2 of 3

AD 603 304

VOLUME II

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

APPLICATION TO OPTEVFOR OPERATIONS

June 1964

Report No. ND 64-68

64 p
k-3.00
mf-0.75

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

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 second volume of a two volume 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.

It is the intent of this volume to provide a guide for the OPTEVFOR project officer, in order that he may incorporate human factors testing in the conduct of the tests and/or evaluations for which he has responsibility. The document is organized around the framework established by the requirements of the project officer's duties as they are presently constituted. The chapter headings reflect the main subdivisions of these duties, i.e., planning and preparation, preparation of the project plan, conducting the test, evaluating the data, and preparing the final report. Where appropriate, separate consideration is given to the various types of tests and evaluations with respect to each of the above categories.

In general, the approach is one that recognizes the constraints of testing in the operational environment and attempts to integrate the human factors testing in a manner such that there will be a minimum of added burden. Within the limits imposed by practical considerations, the suggestions and recommendations advanced within will lead to the acquisition of useful human performance data and, at the same time, reduce the effects of variations in human performance upon the tests of equipment adequacy.

SUMMARY AND CONCLUSIONS

As stated previously the purpose of this report is to provide project officers with a practical, workable guide for incorporating human factors tests in OPEVALs, as well as the other activities conducted by OPTEVFOR; or, in instances where practical contingencies preclude human factors testing per se, to insure that human factors effects do not complicate or confound results of the systems tests. The recommendations and suggested procedures outlined here are designed to take into account the many limitations and constraints imposed by the operational environment of the tests, time available to prepare for and conduct the tests, supporting services obtainable, personnel for operation, maintenance and testing--all of which are normally in short supply. These factors combine with the nature of the questions which must be answered, and the complexity of modern systems, to confront the project officer with a most formidable task. In this context the human factors efforts have been carefully structured to blend in easily with the existing duties.

In the project planning and preparation phase, suggestions are advanced concerning possible sources of information on the systems personnel requirements, operator and technician duties, human contribution to system operation, criticality of operator positions, and potential effects of error and/or delays. Because there is no guaranteed source for information on these topics, suggestions for developing supplementary data are also included (e.g., job analyses, consultation with designers, questionnaires, etc.). Also included under this phase are the objectives of the human factors efforts and some indication of the necessity for anticipating testing needs. In this latter category, topics such as data gathering requirements and utilization of available test personnel are discussed.

Preparing the project plan is a crucial phase in the conduct of an OPEVAL. After discussing the usual requirements in the various sections of the project plan, the human factors control considerations are discussed. Control as used here has to do with attempts to reduce or eliminate the effects of individual variations in aptitude, training, skill, or experience upon the evaluation of the systems operational suitability. In particular, the notion of control is related to the differential criticality of operator positions in order that effort be expended where it will produce the greatest return. Another topic considered in the section focuses upon a sample of the many available types of data recording and timing devices. The use of these devices serves a twofold purpose, it reduces interference with normal operation; and, increases the accuracy and reliability of the data. Methods of conducting activity analyses of both operator and technician tasks are also outlined.

In the chapter dealing with the actual conduct of the tests the major concern is to provide project officers with some cues regarding the implementation of some of the human factors tests. Included are discussions of questionnaire and interviewing techniques, activity analysis data collection, and the problem of omitted or missing data. This latter section is of particular importance since the very nature of an OPEVAL predisposes it to be susceptible to data losses. These losses need not always be detrimental to the experimental design, and the emphasis of the discussion is on appropriate ways of overcoming these problems.

The evaluation of data from an OPEVAL can be divided into four phases, screening, cataloging, reducing and analyzing. The main theme of the chapter concerning evaluation deals with analysis. By means of two numerical examples, one dealing with an operator function, the other with the technician effect, it shows the importance of keeping track of human factors effects. Further, it shows how knowledge of these effects can aid in the analysis of systems performance data. The implications of screening, cataloging and reducing are also covered and their purpose made explicit.

The final chapter deals with the contributions that knowledge of human factors effects can make in the interpretation of results and the formulation of conclusions. The tolerance limit approach can be quite effective in facilitating decisions concerning potential systems suitability since it will permit extrapolation to "best" and "worst" obtainable performance. The questionnaire data can also be of further use in making recommendations for necessary changes and improvements to bring the system up to necessary standards.

RECOMMENDATIONS

The general content of the suggestions contained in this report center around the project officer's goal and the means by which he can circumvent the numerous practical limitations imposed by the OPEVAL situation. Briefly, the goal is to determine empirically the relationship between system performance and numerous human factors--e.g., state and amount of training, particular equipment configurations, personnel aptitude, etc. Once such data are available, it is possible to state with some certainty the personnel demands imposed by the mission that the tested equipment must fulfill.

Based upon the authors' review of the OPEVAL situation and the orientation discussed above, the following conclusions are in order:

(1) The project officer can gather important human factors data during an OPEVAL simply by including in his plans a cognizance of the ways in which human performance can influence system performance. These data are, in turn, of considerable value to him when he is required to answer such questions as "will this equipment work satisfactorily in the fleet with existing personnel or will special training be needed?"

(2) The data themselves take the form of relating personnel characteristics and attributes, such as aptitude scores, experience, etc., to systems performance measures. Ideally, these relationships would take into account the whole range of possible personnel variations and their resultant influence on system performance. In order to reach this ideal state, considerable time must be spent gathering the data during an OPEVAL, numerous maintenance technicians and equipment operators must be studied, and many data gatherings runs must be scheduled. When such data are collected, however, the project officer can state unambiguously just how well the system will perform using different personnel complements, etc.

(3) Often, OPEVALs must be conducted within the context of small samples of data and simply cannot be extended to gather the types of data described above. In such cases it is best for the project officer to reduce his task to workable portions by first concentrating his efforts on those human factors which are most crucial to system performance, and secondly, to study "extreme groups" of personnel so that he knows approximately how much variation in system performance accrues from manipulation of a particular human factor. After reviewing such data the project officer is in a position to call specific attention to key trouble spots in the human factors area which must be carefully watched and/or corrected. (Means of identifying crucial factors and implementing this tolerance limit approach are discussed in detail in the report.)

(4) Even when smaller amounts of data are collected. project officers can expect data losses and certain difficulties inherent in measuring human factors effects. Techniques are available to circumvent these problems, and when used, the project officer can still satisfy his goals despite these adversities

(5) By gathering date of tye type described above, the project officer's final report answers questions of particular interest to his current problem. In addition, when taken in conjuction with other similar project reports, it can contribute valuable knowledge about such human factors problems in general. This accumulated knowledge will allow a reduction in the amount of human factors testing required in each future OPEVAL.

TABLE OF CONTENTS

		Page
ACKNOWLEDGEMENTS		ii
ABSTRACT		iii
SUMMARY AND CONCLUSIONS		iv
1.0	HUMAN FACTORS AND THE OPEVAL PROJECT OFFICER	1
1.1	THE OPEVAL SITUATION	2
1.2	IMPORTANCE OF HUMAN FACTORS IN OPEVAL	3
1.3	EFFECT ON OPEVAL ADEQUACY	6
1.4	OPEVAL CONSTRAINTS ON HUMAN FACTORS TESTING	7
2.0	TESTS OTHER THAN OPEVAL'S	9
2.1	INTRODUCTION	9
2.2	TECHNICAL EVALUATION	9
2.3	CONCURRENT EVALUATION	11
2.4	DEVELOPMENT ASSIST TEST	13
2.5	FLEET RESEARCH INVESTIGATION	14
2.6	FLEET OPERATION INVESTIGATION	14
3.0	INCORPORATING HUMAN FACTORS TESTING IN THE OPERATIONAL TEST AND EVALUATION CYCLE	15
3.1	FORMAT DESCRIPTION	15
3.2	OPEVAL PHASES	15
4.0	OPEVAL ACTIVITIES IN THE PLANNING AND PREPARATION PHASE	17
4.1	ACCUMULATING KNOWLEDGE CONCERNING HUMAN FACTORS CONSIDERATIONS IN THE PREPARATION AND PLANNING PHASE	18
4.2	SPECIFYING HUMAN FACTORS TEST OBJECTIVES	23
4.3	ANTICIPATING NEEDS FOR HUMAN FACTORS EVALUATION	25
4.4	SUMMARY	26

TABLE OF CONTENTS (Cont'd)

		Page
5.0	PREPARATION OF THE PROJECT PLAN	28
5.1	DISCUSSION OF THE TEST PLAN SEGMENTS	28
5.2	HUMAN FACTORS CONTROL CONSIDERATIONS IN THE DEVELOPMENT OF A PROJECT PLAN	30
5.3	RECORDING HUMAN PERFORMANCE DATA	30
5.4	PURPOSES OF HUMAN PERFORMANCE DATA COLLECTION	32
5.5	ACTIVITY ANALYSIS OF MAINTENANCE TECHNICIAN TASKS	33
5.6	ACTIVITY ANALYSES OF OPERATOR TASKS	35
6.0	CONDUCTING THE TEST	37
6.1	PROBLEMS IN QUESTIONNAIRE ADMINISTRATION	37
6.2	PROBLEMS IN ACTIVITY ANALYSIS DATA COLLECTIONS	38
6.3	THE PROBLEM OF OMITTED RESPONSES	40
7.0	EVALUATING OPEVAL DATA	42
7.1	IMPLICATIONS OF THE SCREENING, CATALOGING, AND REDUCING FUNCTIONS UPON DATA ANALYSIS	43
7.2	FIRST NUMERICAL EXAMPLE	44
7.3	SECOND NUMERICAL EXAMPLE	47
7.4	SUMMARY	51
8.0	PREPARING THE FINAL REPORT	52
8.1	USE OF THE TOLERANCE LIMIT NOTION	52
8.2	USE OF QUESTIONNAIRE DATA	54
DISTRIBUTION LIST		55

1.0 HUMAN FACTORS AND THE OPEVAL PROJECT OFFICER

> *Operational Evaluation. An Operational Evaluation is the test and analysis of a weapon system, support system, component or equipment conducted by the Operational Test and Evaluation Force, under service operation conditions insofar as practicable, to determine the ability of a system, component or equipment to meet specified operational performance requirements and/or to establish suitability for service use. When appropriate, an Operational Evaluation may be ordered solely for the development of basic tactical doctrine, training procedures, and requirements for training aids and/or countermeasures.*[1]

Therefore a project officer's primary responsibility is to determine whether or not the system undergoing operational test *can* accomplish its assigned mission under realistic fleet conditions. One important facet of realism to be considered is the "enlisted man" who operates and/or maintains portions of the system undergoing testing. Thus, human problems are a crucial aspect of operational realism and, as such, merit study.

The purpose of this report is to provide project officers with a *practical, workable* guide for incorporating human factors tests in OPEVAL studies as well as the other activities conducted by OPTEVFOR (these are discussed separately in Section 2.0); and to insure as well that human factors effects do not complicate or confound results of the system tests.

For example, failure to control learning, the field of *human factors* itself is concerned with a host of psychological, sociological, cultural, situational (work space and equipment arrangement), and procedural (method of work) considerations which cause human performance to vary in either a positive or a negative manner. Such variations are of considerable interest to project officers because of resultant effects on the performance of the system in question.

---

[1] COMOPTEVFOR Inst. P. 3930, 1c. Vol. 1., 1 April 1961.

Thus, while the project officer may not be able to trace all of the causes of variations in human performance, he is vitally concerned with how much deviation in system output is contributed by human factors. Then, the OPTEVFOR Staff, in conjunction with human factors specialists from the Bureau of Naval Personnel, can derive ways of improving system performance related to human factors.

The importance of interplay between OPTEVFOR project officers and human factors specialists can readily be seen. *In order to achieve a meaningful solution to human factors problems, it is necessary to know not only how individual and equipment characteristics influence human performance, but also to know how human performance, in turn, relates to system performance. Thus, it is of limited value to know how human performance might be influenced in general as well as how systems perform in general, because what results is a set of general guidelines. In order to go beyond this step, what is needed is detailed, concrete information (and data) relating these two links, not in general but in specific fashion; e.g., relate particular aspects of individuals and equipment to specific system performance data of the type readily available to OPTEVFOR but usually unavailable to anyone else.* When this linkage is complete at a practical everyday working level, OPTEVFOR personnel can bring human factors aspects of their OPEVALs into tighter and tighter control and make more definitive statements of just what human resources are needed to make a system succeed in an operational environment.

1.1 THE OPEVAL SITUATION

In an OPEVAL, project officers are required ultimately to make a "yes-no" judgment concerning whether or not a system undergoing test will actually work in the Fleet. Further, because of the high costs of equipment development before operational testing, OPTEVFOR personnel are also required to point out the nature and extent of system revisions necessary to bring a system from the "no" to the "yes" category. This latter effort of development assistance occurs as a result of two practicalities: first, the Navy wishes to recover as much of the development costs as possible from any system--these costs are staggering in modern weaponry; second, since OPTEVFOR project officers base their "yes-no" judgment upon a body of data and information concerning the system, they are in a position to offer to bring a system up to operational standards.

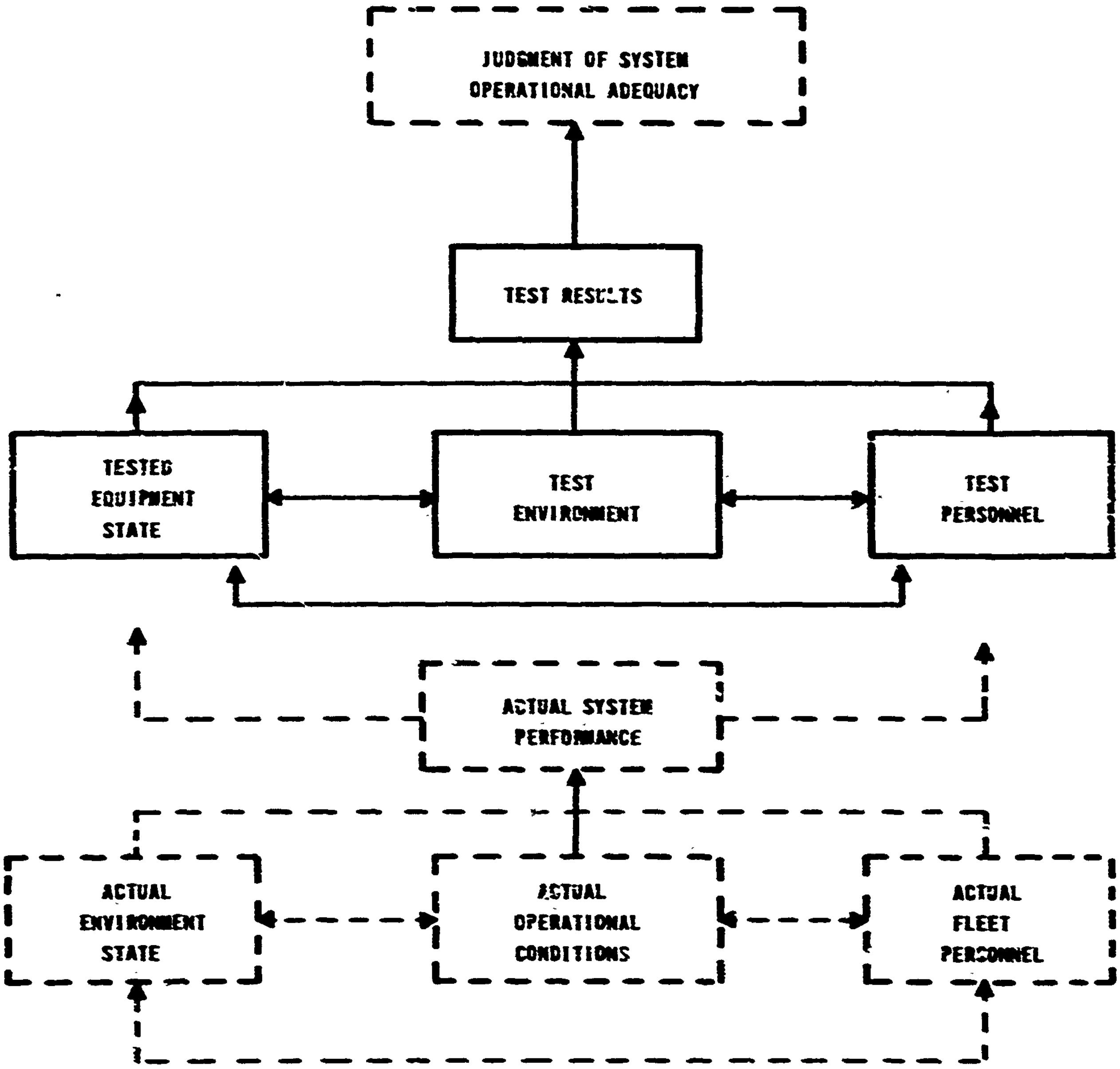
Because of the extreme importance of the type of judgments a project officer must make, OPEVALs are conducted in as rigorous a fashion as possible. Ideally, the entire test would resemble a large-scale experimental evaluation conducted within realistic environments while employing all of the rigor, control, and precision of the scientific laboratory. As anyone with field data-gathering experience will recognize, however, the introduction of realism is accompanied by losses in experimenter control. What complicates the project officer's task even more is the fact that he has no real assurance that the equipment being tested will remain operational throughout the test. When the system is "down" for extended periods of time for maintenance, important operational data cannot be collected.

The basic characteristics of an OPEVAL situation are shown in Figure 1.1. It can be seen that while the project officer is interested in the actual state of affairs, his data are limited to the test situation; i.e., he cannot include all possible operational variations within a single test. Thus he has the task of examining a complex territory guided by a map that is composed of information concerning only the main features of the terrain. He must use his background information and experience to the best possible advantage in making estimates of features not shown in complete detail. To counteract this problem he uses the test plan to insure that the improved map he is attempting to develop in the course of the evaluation will be as accurate and detailed a representation of the territory as possible.

A further complication arises because all of the major sources of influence (shown as boxes in the figure) *interact*; i.e., what happens to one, influences what happens to the others. Therefore, development of a test plan which allows the derivation of a sound judgment is an extremely tricky business where complex interactions need to be anticipated, examined, and controlled. In view of these considerations, the need for broad operational experience becomes even more readily apparent.

1.2 IMPORTANCE OF HUMAN FACTORS IN OPEVAL.

The diagram shown in Figure 1.1 and the associated discussion show rather clearly that *each of the major areas of consideration (the boxes) not only contributes a direct influence of its own, but an indirect influence as well by its interactions with other major areas of consideration.* This indirect



NOTE: DASHED LINE BOXES INDICATE WHERE THE PROJECT OFFICER'S JUDGMENT AND ESTIMATES PROVIDE NECESSARY INFORMATION. NORMAL TYPE, SOLID LINE BOXES INDICATE WHERE CONCRETE OBSERVATIONS CAN BE MADE

FIG. 1.1 - ILLUSTRATION OF THE OPEVAL PROBLEM

influence is one of the factors that serves to complicate the project officer's task even more, for if it is not controlled, it becomes impossible to say what results can be attributed to each source of influence acting alone and what results can be attributed to interactions.

For example, consider the case where the project officer is focusing his complete attention on a surface-to-air missile (SAM) system, one of the chief advantages of which is an ability to engage low-flying targets. In conducting a test of this capability, many of the necessary environmental characteristics are defined, especially the need for flights "on the deck." Despite this rather high degree of environmental specification, other environmental variations can arise to complicate matters; e.g., the weather might change drastically. In the same manner, human factors variations can arise to make interpretation of test runs difficult; unfortunately, these are more subtle in nature and more difficult to observe but in no case are they any less important. Consider, for example, that the tracking radar is "peaked" (adjusted with maximal precision) for high flights but not as well "peaked" for low flights. This source of variation which is contributed by the maintenance technician is not easily noted but can cause wide variations in the time to "lock on" a target. Another instance can result from unnoticed crew variations. It may be that the tracking console operators are changed systematically but that no control is exerted over the manning of other system operating positions. These variations may influence tracking console operator performance since it is quite likely that these men have different amounts of test experience and differing amounts of practice working within a particular team--not to mention their prior basic experience and skills. A particularly striking example of this influence arises with a search radar operator who has learned the value of "preping" the tracking radar operator; i.e., giving him precise clues on where to pick up the target. If a substitute is not as adroit in this "preping," lock-on times can increase drastically--as has been noted in OPTEVFOR tests in the TARTAR. Finally, the project officer might get an overly optimistic view of the system's ability to handle high altitude flights because of radarman's unique ability to use auditory returns of the radar output to distinct advantage before the target even appears on the screen, i.e., do his own "preping."

All of the above examples were gathered from specific OPTEVFOR test experience; they are not hypothetical cases. According to the estimates of the project officer involved, such human effects could *easily* cause lock-on time performance to vary by a factor of three to one, or even more. Similar operator influences have been noted in many OPEVAL's . Therefore, care must be taken to insure that when these effects are *not* being studied, they are controlled in such a way that they do not produce results which will be attributed to something else (the equipment capability, for example).

The importance of human factors effects, as such, is equally worthy of consideration, since the project officer watns to know what causes wide changes in system performance. In Volume I of this report it was shown analytically that for a SAM system, if it is postulated that everyone behaves optim. 'lv except a single tracking radar operator who is only 10 percent higher than the standard on his lock-on times, in many raids system performance (as measured in kills) will drop 33 percent. The important thing to remember here is that 10 percent variations in performance between men are routine; much more sizable variations often occur. When these are added to variations in maintenance technician proficiency and other operator station proficiency, the results on system performance become dramatically striking, *even when examples are chosen which are deliberately conservative estimates of effects noted in research and in OPTEVFOR project officers' experience.*

1.3 EFFECT ON OPEVAL ADEQUACY.

The meaning of what has been illustrated above can be summarized quickly by reference to what takes place in an OPEVAL. As Figure 1.1 indicates, there are at least three important sources of influence present, not to mention effects occurring as a result of source interaction: (1) the equipment; (2) the environment; and (3) the humans. If an OPEVAL accumulates systematic data about *only* the equipment, all test data can be parcelled into only two components: those performance variations associated with equipment and variations not attributable to any specific source. Obviously, environmental and human influences have not disappeared by choosing this approach; their effects appear as part of the residual variation component. If environmental effects are studied systematically, the residual component (performance variation unaccounted for) shrinks but the effects of human influence, the multiplicity of interaction effects and errors remain.

The point to be emphasized throughout this report is that we propose to show how the residual variation component in OPEVALs can be reduced still further by systematic studies of human factors. In this manner, more of the system performance observed can be related to specific types of problems and, more importantly, pointed toward different kinds of solution.

1.4 OPEVAL CONSTRAINTS ON HUMAN FACTORS TESTING

Having discussed at some length the complications and importance of human factors (HF) in OPEVAL studies, the demands of the problem must be tempered with practical constraints levied by the project officer's situation. These can be listed as follows:

(1) While human factors are important, they are only one of a number of important factors that need to be considered by the project officer.

(2) It is unreasonable to ask that the project officer become a skilled human factors specialist. He cannot trace all of the implications of what he observes, nor should he be expected to comprehensively examine exactly how some of the important human effects he notes can be brought under control.

(3) Most test environments do not have a large enough sample of men for the project officer to get a clear indication of how all of the human factors problems arise. He cannot duplicate the fleet personnel problem on a single ship.

(4) Many human factors tests are extremely time-consuming and expensive to run because all of the situational conditions must be repeated *exactly* with different men working within the system.

(5) Practical aborts can be expected which will complicate and sometimes negate the project officer's attempts to complete a test of anything, including human factors.

(6) The most efficient use of the project officer's time would be to concentrate on the most important human factors effects and to gather data on these effects. Additional associated problems must, by necessity, remain the province of the HF specialist who can assemble large amounts of data, carefully study the personnel situation in general, and draw needed conclusions.

In recognition of the above, the technique used in this report is to concentrate on a step-by-step approach to IIF which parallels the steps taken in OPEVALS by project officers. These are discussed in more detail in Section 3.0. Throughout this implementation discussion, care is taken to emphasize what can be done quickly and easily and what can be done in face of the many "unexpected" contingencies which arise in testing.

In addition to the points raised above, one further practicality must be recognized; (7) OPTEVFOR Staff Personnel are sometimes concerned with OPEVALs in a slightly different fashion; i.e., OPEVALs conducted by OPTEVFOR subunits such as NORTEVDET, where an OPTEVFOR Staff Officer acts as a monitor and coordinator of the test.

In this case, the OPTEVFOR staff is not concerned with performing each and every step in OPEVAL. Rather, their problem is to insure (by supervision and guidance) that all of the necessary actions are taken to insure success of the test. In this case, the material presented in Section 3.0 and following can be used as a guide and checklist for the subunit and for the OPTEVFOR staff coordinator.

2.0 TESTS OTHER THAN OPEVAL'S

2.1 INTRODUCTION

Responsibility for a number of other types of test and evaluation work is charged to OPTEVFOR. These include technical evaluation, concurrent evaluation, logistic evaluation, development assist tests, fleet research investigations, and fleet operational investigations. Because the main concern of this entire document is the operational evaluation, this chapter will be limited to a brief discussion of each of the above and a short summary of human factors work that can be accomplished in conjunction with each. These summaries will reflect two major considerations: first, the accumulation of information and background material useful in the OPEVAL, and second, where appropriate, the minimizing of human influence on the conduct of the test or evaluation.

2.2 TECHNICAL EVALUATION

A technical evaluation is the test and analysis required by a developing agency to determine whether a weapon system, support system, component, equipment, or material meets design specifications and is technically suitable for operational evaluation or service use. Where an operational environment is desired or required in connection with such evaluation, tests incident thereto are conducted for the developing agency by the Operational Test and Evaluation Force utilizing mutually agreeable plans.[1]

Under this definition, the responsibility of OPTEVFOR is limited to operational planning and conduct of tests that utilize operating forces. The requirements and specifications of the tests are generated by the developing agency. Another limiting factor involved in the technical evaluation is the degree to which special technical assistance is employed during the test itself. Although regular Navy personnel work on the gear, civilian engineers and other contractor personnel are on board and do take a direct hand in the proceedings.

---

[1] Definition taken from COMOPTEVFOR Inst. P3930, 1 c. Vol. I, 1 April 1961.

The project officer assigned by COMPOTEVFOR must complete a cycle of activities much like those undertaken in OPEVAL. To the extent which these activities are the same or highly similar, they will not be discussed here; they are given extensive treatment in later chapters. Essentially the steps are as follows: (1) accumulation of background knowledge about the system; (2) development of a test plan (in this case the planning is done in close conjunction with the developing agency); (3) concurrent with these activities, arrangements are made for the needed services from operational units; (4) during the test phase the project officer coordinates the activities and aids in supervision of tests involving the operational environment.

In the data evaluat:-n phase, although COMOPTEVFOR can and has in the past reported more extensively,the responsibility is officially on the developing agency; only a report of services rendered is required from the OPTEVFOR officer.

2.2.1 Human Factors in a Technical Evaluation.

The OPTEVFOR representative is presented, during a technical evaluation, an excellent opportunity for observing and studying the sources of human influence on the operation of the system. While it is true that he can do little about such influence at this stage, he can identify those points at which the process is channeled through a human operator. It is further true that little can be gained from observation of maintenance technician performance because of the presence of engineers and equipment experts who are often called upon for the really difficult corrective maintenance tasks (particularly fault diagnosis).

The exact nature of the observations that can be made at this time is difficult to specify independently of the particular system or equipment being evaluated, but the following are some clues as to the possibilities:

(1) Identification of operator functions with emphasis on those required to initiate or direct action.

(2) Identification of tasks that appear to be most demanding and difficult (e. g., most fatiguing, most prone to error or delay).

(3) Estimation of level of training of operating crew (e. g., have they received special training on this particular gear, or are they attempting to transfer skills learned on the previous generation of equipment?)

(4) Estimation, in the case of on-the-job training, of the amount of practice required to produce adequate performance.

(5) Identification of points available for automatic data recording concerning operator or component performance.

(6) Identification of special problem areas such as interfaces between equipments, between operator and equipment, or between operators.

The above list is not exhaustive and there can be no substitute for the project officer's insight and ingenuity. The important concept here is that *any* information that can be gathered at this time will greatly facilitate planning when the system, component, or equipment is finally brought up for OPEVAL.

2.3 CONCURRENT EVALUATION

A concurrent evaluation is the joint test and analysis of a weapon system, support system, component, or equipment conducted by the developing agency and the Operational Test and Evaluation Force to determine the system, component, or equipment meets design.specifications and specified operational requirements and is suitable for service acceptance. A concurrent evaluation is conducted in accordance with a plan prepared jointly by the developing agency and the Operational Test and Evaluation Force. The plan is fully integrated to achieve maximum economy of time, funds, and fleet services.[1]

In effect, the concurrent evaluation is two evaluations at once (technical evaluation and operational evaluation) and as such presents some rather unique problems to the project officer. First of all, is the division of responsibility between the developing agency and COMOPTEVFOR, the former being responsible for technical aspects and the latter, the operational. The matter of technical assistance also can operate to complicate both the planning and the conduct of a Concurrent Evaluation since the expert's services may well be needed in the technical tests but would serve only to contaminate operational tests. These factors, when viewed against the background of the requirement for total integration, place severe demands on the personnel who plan and conduct the tests. In reality, the situation is more complex than one would expect from merely adding an OPEVAL to a Technical Evaluation since the activities in the two different phases interact or have an influence

[1]COMOPTEVFOR Inst. P3930, 1 c. Vol. I, 1 April 1961.

upon one another. The procedures of the project officer are, on the surface, a simple combination of his duties in each separate type of test and, therefore, do not require restatement here. The complications arise in the details involved in the carrying out of these duties.

2.3.1 Human Factors in Concurrent Evaluations.

The control of human influences on the performance of the system presents a considerable challenge in this context. As in an OPEVAL, there can be no substitute for careful planning, but in this case the project officer may be severely restricted in the degree to which he can employ the techniques discussed in the following chapters. It is quite possible that he will have even fewer data available, since some of the services will, of necessity, be devoted to technical tests. Some of the precautions that can be taken are as follows:

(1) Build into the test plan as much separation of operational and technical tests as possible. (For example, the data gathered on an operational test could be unduly enhanced or degraded if the same personnel had just completed a long series of runs of the same type for a technical test.)

(2) Make use of the available operator skill level differences in order that not all of the data taken in a particular test is influenced only by a very expert, or inexpert operator.

(3) Carefully segregate maintenance data in order that no contamination by technical experts occurs.

(4) Be particularly sensitive to the effects of factors such as practice and fatigue (i.e., alternate personnel at key positions so that data are not biased by the facilitative effects of practice nor by the degrading effects of fatigue).

(5) Attempt to automate within cost limitations imposed as much of the data-gathering as possible on all tests in order to minimize errors and direct influence of data takers.

(6) Be alert to possibilities for combining the portions of data from technical tests with that of operational tests in order to obtain additional information concerning performance of certain tasks (e.g., although the primary concern of a test may be the slew time of a director, data may concurrently be collected concerning the performance of tracking operators).

In summary, it can be said that the main human factors concern in a Concurrent Evaluation may best be focused on those measures that would tend to reduce the variability observed in equipment parameters that is traceable to human performance. The major reason for this lies in the already stringent requirements arising from the necessity to conduct two types of tests at the same time.

2.4 DEVELOPMENT ASSIST TEST

A Development Assist Test is the testing of a weapon system, support system, component, equipment or material in any stage of research and development, wherein the assistance of ships, aircraft and/or other appropriate fleet units is requisite to the collection of data necessary for the determination of the direction in which an established development should advance. Such tests are conducted by the developing agency, utilizing the services of fleet units arranged for and coordinated by the Operational Test and Evaluation Force.

The role of OPTEVFOR in such tests is, by this definition, a limited one since both the planning and conducting responsibilities lie solely with the developing agency. It is, nevertheless, possible for the OPTEVFOR project officer to have an influence upon the tests since his knowledge and experience are related to the operational environment and missions involved.

2.4.1 Human Factors in a Development Assist Test.

When the Development Assist Test is conducted at an early stage in the Research and Development, Test and Evaluation, (RDT & E) cycle the main human factors consideration that should concern the COMOPTEVFOR representative is the accumulation of information about the functioning of the system or equipment. This information concerning system operation, information and action flow, manning requirements, special task requirements, etc., should be recorded and maintained for later use and updating during a technical evaluation or for later employment directly in the planning of an OPEVAL.

In those cases where the Development Assist Test occurs late in the RDT & E cycle (such as was the case with TARTAR), much human factors work of the type described later in the OPEVAL chapters of this report would be applicable.

2.5 FLEET RESEARCH INVESTIGATION

A Fleet Research Investigation is an examination into the military application of natural or special phenomena in the operational environment, as required by a developing agency in prosecution of research and for which the operating forces provide assistance for the conduct thereof.

2.5.1 As can be seen from the above foregoing material, this is a situation that is quite similar to the Developmental Assist Test at an early stage in the RDT & E cycle. This point, coupled with the fact that there is no system or equipment *per se* involved in this type of operation, precludes the discussion of any human factors work.

2.6 FLEET OPERATIONAL INVESTIGATION

A Fleet Operational Investigation is an examination or a comparison by the operating forces of concept, procedures, techniques, equipments, or material aimed at enhancing fleet readiness and with the concomitant aim of determining the adequacy of the RDT & E program in the area investigated.

2.6.1 For the most part, this type of testing is functionally nearly the same as an Operational Evaluation. The questions to be answered and the methods employed to obtain the answers are nearly identical. For these reasons the material in the following chapters will be, in large measure, directly applicable.

3.0 INCORPORATING HUMAN FACTORS TESTING IN THE OPERATIONAL TEST AND EVALUATION CYCLE

3.1 FORMAL DESCRIPTION

Describing in detail the activity cycle of a "typical" project officer handling a "normal" OPTEVFOR project is much like describing the "average" American man who has 2.4 children, 11.7 years of formal schooling, etc. The reason for this is that the project officer's job demands *flexibility*, and a facility to tailor assigned responsibilities to the unique features of each system, the testing environment, etc. For this reason, it is impossible to proceed beyond a crude functional analysis of what the project officer does without full realization that exceptions can be expected from time to time even at this level of analysis. By supplying such a description, however, two major achievements result: first, it immediately becomes clear that the task of the project officer is sufficiently complex and unstructured that he needs to draw often from his fund of general knowledge and experience; and second, it becomes possible for us to supply a description of how to derive and implement a human factors test at each stage of OPEVAL activity.

In short, the goal is to present information needed to broaden the scope of operational testing so that human factors problems are tackled in as rigorous a fashion as possible with minimal interference in the normal test cycle. Of course, additional effort is required to gain these extra advantages; further, the material provided is not likely to cover all contingencies which arise in operational testing. *This presentation aims to provide a means of doing the best job possible concerning handling of human factors problems and to do this in a way that human factors testing becomes an integral part of all operational tests. If this goal is met, human factors testing will not be an additional burden placed upon the project officer; rather, it will become a source of a assistance to him in obtaining realistic results and workable solutions.*

3.2 OPEVAL PHASES

Keeping in mind the restrictions described above, the authors' observations and the comments of OPTEVFOR staff members indicate that the OPEVAL cycle can be considered as being composed of five major phases: (1) preparation and initial planning; (2) devising and writing the test plan; (3) conducting the test; (4) evaluation of data from the test; and (5) derivation of conclusions and preparation of the final report (including "cutting board" activities). Each of these phases will be discussed in turn, beginning with a description of what the project officer is already required to do, and ending with a description of what steps are required at each *stage* within each *phase* in order to provide adequate coverage of human factors questions. As an additional assist, the latter material ends with a set of questions which, when answered, gives the project officer reasonable assurance that he has not overlooked important points in his preparation for human factors testing.

4.0 OPEVAL ACTIVITIES IN THE PLANNING AND PREPARATION PHASE

The project officer has three main functions in the planning and preparation phase: (1) accumulating sufficient knowledge about the system to be tested; (2) using this knowledge, plus the text of the assignment letter (as well as past experience within OPTEVFOR) to specify a set of test objectives; (3) insuring that the test can be accomplished during a specific time interval (the due date) by anticipating support needs from other groups, data recording needs of the test, etc., which require time to program. Having fulfilled these functions, the project officer has a good idea of what he has to do and has taken steps to insure that his job can, in fact, be done. Thus, he is in a position to begin to write the test plan.

The striking thing about each of the three preparation and planning functions is the complexity of job demands. For example, in accumulating knowledge about the system, project officers are required to search for information in any or all of the following sources: contract development documents; early drafts of technical manuals (finished manuals are rarely ready at this point); technical evaluation test results (when available); discussions with OPTEVFOR technicians sent to factory training schools; Technical Development Plans (TDPs); reports from the Bureau of Weapons and Bureau of Ships; reports having system descriptions from the Bureau of Personnel; and discussions with contract engineers and other OPTEVFOR staff members. While the above enumeration is not exhaustive, it suffices to demonstrate the difficulty of integrating information once it is obtained; more importantly, it demonstrates the difficulty of getting needed information in the first place. In deriving the test objectives, the project officer's task is equally complex; he must begin by taking the objectives as broadly stated in the assignment letter; he must, after extensive reading and numerous meetings, identify, isolate, and determine how to measure critical performance characteristics; he must study and weigh all kinds of "environmental" effects (weather, attack conditions, state of equipment readiness, extent of countermeasures employed, etc.); and, finally, he must not only draw upon the experience of OPTEVFOR staff members who are familiar with "earlier generation" systems aimed at achieving the same general mission, but if it is at all possible, he must try to tailor his plan so that his results can be meaningfully combined with earlier tests to provide an expanding fund of recorded knowledge at OPTEVFOR.

Finally, when the project officer reaches the third and last step in the preparation and planning phase, he must, by necessity, make early commitments of considerable significance *before* he has had time o complete his plans in full. For example, if a Surface-to-Air Missile System (SAM) is being tested, he must arrange for air support to simulate the targets. He must give some indication of how long the support is needed and, roughly, what is required of it so that supporting commanders can arrange to have the proper aircraft and equipment available. He also has to anticipate personnel requirements on board the test ship, decide who must be recruited from the ship's company, and anticipate his personnel and general equipment requirements from OPTEVFOR so that efficient scheduling of OPTEVFOR resources results.

The complexity inherent in the preceding discussion is illustrated in Figure 4.1 which describes in crude form some of the sources of information used in planning, how these are integrated, and how the results interact and combine to complete the preparation and planning phase activity. No attempt has been made to cover all possible sources of information or all likely interactions among sources of information.

4.1 ACCUMULATING KNOWLEDGE CONCERNING HUMAN FACTORS CONSIDERATIONS IN THE PREPARATION AND PLANNING PHASE

While the project officer is going through the three basic steps necessary to complete the preparation and planning phase of OPEVAL's, concurrent activities in the human factors area are needed. The first step is the accumulation of knowledge concerning where and how human factors effects exert themselves in the system. It is necessary to recall here that the variations in operational and maintenance proficiency occur primarily as a result of factors such as differences in basic aptitude and motivation, differences in training, and differences in the complexity of the equipment. In order to take such factors into account, it is necessary to collect information bearing upon certain fundamental questions. These questions form the backbone of the human factors efforts in this and subsequent phases of the OPEVAL cycle.

(1) How many men are required to operate and maintain the system or equipment to be tested?

(2) What are the exact duties associated with each operator station and each maintenance technician assignment?

(3) What are the form and extent of each man's contribution to the operation of the system?

(4) Which, if any, of the operator and technician positions are particularly crucial to the system performance?

(5) What is the potential effect upon the system of a delay or an error by the operator or maintenance technician at any one of the positions in the system?

If answers are obtained to these five questions, the task of accumulating the necessary background information about the human factors effects is satisfied. However, gathering such information is not easy; for this reason, it is necessary to discuss in some detail how answers can be obtained.

The first two questions concerning how many men are involved, and what each of them does, can best be answered by reference to systems (personnel and training information) furnished by the Bureau of Naval Personnel. If this information is not yet available in published form, it is possible to question technical personnel from the Bureau of Naval Personnel to discover whether such a study is in progress. If a study is underway, conferences can be held with the personnel research specialists to obtain the necessary data. If no study is in progress, it is necessary to turn to alternative sources of information. One of these sources is the Technical Development Plan which generally contains in crude form the basic human tasks required, as well as the manning statements for the subject system. From here, concurrent with discussions with contractor personnel who have developed the system as described in 3.3, queries can be made to define each task specification. As an added precaution, however, this information should be cross-checked with operational Naval personnel on board the test ship to insure that the normal complex of "field modifications" have not changed the tasks significantly. If this latter source of information is not yet available, the contractor's information can be preserved and checked by personnel on the test ship as soon as it is possible.

Having accomplished the preceding activities, *most of which are done concurrently with the activities needed to perform any sort of OPEVAL the additional load on the project officer can be decreased by an appreciable extent.* The project officer then has a good idea of what each man is supposed to do in the system. If the number of human tasks is small, study of the human factors problem stays within reasonable limits. If, however, as is often the case with

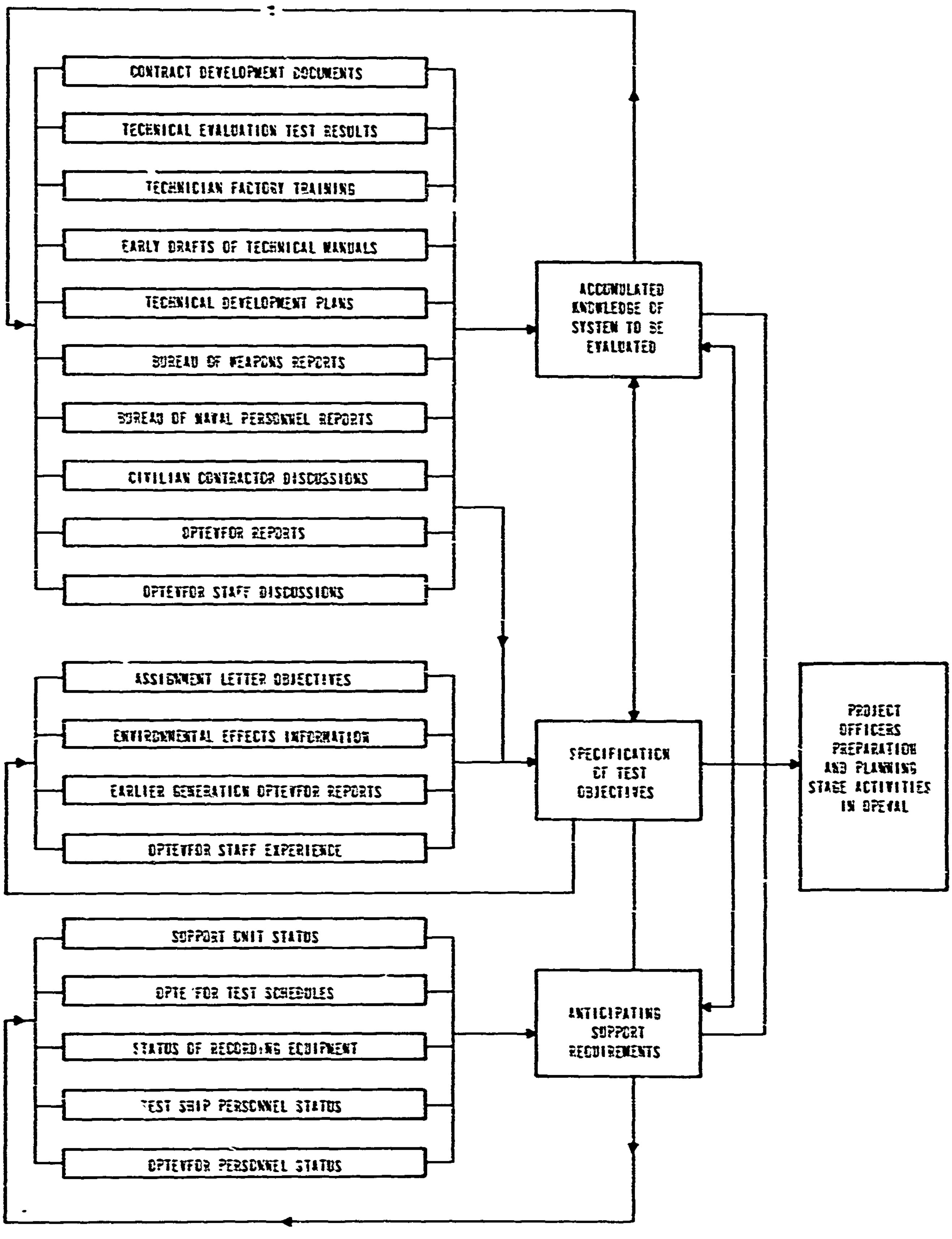


FIG. 4.1 - ILLUSTRATION OF PROJECT OFFICER'S ACTIVITIES IN THE PREPARATION AND PLANNING PHASE OF OPEVAL

large systems such as the TARTAR, there are more task positions to study than there are OPTEVFOR personnel available to study them, some compromises are in order--the most likely one being to choose to focus attention on the "critical" or most important tasks.

Identifying critical human tasks within the system is no easy matter. For example, one answer almost certain to be heard from design personnel is that "the *only* critical task is that of maintenance. If the equipment is functioning properly, there is no operator problem at all." Unfortunately, such replies often come from a heavily biased set of experiences. The first of these is the designer's effort to make the system as fully automatic as possible, usually in line with his contract specifications, and the second is that practically all new systems are beset with malfunction problems at the prototype stage. *What is forgotten in the process is that no operator effect on system proficiency can show itself when the system is "down" for maintenance. Thus, operator effects may well be there but there has been no opportunity to observe them yet. It is fallacious to conclude that such problems are not there simply because they have not yet been observed.*

Rather than become embroiled in a series of emotionally charged discussions concerning whether or not operator performance can really influence this "foolproof" system, it is better to approach the operator problem systematically and obtain data necessary for deriving conclusions. To begin, it is necessary to trace early potential operator effect through the system to see how it influences performance. Usually it is best to begin by considering the information supplied to the operator. Then, it is also possible to specify what actions the operator takes in view of the information. Finally, the question of how operators make decisions concerning action or no action, or action No. 1 versus action No. 2, etc., must be considered. In the forseeable future some such data may be available from studies conducted under the auspices of the Bureau of Naval Personnel. Meanwhile, the information must be obtained from the design and operational personnel available. With these types of data as a background, the next item of concern is to identify those points at which it may be possible to observe the effects of the actions selected by the operator and additionally determine the adequacy or appropriateness of the action. For example, it may be possible to monitor the opening or closing of certain relays or the arrival of an action signal at the next position in the sequence. In this way, the foundation has been laid for determining the effects of various amounts of operator delays or decision errors. This leads directly to a determination of the potential importance or criticality of each position and permits the project officer to concentrate his efforts on those that are most important. This rational analysis can be substantiated by means of direct questions asked of designers and operational personnel in an attempt to isolate those positions having critical importance. One simple way of doing this is to ask what the respondent would do if he had only one outstanding man--where would he like to see him assigned? What if he had two, etc.? Answers to these questions will probably serve to identify at least what is felt to be important. An additional part of this problem is the likelihood of errors or delays at these positions. One way of achieving this is to ask Bureau of Naval Personnel specialists[1] how likely it is that wrong decisions will be made at each position or what sizable time delays might appear. His replies, when translated into their effects on systems performance, will allow a comparatively objective determination to be made concerning what operator positions are really crucial to the system. At this point all five of the basic questions proposed earlier have been addressed but only in terms of the operator aspect of human factors.

It should be noted that much of the preceding discussion has focused on the operator effect and disregarded the technician effect. The reasons for this are two in number; first, the importance of the operator tends to be underplayed for the reasons given earlier (if the equipment isn't working how can we tell what the effect is?); second, the task of determining the importance of the maintenance technician is much easier to handle; i.e., the importance of maintenance is readily acknowledged by practically everyone and it is easier

[1] Contact can be made most directly with the Bureau of Naval Personnel Liaison Officer at OPTEVFOR.

to trace the effects of maintenance on system performance. For example, in many cases, it would be relatively easy to assess the degree of degradation resulting from operation in a casuality mode as compared to the fully operative system. *These statements should not be construed as meaning that there are no cases where operator effects are negligible; rather, the tendency has been to jump to this conclusion prematurely. As shown earlier (see Section 1.2) as the equipment remains operational (is down less often) these operator effects demonstrate themselves more often and become more and more important. Therefore, it is necessary to look beyond the usual jungle of maintenance problems to discover what can happen when the system is reliable.*

The logical starting point for considering human factors problems in system maintenance is the system description. For example, what happens when various elements are down for maintenance? Can the system still operate even with a reduced capability or is it out of action altogether? The next step is to consider reliability data to determine how likely it is for each element to fail. Finally, by projecting variations in the amount of down time, it is possible to quickly calculate the likelihood that the system will be up at any particular time. For example, if failures are expected in element X, which when down, renders the system inoperative, at a rate of two per thousand hours and, if we are interested in the effects of taking 30 minutes, 60 minutes, or 6 minutes to correct these on the average, we can see that in the 6-minute case this *one* element will cause the system to be down 1/5 hour in a thousand; in the 30-minute case one hour in a thousand; and in the 60-minute case, two hours in a thousand. When reliability figures diminish, the consequences of such differences can become especially striking.

However, not all maintenance problems lead to total system failure. What remains to be explored is the effect that misalignment has on the system performance and the likelihood that misalignment might occur at each operational position. For example, a simple operator task might suddenly become extremely difficult if the equipment is badly aligned. Such possibilities should be investigated while gathering background material concerning the system.

4.2 SPECIFYING HUMAN FACTORS TEST OBJECTIVES

This step begins by acknowledging what can and cannot be done in general within the context of an OPEVAL study. For example, it would be extremely useful to know exactly how variations in aptitude, motivation, and training affect all types of human performance in the system being tested. To answer such questions, however, large-scale studies are necessary using large numbers of Navy operational personnel with numerous duplications (replications) of expensive test conditions (e.g., flying aircraft). Further, it is necessary to insure that a sufficient sample of data is taken from each person so that legitimate inferences can be made concerning what each person can do consistently. A good illustration of this last point occurs in analyzing maintenance activities. Here, in order to eliminate unwanted effects of trouble difficulty, it is necessary to average repair time scores for at least eight troubles for each person studied. *For these reasons, it is well to acknowledge that full-scale testing of human factors effects cannot be accomplished in a single test. What is more reasonable is to use simulation (which is much less expensive) to study such questions and to count on a growing body of data from many tests to point toward generalizable relationships and conclusions.* In some cases, however, large numbers of persons can be studied using simulated operator inputs and simulated troubles. When such opportunities arise, it is worthwhile to contact the Bureau of Naval Personnel representative to obtain assistance on planning, designing, and conducting the test. In this manner additional value can be obtained from the available data. The reason for this suggestion is that the problem of designing studies of human proficiency is complex, usually requiring a timed study and practical experience in order to avoid numerous subtle pitfalls.

In view of this situation, a "tolerance limit" approach is suggested where the project officer attempts to differentiate between what the best trained, most capable men available can do as opposed to what the less trained, less capable men can do. Judgments as to who in the ship complement belongs in either extreme group can generally be obtained from the Supervising Chief Petty Officer (CPO) simply by asking "If this were combat, who would want on this gear and who would worry you most if he were here?" While it would be better to have more carefully defined extreme groups in an absolute sense, taking this simple step will provide much additional information. Then, in order to reach any conclusions at all, the following conditions must be satisfied:

(1) No matter what the size of the group being teste ? s, it is necessary to insure that the data from each person are actually representative of performance; if the same data were taken again, similar results would be obtained (insure that an adequate activity sample is taken).

(2) No comparisons between different individuals or extreme groups can occur without assurances that the task of each person is as similar to the others as possible.

(3) It is necessary to insure that data recorded combine not only what was done by the person, but also who the person is, what time it occurred, and what test conditions were in force at the time.

(4) Precautions must be taken to insure that the data recorded are not contaminated by such factors as practice effects, e.g., one operator works "cold" while a second works with people who have just completed a series of highly similar tasks (the effects of the adequacy of help given by other persons will surely show itself in a person's performance score).

If the above precautions are taken, there is cause to believe that the general objective of obtaining a reasonably workable, reliable, and valid set of data will be met. What need to be added to these generalizations are the specific goals to be met in each test.

4.3 ANTICIPATING NEEDS FOR HUMAN FACTORS EVALUATION

The last step in the preparation and planning phase involves anticipating testing needs. These are of three basic types: first, determining how the test ship's complement meets the requirements for setting up "extreme group" or "tolerance limit" tests; second, on the basis of material gathered in Step of this phase, to determine how information concerning the results of human activities can automatically be recorded and, thereby, reduce demands on observers; and third, to reconcile recording equipment demands with supplies available on board ship and at OPTEVFOR supply.

The second item mentioned above deserves elaboration because of its essential importance. Anything that can be done in the way of reducing the load on the OPTEVFOR observers is worthwhile, especially when it adds refinement and precision to data gathering procedures without greatly increasing costs.

Consider the case that one wishes to monitor the performance of the TSTC operator in the TARTAR system. There are two aspects of his performance that are of importance; the timeliness and consistency of his updating of the data stored in the target tracking channel. Both of these factors can influence the speed with which the target can be acquired. The most apparent way of obtaining data on these functions is not, in this case, the most appropriate one. Because of the press of activities already conducted in the CIC/WDS complex, the addition of an observer with a stop watch is just not feasible. Furthermore, even if such an observer could be placed in the CIC/WDS, his observations would be open to serious question on several grounds, e.g., his reaction time, his influence on operator performance (can be either facilitative or deleterious), etc. This being the case, it would be desirable to obtain those data in an automated fashion--preferably at some remote location. In the specific case in question, it is possible to do this by means of an events recorder located in the unattended equipments space and connected to certain relays in the target-tracker unit. This method will produce complete permanent records of operator performance and allow for accurately timing to 0.1 second without interfering in any way with the operator's task or the general conduct of the CIC/WDS.

4.4 SUMMARY

In the preparation and planning phase of OPEVALs, steps taken by the project officer to account for human factors are all of a practical nature; i.e., they are centered about the potential influence men have upon the system's ultimate mission and the extent to which practical limitations of the OPEVAL situation allow empirical assessment of these potential effects. Therefore, what needs to be done in the preparation and planning phase can be summarized quickly by attempts to answer the following questions:

(1) How many men are involved in operating and maintaining the system to be tested?

(2) What does each of them do and what are their primary responsibilities?

(3) How do these human actions effect system performance?

(4) What happens to system performance if human performance efficiency varies considerably in either a positive or negative manner?

(5) Which of the operator and maintenance technician positions have the greatest *potential* effect on system performance?

(6) According to early, crude estimates, which positions are *likely* to have significant effect?

(7) How might these effects be monitored without increasing the load on project personnel?

(8) How much of the human factors problem can be studied within the practical constraints of the OPEVAL (limited number of personnel, limited time, etc.)?

(9) What is the general plan of attack aimed at assessing influence of those human factors selected for study?

5.0 PREPARATION OF THE PROJECT PLAN

In this phase, project personnel face important and critical aspects of the OPEVAL cycle. The project plan is a guide to all that follows, and it is axiomatic that omissions or misjudgments at this point have far-reaching consequences. The substance of the plan itself has been thoroughly outlined in COMOPTEVFOR Inst. P3930. 1c, 1 April 1961, but a brief consideration of each section in the plan and its implications will be undertaken here in order to stress the many interlocking decisions and considerations involved.

5.1 DISCUSSION OF TEST PLAN SEGMENTS

Purpose of the tests. General objectives of the tests are defined by the CNO assignment letter. These must be further refined and made explicit in terms of the mission or task the equipment under test is designed to accomplish. At this point much of the experienced judgment of the project personnel comes into play, as well as the background information obtained in the previous phase. The combination of these three elements produces objectives that are reasonable, consistent, and achievable.

Previous work, background, and material. Although this may not result in the actual preparation of a section in the project plan, it must be remembered that though the effort in accumulating necessary background information is laborious, it is vitally important to the development of the necessary understanding of and competence with the system. Many aspects of the actual testing will be influenced by the knowledge collected and integrated in this phase.

Supporting activities involved. Here again this section as it appears in the actual project plan document has its main importance insofar as it reflects adequate advance planning in terms of the types and amounts of services scheduled and the provisions made for the various functions required in the conduct of the tests. For example, have appropriate provisions been made for specialized data taking activities such as photographic coverage, or for on board "quick look" data analysis, etc.? An additional important function of this section is to delineate the areas of responsibilities of the various personnel and/or commands participating in the evaluation.

The operations schedule. This section requires a great deal of careful planning and work in order to utilize available services in the most efficient and effective manner. This often necessitates much trial and error tentative scheduling which is continually revised and modified in response to the operating realities of the units providing the support. These operating requirements may severely limit the flexibility of the final schedule--as, for example, when aircraft having special equipment or capabilities will be available only on certain days. A further restriction is often introduced by the operating areas assigned for the tests since the distance from the source of air services has a direct influence upon the available aircraft time in the test area and, therefore, upon the amount of data that can be obtained per run.

Plans for the tests. This section is, of course, the heart of the project plan, and it will receive most of the attention in this chapter. The central objective of the test plans section is to describe the tests required and the procedures to be followed to obtain data which will produce valid conclusions. This objective is approached with two underlying considerations in mind. The first of these is that the finished plans will be in sufficient detail so that someone *other than the author* can be entrusted to conduct the tests. Secondly, because of uncertainties imposed by the operating environment, plans must contain much inherent flexibility. It is almost certainly true that in this type of testing unforeseen events will occur that will necessitate changes in plans. It is incumbent, therefore, upon the planner to provide reasonable contingency plans and to so construct and schedule the tests that modifications can be made "on the spot" which will not adversely affect the accumulation of the required data.

In connection with outlining procedures for the tests, ample detail must be supplied not only concerning conditions of the tests (e.g., mode of operation, flight profile, ships maneuvers, etc.) but also on the methods of collecting and recording data. Sample data sheets must be supplied. Starting, marking, and stopping procedures for automatic recorders must be given. Detailed instructions for photographic coverage and provisions for "dry runs" must be included. Care must also be taken at this point to plan data reduction and analysis procedures, since the requirements here will have direct bearing on the type and amount of supporting data that will be required. For example, if it is planned to correlate measurements taken at two different points in the system, it will be necessary to insure that they are taken at similar times. Such a stipulation would undoubtedly require special procedures to coordinate the data takers.

5.2 HUMAN FACTORS CONTROL CONSIDERATIONS IN THE DEVELOPMENT OF A PROJECT PLAN

As with other aspects of the OPEVAL, the planning phase is the point at which the crucial decisions are made. Provisions must be made here for whatever controls and/or data gathering procedures will be employed in the tests to deal with human factors variables. Because this is true, it seems appropriate to go into considerable detail concerning the sort of provisions necessary in order to cope adequately with these variables. There are two main areas of consideration: control procedures and data gathering procedures.

Control procedures have to do with the attempt to eliminate, or at least to minimize, the effect of individual variations in aptitude, training, skill, or experience upon evaluation of the systems operational suitability. One of the primary questions asked concerning the gear being evaluated is, "Will it do the job under fleet conditions when operated and maintained by regular Navy personnel?" It is important to note that the procedures to be outlined here do not have as their goal elimination of the effects of different operators or technicians, but rather the freeing results from the influence of a particular operator or technician. In short, the control procedures are introduced so that a particularly skilled or unskilled crew cannot cause an erroneous conclusion to be reached concerning operational suitability.

The project plan, employing the information collected in the preparation phase, must take into account the possible effects introduced at certain previously identified critical points. For example, if two out of six possible positions to be studied not only have a large *potential* effect on system performance but are indeed *likely* to have this effect as well, care must be taken to insure that human factors at these two positions are either studied directly or at least sufficiently well controlled to eliminate contamination of observed results. In order to study such effects directly, it is necessary to time-correlate human actions with system functions on the observational record and assign operational personnel in a fashion which allows these objectives to be met.

5.3 RECORDING HUMAN PERFORMANCE DATA

The focus here is on preventing contamination of data by human errors. Because of the necessary reliance on human observers for portions of the data, this will always be an area of importance. Certain functions cannot be automatically recorded. In addition, because of space or power requirements or cost, it may not be feasible to automate human performance data recording in all instances. As a general rule, however, it will probably be advantageous to automate as much as possible in order to minimize human error. Therefore, some suggestions are made here as to instrumentation that would be useful for this purpose. These are given below in the form of a list of potential recording devices some of which now exist in the fleet and others which would not be difficult or expensive to obtain.

(a) Operations events recorder. There are many examples of this type of recorder, but perhaps the most common is the 20 pen Esterline Angus. When the desired data is the occurrence or nonoccurrence of a specific event, (defined as any discrete happening, e.g., onset of a signal, closing of a relay, etc.) the time between two separate events, or recurrences of the same event, this is a most useful piece of apparatus. It will display and record in a time-correlated fashion up to 20 separate functions simultaneously. Depending upon chart speed, the timing of intervals between events can be accomplished with an accuracy of ± 0.05 sec. It can be set up and monitored at a site remote from the actual scene of the activity, thereby minimizing the chance for interference with operational activity. Its major disadvantage lies in the time requirements for data reduction.

(b) Recycling timers. If a function or task of some length requires timing, it is often advisable to use an instrument of this type rather than a hand held stopwatch. The advantage here accrues from the fact that the timer is started and stopped by a signal from the equipment itself, thereby eliminating the reaction time variable introduced by the stopwatch operator. The data taker has only to record the time displayed on the timer, recycle it, and await the next reading. The remote site advantage also applies here.

(c) Tape recorders. In addition to their more obvious uses to record communications, sonar signals, doppler returns and the like, tape recorders can be an important adjunct to many other data taking activities. Properly employed, they facilitate the taking of notes and keeping of records concerning the conditions of a test and potentially important occurrences during a test. Notes taken by tape recorder are much more likely to be complete and accurate than written notes simply because of the increase in the speed with which they can be recorded. An additional benefit can be derived by making use of the fact that tape speed is known and fairly constant--a distant advantage in time correlating data points.

(d) Counters (hand operated and automatic). Either of these rather simple gadgets represent a substantial improvement over the tally mark method of recording frequency data such as is required for blip-scan ratio computations. Of course, where feasible, the automatic counter is to be preferred, but even the hand operated models will improve data accuracy.

(e) Function recorders. Although this class of instruments is large and diverse, it can be characterized in general as consisting of various types of motor driven charts upon which a pen or pens automatically trace the shape of the monitored function. For example, employing an instrument of this type it would be possible to monitor the smoothness of a fire control director tracking function or to produce a permanent visual readout of a sonar signal or a doppler return. There are, of course, other types of special purpose data gathering apparatus but the above are judged to be the most likely to be useful, and they serve to exemplify the possibilities for improving on human observers.

5.4 PURPOSES OF HUMAN PERFORMANCE DATA COLLECTION

The devices mentioned above are introduced to minimize error in data collection which may be attributed to human errors in recording and to maximize accuracy and reliability in data collection as well as to provide a means for such collection which does not interfere with the basic conduct of the test. One point should be stressed when employing such devices, however: A careful note should be made of personnel operating and maintaining the equipment at various times in the test period. This should help define the limits of variation in system performance attributable to human factors effects and aid in determining variables between personnel which are significant in terms of systems operations.

As was stated previously, while automatic data collection is desirable, certain important information cannot be gathered utilizing mechanical devices. The project officer's task requires that he make recommendations about ways in which the system could be improved, not merely state that the system does or does not reach certain specifications.

While, ideally, one might wish to take the approach utilized by experimental personnel working in a laboratory and experimentally manipulate each portion of the equipment to determine how each should be designed for maximal efficiency in terms of maintenance and operation, such procedure is usually precluded by factors of time and cost. Pragmatically, then, one must work within the confines of a given piece of hardware and seek to improve its performance.

Under these constraints, the study of human factors variables become crucial because (1) man is the most flexible part of the system, and (2) a study of the critical functioning areas which seem most difficult for those concerned with the operation of the equipment will reveal subtasks where changes introduced into the operation and technician tasks would have maximum effect.

In view of these considerations, the project officer should seek to determine several types of factors about the personnel involved with the maintenance and operation of the equipment. The first concerns the prior training and alleged job effectiveness of such individuals. This information in general form can be gathered from the personnel officer and from interviews with such persons' superiors on the ship. If differences between individuals exist in either dimension, as mentioned earlier, care should be taken to insure that conditions under which persons with more training or competence operating at a given work station are similar to those which persons with less training or competence encounter. An example of the way in which such procedures could be implemented is given in Section 7. If planning along such lines has been introduced into the original design and is implemented in the data gathering phase, the project officer will have some basis for stating in his report the effects which he believes are attributable to personnel training and competence. He [illegible], therefore,make recommendations about the type of personnel who should be assigned to operate and maintain this equipment.

5.5 ACTIVITY ANALYSIS OF MAINTENANCE TECHNICIAN TASKS

It was mentioned previously that a job analysis or study of the work involved in operating and maintaining the equipment would be valuable in terms of determining where the introduction of change in the hardware itself could be expected to produce maximal change as far as overall operating efficiency of the system is concerned. Previously it was pointed out that in the planning phase the project officer should determine those aspects of the system which seem most crucial to its operation. Having determined these, he should then attempt to narrow the field by determining which particular subsystems and particular tasks within the subsystem are crucial. The empirical approach taken to gather information about this would be similar to that which would ideally have been taken to determine what aspects of the system one would study, if practical considerations would not preclude such steps (e.g., complete experimental manipulation of all key variables.)

In this phase of the date gathering process, it may be helpful to differentiate potential methods of gathering data from those concerned with the maintenance of equipment and from those concerned with the operation of equipment.

Date concerning the maintenance technician can be obtained from failure report forms (such as Bureau of Weapons Form 8000/13) completed while the actual test is underway. It should be emphasized that an accurate record should be kept of the technicians who actually performed the maintenance, remembering that sometimes these reports are signed by the senior technician on duty even though he did not actually perform the work. It would also be useful, for future planning, to have the technician indicate on the reverse side of the form, the effect of this failure on the operation of the entire system.

The advantage of collecting these data while the test is underway is that the project officer knows that the forms are being completed in a way which will allow him to make use of the data. The disadvantage of this method is that the sample of failures may be so small as to make the drawing of conclusion impossible.

There are two ways of increasing the sample size to a level which makes analysis possible. The first is utilization of reports of failures which occurred before the actual test began. In order to take maximal advantage of this source, it might be wise for the project officer to request that copies of such reports, including the names of the technicians who actually worked on the failure and the effect of the failure on the whole system, be kept for him for a period of about two or more weeks of operation before the actual test begins. In lieu of this, he could try to obtain copies of past failure reports submitted on the system in which he was interested or resort to a study of the equipment log.

Both of these latter mentioned sources have the disadvantage of not having available as much detailed, unbiased information as might be desired. While the possibility exists of asking the technicians involved to try to remember what failures they dealt with and how they affected the operating system, such reliance on memory is tenuous and may result in misleading conclusions.

From failure report data information concerning the probability of failures in certain subsystems can be gleaned as well as information concerning the length of time required to correct certain failures. This, combined with information concerning operator training and competence, can lead to specific recommendations concerning what aspects of a job consume the largest amounts of time, are most important, etc. In order to consolidate such data from a number of individuals operating at a number of positions within a single system, the format of the task analysis form must be comparable. In this case, since a standard maintenance form already exists, there is no problem. In the case of the operator study, however, the project officer must generate the necessary forms.

5.6 ACTIVITY ANALYSES OF OPERATOR TASKS

With regard to the task of the operator, the chief methods of collecting data would be by the employment of mechanical recording devices such as those discussed earlier and by the employment of questionnaires. If questionnaires are to be utilized, it is necessary to prepare them in advance. One of the most important aspects to remember about the use of such devices is that care must be taken to insure that responses obtained in such a manner from different individuals are comparable. This should have been considered in the planning phase if utilization of questionnaires was contemplated in the data gathering phase. With respect to the use of questionnaires in the data gathering phase *per se*, only a few obvious precautions must be taken: (1) unambiguous responses should be elicited; (2) questions should be phrased to deal with specific rather than general problems; (3) respondents should be reminded of the importance of making independent answers; (4) respondents should be urged to answer as completely and as accurately as possible; and, (5) respondents should be assured that replies are used for research purposes and not to appraise the respondent.

Such questionnaire data can be obtained during times when equipment is down, runs are cancelled, or no runs are scheduled, thereby utilizing unscheduled "free time" for data acquisition.

The questionnaire itself should probably begin by (1) identifying those areas of communality between jobs, and (2) dividing each job into a series of nearly independent component tasks performed in sequence. Then, by asking questions as to the time spent on each component as well as to the estimated importance of each, estimates of relative criticality of the various components can be obtained. When such data are available, a priority system exists by which each difficulty associated with each job can be attacked with assurance that these early efforts will have the greatest chance of upgrading system performance. When efforts later in time yield smaller and smaller returns, a logical stopping point can be identified.

Two further items need to be considered in devising project plans. First, in addition to job analysis procedures as described above, questions should be added to solicit opinions as to what simple steps could be taken to improve performance (change in training, emphasis, etc.) as well as the consequences of failing to perform adequately on each task component. Second, there are cases where cross-comparison of operator task components becomes impossible throughout the entire range of tasks studied. Data bases within each task should then be made as large as possible to insure that identification of critical task components is accomplished as accurately as possible.

6.0. CONDUCTING THE TEST

In a laboratory testing situation, it is sometimes possible to attain an "ideal" state where the conduct of a test offers no problems, even though it might consume sizeable amounts of time. Since it is recognized that the "ideal" state is *not* normally met even in the laboratory, however, a "trial run" or pilot testing phase is inserted to insure that the actual test itself is completed with few, if any, difficulties. In the case of an OPEVAL, or testing under operational environment conditions, a project officer can anticipate numerous problems often to the point of substantiating Murphy's Law of "what can go wrong, will go wrong!"

In overcoming these practical difficulties, the project officer has to amend his plans, shift his emphasis at times, and, in essence, be extremely "flexible." Much of this "flexibility" evolves from a clear recognition of the fact that what he wishes to test cannot be tested completely. Therefore, he must salvage what he can from the situation by conducting subtests in order of their priority (which he has already established). In addition, by shifting features of "test runs," he is able to get the maximal amount of data out of what he has available for use.

The authors will not presume to describe how a project officer accomplishes these feats since, by definition, they are rather clearly an art and not a systematic (planned, controlled) science. However, since the problems of human factors testing may be somewhat novel to project officers, some time should be spent indicating things likely to cause trouble and what can be done in the face of these difficulties.

6.1. PROBLEMS IN QUESTIONNAIRE ADMINISTRATION.

One common reaction on the part of human subjects is to immediately "pick holes" in any questionnaire--no matter how carefully it was developed and pretested. Usually, pretesting on other operational personnel serves to eliminate some of the resulting criticism, but the project officer should recognize that a number of points of contention will remain in the instrument he carries on-board the test ship. The reasons for this human reaction are many. First, it is impossible to use written English and be perfectly lucid.

Second, much of the information desired requires personnel to make difficult decisions and choices (for example, "____that trouble listed in your log as Number 8, about how long did it take to find the source of difficulty?"). People do not like to make difficult decisions or choices or to be exact in their answers. Third, since all people are human, it follows (to them) that they are experts in understanding human problems. If your approach happens to differ from theirs, trouble can be expected. Finally, asking detailed questions of anyone tends to make him suspicious. Persons often feel that the questioner's purpose is one of fixing blame on the respondent.

All of these difficulties are compounded in an OPEVAL by the fact that the project officer has a higher military rank than the people he questions. Because of military discipline, many obvious reactions of the type described above will not appear in a way that the project officer can see them. What might happen, however, is that the respondents do a haphazard job of answering questions, thereby making the carefully-controlled, well-analyzed responses meaningless.

To overcome this type of reaction, one should anticipate it in advance and take steps to eliminate it before it has a chance to arise. Based upon the author's experience in collecting field data, it is possible to circumvent problems by doing the following: First, let the respondent know that you are aware that they can "beat the system" and probably fool you in the process. Second, admit that while the forms have faults, they are the best things available at the time. If they have ideas for future improvement, you would be happy to get them. Third, let the respondents know that if the forms are going to be carelessly completed, you would rather not have them completed (a small *valid* sample is better than a large invalid one). Fourth, have the questions prepared and presented verbally by an assistant (the personal attention often gets large amounts of cooperation from an otherwise indifferent source).

6.2 PROBLEMS IN ACTIVITY ANALYSIS DATA COLLECTION

Collecting questionnaire data is not especially difficult since this can begin during the trip to a testing site and continue on the return trip. However, it is best to quietly review responses to questions to insure that answers are complete, the respondent is properly identified, etc.

Activity analysis forms, which describe what each man did when correcting a trouble or reacting to his operator demands during a test run, are more prone to error. These should not only be reviewed for completeness but checked as to accuracy as well. For example, consider that the project officer on his return trip to Norfolk finds that all activity analysis forms are complete He can then select a few in a random manner and question the men who completed them. If it were a record maintenance activity sample, the questioning might be along the following lines.

"Jones, I notice that you had trouble with a thumbgadget on Wednesday when we were conducting high altitude test runs." "Do you remember that one?"

"Very much, sir, because the chief was on my back to get the computer operating as quickly as possible."

"Did it take you more time to locate the trouble to a cabinet or to locate the trouble to a subassembly within a cabinet once you knew which cabinet was at fault?"

"Working inside the cabinet was what really took the time, sir."

By comparing Jones' answers to such questions with what he wrote on the form, an estimate can be obtained of how accurately the forms were completed. By selecting a report which Jones remembers well, it is reasonable to expect that his answers to questions will be consistent. If they are not, a question arises concerning which answer is correct. Instances where the man completing the form differs from the man who did the work can also be noted by this process, thereby correcting a potential error in associating an individual with a particular task during the test. Because of the massive pressure and the resulting confusion during a test run, some errors of this type can probably be expected.

Collecting more objective data can also cause problems. For example, the preceding chapter extolled at some length the numerous virtues of automatic recording apparatus. Obviously, such automatic equipment is prone to failure and should be checked periodically during runs, as well as before and after such activities. In addition, a small pilot test is in order to insure that the response recording system is accurate.

The "pilot test" mentioned above is really a simple checkout procedure. During this "test," operators push certain buttons in a fixed sequence and a monitor determines whether each is accurately and clearly recorded. Then, tests are made to insure that time-correlated data can be inserted on the same record. In this manner, before the data are returned to OPTEVFOR, the project officer has appraised himself of how much data he has available for processing and analysis, as well as having estimated the worth of the data accumulated in cross correlation of one type of response with another.

6.3 THE PROBLEM OF OMITTED RESPONSES

Despite all precautions taken to the contrary, the project officer can be assured that some data points will be lost in the testing process. Causes of missing data points are numerous; by taking steps outlined in Segments 6.1 and 6.2, some of the lost data can be retrieved. However, if vacancies still remain, one of two possible alternatives is available to the project officer. First, he can simply leave the space blank and analyze the data that he has (the more defensible option). Second, he can insert data in a carefully controlled fashion. For example, the method of dropping data is sometimes precluded by the need to have all cell entries filled for the more orthodox types of statistical analysis. Consider the case where a project officer is going to have each of two operators perform at a weapons assignment console during three successive testing days where numerous target runs are presented to each man on each day. If he has taken the precaution of giving each man equivalently difficult runs on each day, his data chart might look like this:

	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
Operator A			*																					
Operator B			*																					

Within each cell, performance measures can be entered, e.g., target intercept range in yards.

If all data points (cells) are filled, the project officer can employ difference scores (Cell 1 scores minus Cell 2 scores, etc.) and test the performance differences between operators A and B by a simple t test having 23 degrees of freedom. If, however, data points are lost, the number of available degrees of freedom decreases, and concurrently the chance of detecting significant differences when they arise (the power of the t test) decreases also. In this case if the t test is to be used, the most appropriate thing to do is to accept the loss in test power.

If, however, the project officer had needed to conduct a more complex F analysis where he also analyzed the effects of eight different types of runs and the three different days of runs, the problem of overcoming missing data points becomes significant. The only way to do this in any realistic fashion is to insert data based upon best estimates of what would be likely to happen. Consider that the two starred cells in the figure above are vacant. The first step is to determine whether data can be inserted at all. As a rule of thumb, if the missing data points constitute less than five percent of all data points for small samples of data or with a lower percentage in larger samples, it is reasonable to insert data. That is, reasonable if other conditions are satisfied and conclusions which are based upon the data, consider that this step has been taken. In our case, there are a total of 16 X 3 = 48 data points possible, and two missing, for a percentage of 4.2% (1/24). Second, the two missing data points occurred on separate days, on different types of runs, and to different operators. Therefore, whatever errors might result as a consequence of the data insertion process will be reasonably distributed. If, for example, they were both for the same run, that run would have to be dropped from the analysis with an accompanying loss in power of the F test. With the first two conditions satisfied, however, it is reasonable to insert for each star the average of the scores for the other two days for that run for that person. Then the F test may be conducted. If it results that there is no significant performance variation from day to day, the data insertion process probably had little practical effect. If there is significant "day" effect, the inserted data will probably be a conservative estimate of what would probably have happened. In either case, as long as results of the F test are not at the level of borderline significance, the data insertion process should not cause trouble. Certainly, by its use, an otherwise impossible method of analysis becomes possible.

7.0 EVALUATING OPEVAL DATA

In the evaluation phase of OPEVALs, a project officer is concerned with the following:

(1) *Screening* data recorded during the test to insure that they are as accurate and complete as possible.

(2) *Cataloging* data so that appropriate associations are made and the data are prepared for processing.

(3) *Reducing* the screened and catalogued data, usually by computer or calculator operations, to insure that selected descriptive statistical summaries describe adequately major trends in the data.

(4) *Analyzing* the summarized data to derive implications of results which lead to as firm conclusions as possible.

The above mentioned steps are standard operating procedures for anyone in the research and evaluation field, and the numerous complexities and difficulties involved are readily recognized. This description will *not* attempt to provide what is, in essence, a manual of scientific procedure; rather, it will focus on peculiarities of the OPEVAL procedure which make these four operations especially difficult. For example, a project officer rarely has the options of the laboratory scientist; he usually cannot run the test again because it costs too much. Furthermore, the project officer cannot hedge in his conclusions because he *must be able to defend this decision against criticism.*

There are certain similarities between the tasks of a project officer and of the laboratory scientist, however. Neither can be absolutely certain of his interpretation of the data. Our current fact-finding procedures are such that a statement of absolute certainty cannot be substantiated. In addition, neither can one hope to cope with all possible criticism of his conclusions. Constructive criticism may be utilized by each to improve interpretations, later test plans, etc. ("Destructive" criticism may also be valuable, but all too often criticism of this type is merely the pointing out of the failure of obtaining absolute results, a panacea for all ills, a goal which any trained researcher recognizes as being unattainable.) However, both the project officer and the laboratory scientist should be reasonably well assured that the conclusions which are derived from the data are accurate.

In the remainder of this section, a new method of presentation will be used. Throughout earlier chapters, human factor steps were presented separately from other OPEVAL steps. By the evaluation stage in OPEVAL, human factors effects are *intimately* and *prematurely* attached to data showing other effects. Therefore, it is best to present both the "normal" *and* the human factors considerations simultaneously. Otherwise, a false division results which may tend to be confusing.

7.1 IMPLICATIONS OF THE SCREENING, CATALOGING, AND REDUCING FUNCTIONS UPON DATA ANALYSIS

The steps above are descriptions of what a project officer does. Naturally, these operations are performed for a reason; in this case, there are two purposes involved: (1) to insure that the data under study are as reliable, valid, meaningful, and comprehensive as possible; and (2) to transform data into a form which facilitates accurate interpretation. The problem, of course, is to insure that in the often complex process of data manipulation, synthesis, and analysis (including the usual statistical gyrations), important factors having significant implications on system performance do not become overlooked. If such oversights occur, interpretations can be less than optimal or, even worse, they can be erroneous and misleading.

The essence of the above argument is that it is important not to lose track of well known effects in the course of an analysis--including human factors effects. Factors which can be expected to influence performance, such as weather, maintenance and operating personnel, etc., must be duly noted and taken into account in the analysis. This point, which is extremely simple to understand at first glance, is *not* simple to implement in practice--especially when operating under the conditions routinely encountered by a project officer. Often only small amounts of data are available, and a number of factors have to be taken into account. *In this regard, another similarity between the problems faced by the project officer and those of the research scientist becomes apparent; it is extremely unlikely that anything can be done to take account of factors which were not anticipated. In testing, if attempts are made to study factors of known or suspected importance under operational conditions, this list forms an upper limit on the number of effects which can be analyzed and interpreted. Often, because of data loss, only some of the list can be studied.*

Perhaps these points--and the way human factors affect them--can best be illustrated by the use of two examples. Both examples will be as realistic as possible, but in order to keep this report unclassified, the data employed will be artificial. Both examples will include two features which are particularly relevant to the project officer's job. First, the amount of data studied will be sufficiently small to induce problems in interpretation. Second, additional effects not under the control of the project officer will arise to make analysis much more difficult.

7.2 FIRST NUMERICAL EXAMPLE

Assume that one of the objectives of a SAM OPEVAL would be to obtain accurate acquisition envelopes for tracking radars when very high altitude targets are employed. Since aircraft with this capability are not as readily available for OPEVALs as others, practical contingencies might have allowed only six overflights on a two-per-day basis. Also, assume that radar propagation conditions were near maximal on the first day, degraded on the second, and excellent on the third. There is little argument that this unwanted variation in propagation conditions *must* be duly noted in the analysis. Under these conditions the data might appear as follows:

		Acquisition Range
Day 1, Run 1	=	32,000 yds.
Day 1, Run 2	=	48,000 yds.
Day 2, Run 1	=	36,000 yds.
Day 2, Run 2	=	44,000 yds.
Day 3, Run 1	=	38,000 yds.
Day 3, Run 2	=	60,000 yds.

From these data it is simple to compute a mean for each day and to compare the three means to get some indication of the propagation condition effect. On Day 1 the mean (average) performance score is 40,000 yds; on Days 2 and 3 the means are 40,000 and 49,000 yds., respectively. If the first and third days' performances are averaged and compared to the second day's, an advantage of 4500 yds. results (for a good propagation condition), which can be indicative of the effects of the observed amount of propagation change. Thus, the project officer might indicate that while the target acquisition range for high altitude targets is a function of the propagation conditions present, values will tend to range from 40,000 to 50,000 yds.

While the above procedure does what it can to estimate the effects of variations in radar propagation conditions, it overlooks what appears to be a significant human factors effect. It is a common observation that people generally perform better in a situation *after* they have met it once: and it is a fact that the novelty of a high overflight decreases with each experience (run). In fact, it is reasonable to expect a marked decline in novelty after it occurs once. From this hypothesis we can proceed to examine the data to determine whether they support the idea being advanced.

The first support for the idea that learning (decreased novelty) has an effect comes from looking at the variation between means for each of the three days: Day 1 was 40,000 yds; Day 2 was 40,000 yds; and Day 3 was 49,000 yds. Thus, a slight trend can be noted toward increased acquisition ranges, especially when it is remembered that on Day 2 propagation conditions deteriorated. More striking differences emerge, however, from a second indication of learning effects, viz., the second run was always more successful than the first run. The mean (average) target acquisition range for all three first runs was 35,300 yds. while the mean for all second runs was 50,600 yds. Therefore, it is more than likely that a learning effect of the type described above exists and is sizeable. *Also, since it is unlikely that a high altitude overflight can be expected in advance, it might be appropriate to consider the 35,300-yd. value as being more descriptive of operational performance.* If this value is too low to be tolerated, a training recommendation might be in order.

However, there are more direct ways of tackling the training problem--ways that allow data to be accumulated to support recommendations. If, for example, a project officer found that on this test ship he had one operator who had received a special course and one who had not, he could take some steps to assess these effects. In determining what to do, it is obvious that each of the two men should be tested under equivalent conditions. Even after all precautions are taken, however, it is impossible to make the tests completely equivalent because whoever is assigned when the first overflight occurs has the most difficult task. Realizing the advantages of practice and anticipation, the project officer can attempt to neutralize them (by balancing as shown in example below) insofar as possible. In this case, it would seem wise to use both operators during each day (one run each which serves to neutralize the propagation effect) and to alternate which one works first. Since three runs do not allow complete alternation, someone has to get the first run twice. If the first day's first run is likely to be the hardest, the man who gets this assignment should probably get the easier assignment on Days 2 and 3. In short, what the project officer would like to do is follow this schedule (if he had 4 days of 2 runs each):

Day 1, Run 1 - Operator A
Day 1, Run 2 - Operator B
Day 2, Run 1 - Operator B
Day 2, Run 2 - Operator A
Day 3, Run 1 - Operator B
Day 3, Run 2 - Operator A
Day 4, Run 1 - Operator A
Day 4, Run 2 - Operator B

Since he does not have this latitude, he can only follow the schedule through Day 3.

Assuming that this schedule was followed, it is possible to get some estimation of the effects of training (with practice and propagation conditions balanced insofar as possible). In the example, Operator A therefore has a mean acquisition range of 45,300 yds. while Operator B has an average score of 40,700. If Operator A had the special training, it would appear that it was worthwhile and contributed an effect approximately equal to that of the propagation effect noted.

In summary, it can be said that significant human factors effects can be operating in any test; furthermore, it is possible to estimate their effects even when the total amount of available data is small. While the procedure is admittedly crude, it is about as refined as practical circumstances permit it, and it offers new opportunities to the project officer to determine how the system undergoing test can be brought up to operational demands. For example, assume that a 40,000 acquisition range is required for operational use. One way of insuring that acquisition ranges do not settle around 35,000 yds. is to provide practice (perhaps a training device) or to provide special training. *Both of these options are alternatives to system redesign and can be substantiated by OPEVAL data.*

7.3 SECOND NUMERICAL EXAMPLE

The example discussed on earlier pages was one in which data were scarce and the effects to be considered were many. In the second illustration, we will concentrate on a case where data are still in short supply, but are sufficiently available to allow decisions to be made (and substantiated) on the basis of statistical analyses. Also, the example will serve to illustrate how human factors effects enter into any OPEVAL through the performance of maintenance technicians as opposed to equipment operators.

Consider, for example, that failure report data gathered over a five-day test period show that the times for correcting twenty troubles which arose in each of two subsystems were those listed in Table 7.1. For purposes of simplification, we will assume that the system being studied requires only two subsystems which perform their tasks in sequence. That is, in order to accomplish the mission, *both* subsystems must be operative. Therefore, the system can be inoperative under any of the following conditions:

Subsystem 1 is up but Subsystem 2 is down.

Subsystem 1 is down while Subsystem 2 is up.

Both Subsystems 1 and 2 are down.

The remaining possible state is, of course, that both subsystems are up.

The first question a project officer might investigate is the probability of each *subsystem* being down for maintenance and the *system* being down as well, based upon his observations.* In Subsystem 1, equipment was inoperative

---

* Assume that no delays occur because of a lack of replacement parts or that no time is lost detecting the existence of the trouble.

Subsystem 1	Subsystem 2
65	35
90	25
60	50
125	35
30	50
45	15
80	40
120	30
50	55
5	45
15	25
20	35
70	10
20	5
100	25
10	5
15	10
50	20
5	25
25	5

Total time = 1000 minutes

Average time = 50 minutes

Total time = 545

Average time = 27. 25 minutes

Table 7. 1: Illustrative example of time (in minutes) spent by technicians correcting twenty troubles on each of two subsystems within the system undergoing test.

for 1000 minutes during the five-day or 7200-minute period. Therefore, the proportion of "down" time becomes approximately .14 and, as is frequently the case, this is considered as the equipment failure probability--one value of P_d = .14. For Subsystem 2 the value of P_d is lower, being 545/7200, or approximately .075. Therefore, based upon this small sample of data, estimates of "up" and "down" times for the systems can be given as follows:

	Subsystem 1	Subsystem 2
P_d =	.140	.075
P_u =	.860	.925

Recalling that the system can be inoperative under any of three possible states, the probability of system availability can be summarized as follows:

Probability of both Subsystems being up =	.7955
Probability of Subsystem 1 being up and 2 down =	.0645
Probability of Subsystem 1 being down and 2 up =	.1295
Probability of both subsystems being down =	.0105
TOTAL	1.0000

Probability of system being down =	.2045

The second question of interest is whether or not the differences in down time between Subsystems 1 and 2 are statistically significant. If so, it would appear worthwhile to concentrate efforts on improving the availability of Subsystem 1. Taking the difference between means for the two columns shown in Table 7.1 (50.00 - 27.25 = 22.75) and dividing by the standard error of the difference to obtain a *t* ratio (S_d = 9.173), a *t* value of 2.481 results, which indicates that the amount of "down" time in the two subsystems is indeed different ($P < .05$).

On the basis of the analysis illustrated above, valuable information has been gained by the project officer. Now we will proceed to illustrate how taking account of one particular human factor could serve to present to the project officer a whole set of new alternatives by which subsystem availability could be brought up to desired levels.

Consider, for example, that the first ten troubles in each subsystem were corrected by a man with six months' maintenance experience, while troubles 11 through 20 were corrected by a technician with three years' experience on similar equipment. The question now becomes, "Does experience make a difference and if so, what are the implications of this finding?" Here again statistical analyses can be used to test the question. In the case of Subsystem 1 a t value of 2.203 results which indicates that experience does make a significant difference ($P < .05$). In the case of Subsystem 2, a similar result obtains $t = 4.11$, $p < .005$. Therefore, it is reasonable to infer that technician experience does make a difference throughout the entire system. These differences are shown below in terms of average times to correct a trouble.

	More Experienced Technician	Less Experienced Technician
Subsystem 1	33.0 min.	67.0 min.
Subsystem 2	16.5 min.	38.0 min.

On the basis of these additional analyses, we can see that we might obtain approximately a 50% reduction in average "down" time per trouble in each subsystem if only more experienced men are used. Furthermore, it can be seen that the average time per trouble taken by an *experienced* man in Subsystem 1 is very similar to the average time per trouble taken by a *less experienced* man in Subsystem 2. In short, we begin to see possibilities for "trading off" equipment complexity with technician experience. For example, if we have a limited number of experienced technicians, we might choose to put them to work in Subsystem 1.

One final point is worth noting. *Often during a test an "extra effort" situation prevails with the most experienced technician working on all troubles which arise.*. If this were the case in our numerical example, resulting estimates of the probability of subsystem availability would be biased in an upward direction. That is, it might be possible to perform in an extra-effort manner during a crisis period (with the most experienced man getting little sleep, etc.), but can the maintenance crew be expected to act this way *all the time?* If not, two systems' availability estimates are appropriate; one for short-run crisis periods and the other for normal periods. If the second is not available, it should be indicated that the availability figures given are for short-term crisis performance. *But, if data are not coded to reflect these technician effects, they cannot be noted and used to advantage by the project officer. In fact, these hidden effects become a distinct liability.*

7.4 SUMMARY

Throughout this section numerical data have been used to illustrate the point that by taking account of human factors, a project officer gains additional information about the system in question. That is, at the end of the test he has numerous options open to him concerning how to meet system requirements. This additional advantage is gained by taking the residual variation left in system performance data (after equipment effects have been removed) and by removing from it the portion directly attributable to operator and maintenance technician performance. By this process, it becomes possible to determine more accurately numerous combinations of specifications which can serve to insure that system performance standards are met. Also, by the study of human factors effects, it is possible to be more certain that what is labeled as an equipment effect is in fact an equipment effect.

In short, advantages can be gained by the project officer by taking human factors into account in OPEVALs. However, in order to obtain these advantages, care must be taken to insure that in the process of screening, cataloging, reducing, and analyzing data, human factors effects are *not* overlooked. In other words, what comes out of an analysis is not only a function of the preparation and planning that went into it, but is a function of the data handling procedures as well.

8.0 PREPARING THE FINAL REPORT

The project officer's responsibilities and activities at this phase of OPEVAL are covered in some detail in Vol. I, COMOPTEVFOR Inst. P3930, 1c, 1 April 1961. For that reason they need not be reiterated here. What should be mentioned, however, is that *all of the intricacies involved are concerned with a single goal: arriving at definitive conclusions substantiated by data gathered during the OPEVAL.* Therefore, attention will be centered upon how human factors studies in OPEVAL might serve to modify, influence, and assist the project officer's basic task.

8.1 USE OF THE TOLERANCE LIMIT NOTION

Throughout this report, it has been stressed that while project officers rarely have the opportunity to study human effects in considerable detail, they can usually gather important data concerning how much variation in system performance can be anticipated as the result of human components. To illustrate how these data lead to conclusions, the hypothetical data used in Sections 7.2 and 7.3 will be employed.

Assume that one of the requirements placed upon a SAM system is that it must be able consistently to acquire very high altitude targets at a range of, 40,000 yds. Within the range of propagation differences employed in the test, it appears that the following estimates can be derived: first, the lower limit of system performance which is likely to occur, using operators with little practice in searching for and handling such targets, is approximately 35,000 yds. Second, by using an operator without specialized training but having the advantage of practice, it appears that acquisitions can be made at or above the required standard of 40,000 yds. Third, by using an operator who has specialized training but little practice, acquisitions will probably be made at considerably less than 40,000 yds. Finally, by using an operator with specialized training *and* practice, target acquisition ranges can be increased to approximately 50,000 yds.

Thus, a graphic depiction of how the range of human effects center about the desired acquisition range can be shown below, where X_L indicates the lower limit of human effects and X_H the higher limits.

Range of human variation

X_L X_H

10 20 30 40 50 60

Acquisition Range in Thousands of Yards

If this range is subdivided into a below-standard, standard, and high-standard section, the following descriptions are appropriate:

Acquisition Range	Appraisal	Conclusion
35, 000 to 40, 000 yds.	inadequate	Without practice even specially trained operators fall short of the standard.
40, 000 to 45, 000 yds.	marginal	With practice a man can make this range even without special training.
45, 000 + yds.	adequate	Special training *and* practice are required to attain this goal.

All of the above evidence indicates that in order to be reasonably assured of successful attainment of the 45,000-yard objective, both practice and specialized training are required on the present system. Perhaps, by slight job modifications and a different approach to the task, this average acquisition range for high altitude targets can be extended well beyond the 40, 000-year standard. Answers to questionnaires provided by the test ship's crew might be of considerable help in isolating possible points of modification. In this manner, statements and conclusions remain definitive and can be substantiated by data.

A second specific requirement might well be named, e. g., that system availability be greater than 90%. By using illustrative data introduced in Section 7. 3, it can be seen that the system will *not* reach this standard. New tolerance limits can be drawn to indicate the range of performance available simply as a result of considering human factors. First, it is apparent that by using both highly experienced *and* relatively less experienced men in both subsystems, it is doubtful that availability will surpass 80%. Second, if we choose to use only highly experienced men on Subsystem 1 with the mixed experience group on Subsystem 2, this figure can be increased to approximately 83%, which still falls considerably short of the desired goal. Third, if we used only less experienced technicians in both subsystems, availability is likely to plunge to less than 80%--the results of which would be disastrous. If we use *only* highly experienced men on both subsystems, equipment availability reaches approximately 85%. *In short, no matter how well experienced the technicians are, the system availability goal cannot be reached.* Ways of simplifying the job of maintenance should be studied thoroughly by using the results of questionnaires to suggest means of improvement.

Consider, for a moment that in the above example an 85% subsystem availability is the standard instead of 90%. From our analyses, we can see that this goal can only be obtained with highly experienced personnel. If this is the case, an inquiry directed to the Bureau of Naval Personnel's (personnel research specialists) would be in order to see whether personnel with these qualifications could be provided in sufficient numbers to meet the entire Fleet demands. If the number of such men falls short of what is required, we again conclude that improvements must be made to bring the system up to the 85% standard. In short, while it may be theoretically possible to reach the goal under certain special conditions, practical contingencies might preclude meeting the necessary conditions.

ed

8.2 USE OF QUESTIONNAIRE DATA

As was shown in the preceding examples, the tolerance limit notion is useful in deciding whether or not the system can attain its stated requirements. But when it comes to suggesting how to correct existing deficiencies, questionnaire data from the test ship complement are extremely useful. As was indicated earlier, it is possible by means of questionnaires to isolate especially difficult and time consuming aspects of both operator and technician jobs. At the same time, it is also possible to acquire suggestions concerning how present difficulties can be overcome.

With information of this sort available, the project officer is able to pinpoint what aspect of a man's job should receive greatest attention (when corrective measures are needed). Also, he has some suggestions of how to go about reducing existing difficulties. Naturally, the project officer's personal observations and insights are of considerable assistance also.