the system, assuming that other positions than Position X have average maintenance. Equation (5) depends on the following relationship:

$$(5) \quad \Delta R_x = \Sigma P_j R_j/(X = 1)\,[P(X = 1)/X_r - P(X = 1)/X_n] +$$

$$\Sigma P_j R_j/(X = 0)\,[P(X = 0)/X_r \quad P(X = 0)/X_n]$$

The state j is defined as in Table 2; P_j is the probability of state j; R_j is the corresponding rate of fire when the system is in state j. $P(X = 1)/X_r$ is the probability that position X is up given that this position is maintained by a rated man; $P(X = 0)/X_r$ is the probability that position X is "down" if maintained by a rated man. $P(X = 1)/X_n$ and $P(X = 0)/X_n$ are the corresponding terms when maintenance is by nonrated men. For the non-bracketed terms, the R_j values can be read directly from Table 1. In the bracketed terms, computation of the corresponding P_j for each R_j, given that X = 0 or X = 1, is done by inserting the *average* reliability figures from Table 2; e.g., if j = 85(1010101, binary) then for the case that X = A, $[P_j/A = 1] = P(B = 0) \cdot P(C = 1) \cdot P(D = 0) \cdot P(E = 1) \cdot P(F = 0) \cdot P(G = 1)$. The value of R_j for j = 85 can be read directly from Table 1. Thus, to obtain the cross-product sum,

$$\Sigma P_j R_j/(A = 1),$$

it is necessary to sum the 64 cross-product terms ($j = 64$ thru 127) in which the $A = 1$ entry appears. This gives $\Sigma P_j R_j/(A = 1)$, which when multiplied by the first bracketed term in (5) yields the expected difference in payoffs (for A_r vs. A_n) due to differential likelihood of an operable A position. The second difference,

$$\Sigma P_j R_j/(X = 0)\ [P(X = 0)/X_r - P(X = 0)/X_n],$$

gives the expected difference for A_r vs. A_n due to differential likelihood of a *non*-operable A position. Obviously, this last difference will be zero or negative. (Examination of Table 1 will indicate that $\Sigma P_j R_j/(A = 0) = 0.0$, since $R_j = 0$ in any row in which $A = 0$. This reflects the circumstance that operability of Position A is an absolute necessity for the system to function.

C. RESULTS

Table 3 presents mean firing rate, $\overline{R}_{..}$, expressed as a proportion of $\overline{R}_{127}$. Values for the system are given as a function of maintenance technician skill at each position separately. Thus, for example, it can be seen that changing from low to high skill maintenance at Position C or D can affect system performance by 23.9%, compared to a change of only 0.3% at Position E.

TABLE A. 3

Mean Firing Rate as a Function of Maintenance Skill

Maintenance Skill Level	A	B	C	D	E	F	G
High	.292	.276	.304	.304	.273	.276	.276
Average	.272	.272	.272	.272	.272	.272	.272
Low	.252	.269	.239	.239	.271	.268	.268
% Change	14.6%	2.6%	23.9%	23.9%	0.3%	2.7%	2.7%

APPENDIX B. MAINTENANCE TECHNICIAN PERSONNEL DATA AND TARTAR FAILURE REPORT DATA

A. INTRODUCTION

In the development of any truly comprehensive scheme to assess human factors contributions to systems performance, it is necessary to take cognizance of the potential gains to be realized from selection and training. The main body of the effort in other phases of this work has been to relate human performance to system performance; here the concern is with construction of the inferential bridges in the opposite direction, attempting to isolate factors that are related to human performance. The approach taken has been to utilize existing data pool , an effort which, if successful, would effect considerable savings in both time and expense. One area that seems to indicate some promise of useful returns concerns the maintenance technician. It is possible, in theory at least, to obtain actual on-the-job performance scores for technicians. The source of these data is the Bureau of Weapons Form 8000/13 which provides detailed information concerning equipment failures and the time required to locate and correct them. These forms are completed following each failure which resulted in rendering the equipment inoperable. Compilations are available from the Quality Evaluation Laboratory, Concord, California. The information called for on the form is quite complete, giving many particulars concerning equipment status before and after the maintenance activity, the location of the failure, and its symptoms. Of greatest relevance to this effort, however, is the reporting of various times involved in the maintenance activity since these have been shown in other studies to be reliable indices of the adequacy of technician performance. Specifically, it is possible to obtain from the Form 8000/13, diagnosis and repair times in clock hours and man hours and total equipment "down time" in clock hours. For the purposes of this study there is one shortcoming inherent in these data; it is impossible to associate directly a particular maintenance event with an individual technician. The forms, when submitted, do bear a signature but no record of it is kept. As a matter of fact, it would be of no particular value if it were to be recorded because the individual completing the repair is not necessarily the person who fills out the report. A reasonable method of circumventing this problem was devised by the simple expedient of assuming that, averaged over a sufficient period of time, the performance of the maintenance crew assigned to a given equipment group would be an accurate index of the net effectiveness of the various group members.

B. PROCEDURE

It was necessary to determine the composition of the various groups and to obtain data concerning the personnel composing the groups. This was accomplished by requesting a computer run from the Bureau of Personnel, Code 19, to extract the information for all TARTAR maintenance personnel. Not all of the personnel data were relevant, but of that portion that was relevant, not all of the information was available for all of the individuals in the specified study population. In addition to the screening required at this point, it was also necessary to recode some of the data in order to facilitate later computations. The transformations of the data were simple manipulations designed to produce ordinal scales or to combine certain information to yield indices more appropriate for the purpose at hand. The final selection of factors consisted of the following:

(1) Rate (RATE)

(2) Years of education (Y. Ed.)

(3) General classification test scores (GCT)

(4) Arithmetic test scores (ARI)

(5) Mechanical test scores (MECH)

(6) Clerical test scores (CLER)

(7) Electronic technician selection test scores (ETST)

(8) Months in service (MIS)

(9) Months elapsed since most relevant training (MST)

(10) A composite index reflecting amount of training in weeks as a function of relevance of the training (TRI)

A word of explanation concerning judgments of training relevance; based on the PNEC of the individual, an arbitrary weight of 4 was assigned to the C school that leads directly to that PNEC, the A school preparation required for that C school received a weight of 3; schools training for related PNEC's received a weight of 2; and leadership training and other possibly relevant short schools were assigned a weight of 1. To compute the Training Relevance Index (#10 above) the simple sum of products covering all of the individual's training

TABLE B.1 SUMMARY OF 24 ORIGINAL PLANNED CORRELATION MATRICES (EACH MATRIX CONTAINING 50 CORRELATIONS RHO S OF PERSONNEL VARIABLES WITH MAINTENANCE PERFORMANCE INDICES)

	MEAN							VARIANCE	LEAD MAN	DIFFERENCE
			FAILURE INDICES							
			A	B	C	D	E			
SEARCH RADAR	PERSONNEL VARIABLES	1 ... 10								
WEAPONS DIRECTION EQUIP										
COMPUTER										
FIRE CONTROL RADAR										
MISSILE LAUNCH SYSTEM										
MISSILE TEST EQUIP.										

was taken, where each product was the multiple of the length of the school in weeks times its relevancy weight.

These background factors for individuals were then combined into indices of team performance in four ways. First, and most straightforwardly, by taking the arithemetic mean; second, because it is logical to conceive of the amount of homogeneity of the group affecting the overall quality of performance, the variance for each team on each factor was computed. The third and fourth methods involved the notion that the lead man can have, potentially at least, a great influence on the performance of the group. Therefore, scores for the lead man alone were employed as method 3, while method 4 was the difference score obtained by comparing the lead man score with the mean of the remainder of the group.

A correlation coefficient could be calculated for each of the 5 failure report indices with each of the 10 personnel indices for a total of 50 intercorrelations in each matrix. Because each main equipment group was handled separately and there were 4 different ways of computing the personnel indices, the original design called for a series of 24 matrices as shown in Table B. 1

C. RESULTS AND DISCUSSION

The reporting of the results of this investigation cannot be made without first providing certain clarifying information that influence the outcome. Since there were essentially two data pools which the investigation sought to merge and demonstrate the relations between, it appears appropriate here to treat each set separately.

Personnel record data were obtained for 694 technicians who were assigned to TARTAR ships during the study period of 1 September 1963 to 30 November 1963. Upon reduction of these data and computation of the personnel indices, it was found that the two indices which were intended to be sensitive to the effect of the team leader had to be discarded due to lack of complete information on the lead man. This deficiency is thought to be due primarily to the fact that most of this class of personnel entered the Navy prior to the introduction of many of the aptitude tests which were of concern in this investigation.

The failure report data introduced further shrinkage from the original plan. During the three-month study period a total of 860 troubles, spread over five equipment groups, were reported by 18 ships. This did not include the Air Search Radar equipments since these equipments are under the cognizance of the Bureau of Ships and are, therefore, not reported on NAVWEPS Form 8000/13. The distribution of failures reported was such that only 38 out of the 90 cells in the matrix contained the required six or more failures. The criterion of a minimum of six troubles was introduced in order to reduce the effects of chance fluctuations. Of the 38 cells which met the criterion, 14 were in the Fire Control Radar column. The remaining 24 were distributed approximately equally over three of the other four equipment groups, the launcher group having no ships reporting 6 or more troubles. Because the number of degrees of freedom available is determined by the number of cells meeting the criterion, only the Fire Control Radar group was judged to be adequate for further analysis. Table B. 2 summarizes the net effect of the various causes of data shrinkage.

The final results of the correlational analysis were not as definitive as desired. This, however, was anticipated since the best that could rationally be expected from this type of an investigation was to identify promising factors for future follow-up study and refinement. Even if very high positive correlations had been found in specific instances, it would have been necessary to conduct a cross-validation study to verify their true existence. An examination of Table B. 3 indicates that the personnel factors that show potential usefulness as predictors of performance are mechanical (MECH), clerical (CLER), months since most relevant training (MST) and training relevance index (TRI). This statement must be advanced tentatively, however, since it is clear that not only is it based on a small number of data points, in terms of personnel and failure reports, but it is subject to range restriction since the subject population is already a select group; this restriction usually results in spuriously low correlations.

What then can be said to be the value of this attempt to investigate the relationships between personnel factors and performance data? It is, first of all, a demonstration, albeit a tentative one, that it is possible to use existing data pools in such investigations. The fact that the attempt to tease out these relationships, given the crude form of the data, was successful at all is significant. As was demonstrated in Chapter II, the model advanced here can employ this type of result successfully in determining technician effects on systems performance.

A further valuable result can be expressed in terms of indications of how to go about supplementing existing data bases and conducting this kind of investigation. These suggestions are embodied in the following brief outline:

1. Divide the population of ships into two random groups.

2. From both groups of ships select those personnel in each equipment group who will be on board for a period of the next six calendar months.

3. Arrange for duplicates of Form 8000/13 to be filled out *by individual(s) actually conducting repairs* over a period of six months.

4. Obtain personnel record data on the study population and fill in gaps where feasible by interviews and testing (particularly in case of lead man on team).

5. Compute correlations of performance indices with personnel variables, using only data from Group I ships.

6. Enter these correlations into a multiple regression analysis and produce prediction equations.

7. Cross-validate these findings employing data collected on Group II ships.

8. Using those variables that show stable weights construct final prediction equations.

Finally, it should be noted that this investigation provides certain empirical lessons in regard to the difficulties that should be anticipated in employing existing data. For example, the personnel data are also subject to certain unavoidable shortcomings for this type of application. The lack of detailed information regarding the lead man is particularly indicative of the difficulty. Since the main purpose of this data pool is to keep accurate records on all Naval personnel, and, is therefore, restricted by the information available, there is no reason to be particularly concerned with gaps in a given individual's record. Only a very extensive and expensive testing program could fill the gaps--and then only part of them. This would obviously be inappropriate in terms of overall file objectives, and, therefore, subject losses due to missing data will continue to occur in studies of this type.

TABLE B. 2

Summary of Correlation Matrices and Primary Reason for Loss

	Mean	Variance	Lead Man	Difference
Search Radar	No failure report data available	No failure report data available	No failure report data available	No failure report data available
Weapons Direction Equipment	Only 6 ships with sufficient data	Only 6 ships with sufficient data	Only 6 ships with sufficient data	Only 6 ships with sufficient data
Computer	Only 6 ships with sufficient data	Only 6 ships with sufficient data	Only 6 ships with sufficient data	Only 6 ships with sufficient data
Fire Control Radar	Correlations computed; 14 ships with sufficient data*	Variance measure too unstable	Incomplete data on 6 of the personnel variables	Incomplete data on 6 of the personnel variables
Missile Launching System	Only one ship with sufficient data	Only one ship with sufficient data	Only one ship with sufficient data	Only one ship with sufficient data
Missile Test Equipment	Only 4 ships with sufficient data	Only 4 ships with sufficient data	Only 4 ships with sufficient data	Only 4 ships with sufficient data

*NOTE: Only correlation matrix actually computed.

TABLE B. 3

Correlation Matrix - Fire Control Radar
Personnel Factor Group Means by Mean Performance Indices

Performance Factors	DT (MH)	DT(CH)	RT (MH)	RT (CH)	EDT
Rate	.21	.16	.29	.13	.05
Y.Ed.	.11	.15	-.19	-.10	-.36
GCT	-.02	-.01	-.16	-.07	-.23
ARI	-.20	-.15	-.34	-.17	-.46
MECH	.43	.46	.26	.42	.23
CLER	.45	.43	.43	.48	.09
ETST	.25	.28	.16	.29	-.00
MIS	.13	.05	.28	.14	.10
MST	-.53	-.58	-.49	-.63	-.38
TRS	.55	.59	.54	.65	.35

DT(MH) - Diagnosis Time (Man Hours)
DT(CH) - Diagnosis Time (Clock Hours)
RT(MH) - Repair Time (Man Hours)
RT(CH) - Repair Time (Clock Hours)
EDT - Equipment Down Time
Rate - Rate
Y.Ed. - Years of Education
GCT - General Classification Test Score
ARI - Arithmetic Test Score
MECH - Mechanical Test Score
CLER - Clerical Test Score
ETST - Electronics Technician Selection Test Score
MIS - Number of Months in Service as of Dec. 1963
MST - Number of Months Since Most Relevant Training as of Dec. 1963
TRS - Training Relevance Score

Another example is indicative of the present difficulties with the failure reporting system as it is currently implemented in the fleet. This is not to be interpreted as a criticism of the Form 8000/13. The point is an empirical one, not judgmental or evaluative. The variability in numbers of failures reported for the same class of gear over the same period (as summarized in Table 4) seems clearly indicative of wide variability in reporting procedures. This variability can stem from a number of causes, several of which could be at least potentially damaging to the validity of the data. A possible causal factor might be a difference from ship to ship in the operational definition of exactly what constitutes a reportable failure. Another factor may be procedure followed by a particular team in regard to the reporting; if the reports are not completed immediately after the repair, they may be forgotten or some of the pertinent information may be lost. Whatever the causal factor or factors, the end result is a pool of data that is less complete than would be desired.

TABLE B.4

Range of Failures Reported for Period 1 Sept. 1964 to 1 Nov. 1964 by Equipment Group

Air Search Radars	No data
Weapons Direct Equipment	2 - 14 (with 5 ships not reporting)
Computer	2 - 34 (with 2 ships not reporting)
Fire Control Radar	2 - 50 (with 2 ships not reporting)
Launcher	1 - 37 (with 11 ships not reporting)
Missile Test Equipment	1 - 111 (with 3 ships not reporting)

In summary, the main feature of this investigation was the demonstration, although limited, that existing data could potentially be used to discover relationships between personnel factors of training and aptitude and actual field collected performance data. Further, despite the necessary crudity of this first attempt, the results generated were acceptable into the framework of the model to assess Human Factors effects on system performance.

REFERENCES

Banks, Arthur S., and Textor, Robert B. A Cross-Policy Survey. Cambridge, Mass.: The M. I. T. Press, Massachusetts Institue of Technology. 1963

McKendry, J. M., Corso, J. F., and Grant, G. "The Design and Evaluation of Maintenance Packaging Methods for Electronic Equipment," Ergonomics. Vol. 13, No. 3, pp 255-272. 1962.

Siegel, S. Nonparametric Statistics for the Behavioral Sciences. New York, McGraw Hill. 1956.

Tukey, J. W. "The Future of Data Analysis," Annals of Mathematical Statistics. Vol. 33, pp. 13-14. 1962.