INCORPORATING HUMAN PERFORMANCE RELIABILITY DATA IN SYSTEM RELIABILITY MODELS

Joseph P. McCalpin

AAI Corporation

Prepared for:
Human Engineering Laboratory

21 December 1972
INCORPORATING HUMAN
PERFORMANCE RELIABILITY DATA
IN SYSTEM RELIABILITY MODELS

By
J. P. McCalpin

Prepared For
Infantry Team, Systems Performance and Concept Directorate
Human Engineering Laboratory
Aberdeen Proving Ground, Maryland 21005

By
AAI Corporation
Baltimore, Maryland
Under Contract No. DAAD05-72-C-0003
DDC Availability Notice

Destroy this report when no longer needed. Do not return it to the originator.

This document has been approved for public release and sale; its distribution is unlimited.

The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents issued and approved by the Department of the Army.
INCORPORATING HUMAN PERFORMANCE RELIABILITY DATA IN SYSTEM RELIABILITY MODELS

ER-7312

Report No.

December 1972
Date

Approved for public release. Distribution Unlimited.
## TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II. PROCEDURE</td>
<td>3</td>
</tr>
<tr>
<td>A. Weapon Reliability Model</td>
<td>3</td>
</tr>
<tr>
<td>B. Weapon Maintainability Model</td>
<td>8</td>
</tr>
<tr>
<td>C. Resume of Previous Human Performance Reliability Research</td>
<td>11</td>
</tr>
<tr>
<td>D. Establishing a Set of Minimum Requirements for a Human Performance Reliability Program</td>
<td>15</td>
</tr>
<tr>
<td>III. A PROPOSED APPROACH TO ESTABLISHING A HUMAN PERFORMANCE RELIABILITY PROGRAM</td>
<td>18</td>
</tr>
<tr>
<td>IV. SUMMARY</td>
<td>22</td>
</tr>
</tbody>
</table>
An investigation was conducted to determine the minimum requirements for a human performance reliability program which would provide data for system reliability models. Systems reliability models for a machine gun were reviewed and it was concluded that, although no specific provisions were made in the model, human error data could be processed. The state of the art in human performance reliability was reviewed to determine the required components of a human performance reliability program. It was determined that adequate theoretical constructs are available, a suitable data collection vehicle exists and the general procedures to institute such a program are available.
Human Factors Engineering

Human Performance Reliability

Reliability Models
I. INTRODUCTION

Recent military development programs have been emphasizing newer and more sophisticated management tools for decision making. Management concepts such as risk assessment have required program managers to define performance and program objectives more rigorously. The development of quantitative tools for management has placed greater emphasis on precision in predicting reliability, durability and maintainability in emerging systems.

The development of quantitative projective management tools has enabled older goals to be examined in greater detail and re-emphasized with more meaning. Today the materiel development agencies find themselves with an increasing number of programs and requirements, and a decreasing R&D budget. In addition, certain programs have shown that early mistakes can remain undetected until quite late in the development.

One of the requirements in a well managed weapon system development program is human performance data. The purpose of this paper is to address the problem of acquiring useful data on human performance reliability. To accomplish this purpose, three goals have been established. First, human performance reliability will be defined. Second, minimum requirements for the assessment of human performance reliability will be established. Third, a method for incorporating human performance reliability into system reliability models will be provided.

Systems engineers often define a system as a combination of equipment, personnel and procedures which perform a specific set of functions. This definition of a system permits us to view system reliability as the integration of equipment reliability and human performance reliability.
Many authors on reliability tend to establish reliability models which treat only equipment failure rates. The presumption is that error contributions made by man in the system will be reflected in equipment failure rates. The problem with this treatment of human performance reliability is that human errors are treated at the component level, and the rather complex interfaces which can result in modern weapon systems are not adequately treated.

The method used in this paper was to study two recent system reliability models. These reliability models, developed by Bazovsky and Associates (1970a) for predicting reliability of a machine gun, were evaluated to answer the following questions:

a) Does the model consider the human error component of reliability assessment?

b) If the model does not address human error, what changes are necessary to allow for this consideration?

A companion maintainability model proposed for a machine gun, also developed by Bazovsky (1970b), was evaluated to ascertain if that model treated human performance error data.

Relevant research in the area of human performance measurement was also reviewed to identify the minimum requirements for a human reliability assessment, and to establish a standard for comparing these requirements with the reliability and maintainability assessment models indicated above.
II. PROCEDURE

The Bazovsky models cited above were developed for effectiveness assessment of a machine gun. To that end, the models specifically address a nonredundant mechanical system and could be generalized (with only minor modification) to any automatic gun.

A. Weapon Reliability

The Bazovsky models for reliability assessment are mathematical methods for predicting reliability of machine guns. One model is based upon a Weibull distribution. The other is based upon an Erlangian distribution. The models assume that the replaceable items are mechanical subassemblies or components.

The block diagram (Figure 1) was the basis for both models; the diagram was developed from a system functional block diagram closely corresponding to the Work Breakdown Structure for Ordnance Systems defined in MIL-STD-881.

In the models, each of the equipment groups is broken down into components. Using the functional diagrams and a Failure Modes and Effects Analysis (FEMA), Bazovsky identifies the failure modes, probable result of failure, effect of failure, corrective action and recommendations to improve the reliability.
Figure 1. Reliability Block Diagram
(Bazovsky (1970a) p. 18)
In developing the model mathematics, Bazovsky makes the quite valid point that reliability is a decaying function based upon how the weapon has been used.* To account for this fluctuation in reliability, Bazovsky introduces the concept of renewal which permits the estimation of reliability as a function of time and renewal rate. Reliability is expressed as a function of the integrated failure rates of the components. The details of the mathematical expressions can be obtained from the referenced paper.

These models represent techniques for integrating failure rates to predict reliability regardless of the cause of failure. The model does not care whether the operator broke a component, or if the part failed "normally". However, the fact remains that any reliability estimate which does not expressly treat failure probabilities introduced by the human component is not complete. The reason for taking this position is that the man-machine interface is quite complex. Models using only component failure rates will never provide insight into errors caused by complex procedures, pressures of time, etc. Many operator induced malfunctions are not related to component failures. Therefore, as Meier (1971) said, "To ignore human performance in reliability estimates is to assume that operator performance is invariably perfect .... since it is apparent that operator performance is rarely - if ever - perfect, equipment reliability estimates that fail to take account of human performance turn out to be grossly inflated."

* When a weapon is first manufactured its reliability is close to that predicted; after the weapon is used, the reliability may decrease. Once the weapon has failed and is repaired, a new reliability estimate will apply.
If this reliability model is going to include the human component, it must be modified to include a new set of interfaces; the new model would consider all personnel tasks; a block diagram of such a model is shown in Figure 2. To collect failure data for this model, each man-equipment interface must be defined for each equipment group. Defining the man-machine interfaces for each equipment group includes task definition in terms of input (stimulus), processing, output (performance), time requirements, skill requirements, and performance standards. The human error potential is then identified. Each of these potential errors must then be included in the Failure Modes and Effects Analysis. In cases where the results of human error critically affect system performance, failure rates must be estimated. The implications of this procedure are discussed in Section III of this report.
Figure 2. Reliability Block Diagram Modified for Considering Human Performance
B. Maintainability Model

The maintainability model developed by Bazovsky and Associates (1970b) is a mathematical model for the prediction of maintainability characteristics of a machine gun. The applicability of this model is restricted to machine guns with little capacity for generalization.

The maintainability prediction starts with an analysis of the maintenance functions which accrue to each of the four equipment groups shown in Figure 1. Bazovsky describes the maintenance functions by means of functional flow block diagrams (see Figures 3 and 4). The examples shown in the figures are only taken to the second level; this analysis must be continued to successively lower levels to identify the maintenance tasks. Frequency of action and time required to accomplish the task are calculated for each maintenance function.

The mathematics employed by Bazovsky are similar to those employed in most other maintainability models and are fully described in the reference.

This model does not address the potential error contribution of man in the system. This model does, however, provide the hierarchal framework for a major part of the task analysis information required for a human reliability analysis. Beyond this point, the maintainability model will not be of assistance in assessing human performance reliability, since we are not concerned with time to repair, only the probability of error in the repair task.
Figure 3. Maintenance Functional Flow Block Diagrams
First Level Function 2.0 (From Bazovsky 1970b)
Figure 4. 2nd Level Functions for 2.1 and 2.2 Functions (From Bazovsky, 1970b)
C. Resume of Previous Human Performance Reliability Research

The human performance reliability aspects of human performance measurement began as an off-shoot of the ICBM programs. The problem of human error in maintaining an "up" system was discovered to be a significant factor in system availability. Early research described by Rook (1962 & 1964) addressed the problem of human performance aspects of reliability in a production line. Human performance reliability aspects of other military weapon systems were investigated by Knoweles (1959) and Miller et al (1964). Major symposia were conducted on the problem of measuring human reliability as early as 1961 by the Ballistic Missile Division of the Air Force (Majesty 1962). A summary of early human reliability research is presented in Swain (1964).

The problem was well recognized, but until 1964 there was little structure to the research, and as a result only a small portion of the work could be generalized for wider application. In 1964 a symposium was held at Sandia Corporation. The results of that symposium were presented in the December 1964 issue of Human Factors.

From the mid-sixties to the present, the bulk of the human performance reliability research has been concentrated in three areas - the construction of data stores, improvements in modeling techniques and development of standardized methodologies for human performance reliability analysis. Construction of a data store or bank has been, from the outset, one of the major goals of human performance reliability research. Payne and Altman (1962) and Altman (1964) traced the development of the AIR DATA STORE to provide a
general list of tasks and task elements with corresponding failure rates. This work has been applied to many programs, and the shortcomings of this device have been widely discussed in human factors literature. The construction of a data store is a task of epic proportions when one considers the problems of validity; differences in environment, skill levels, and personnel; as well as problems of generalizing from system to system and task to task. Despite all the limitations of data banks, they represent one of the most significant repositories of human error data available. Other attempts at constructing data stores have been made since 1954. Of particular interest are the Sandia Human Error Rate Bank (SHERB) developed by Sandia Corporation (Swain 1969) and the TEPS (Technique for Establishing Human Performance Standards) data bank developed by Blanchard and Harris (1967). The very existence of the several data stores indicates the magnitude of the problems associated with establishment of a general failure rate data bank.

Another area of work between the mid-sixties and the present was the area of modeling. The early works of Swain and Wohl (1961) and Miller et al (1964) represent early modeling efforts in the area of electronics operation and maintenance. Models are developed for specific systems and/or types of systems and their capability for generalization is limited by the structure of the man-machine loop. Several authors, notably Swain (1969) and Meister (1971), have suggested that a uniform methodology to establish error rates is more important than a model to establish human reliability.
Research in the area of developing uniform methodologies for estimating human error started at a simple empirical level, where tasks were performed in a laboratory situation and validated by replication. Correlational techniques were used to determine the variability in error rates from task to task. A description of these efforts can be found in Gagne (1962). One of the major efforts in developing a Technique for Human Error Rate Prediction was performed by Swain (1964b). This technique, called THERP, consisted of the following five steps:

a) Define the failure
b) Define human operations and the man-machine interface
c) Predict error rates for each human operation
d) Determine the effect of failures
e) Recommend changes to reduce failure rates to an acceptable level.

The technique utilized branching or graph theory to estimate the combinations of the conditional probabilities resulting from performing an operation. The technique has merits in that it is objective and system oriented to give answers to engineers and program managers. This technique does, however, presume the existence of a data store to provide failure rates; in this case the AIR DATA STORE. The prediction was subject to the rule of GIGO (garbage in/garbage out); hence, its utility was limited.

Rook (1962) and Irwin (1964) used a methodology of rating tasks to provide estimates of failure rates when no data were available from data stores. The rating technique was adapted from the work of the early psychophysicists, especially from the techniques developed by Thurstone. These techniques and their logical support are described in a paper by Leuba (1964).
The mechanics of employing such techniques can be found in a paper by Meister (1964). These techniques are based upon valid procedures developed in measuring other aspects of human performance.

Recent efforts have produced a method for developing a series of scales for human performance based initially on the rating process. Rating of tasks is used to establish an interval scale of error likeliness; empirical data are used to establish a ratio scale using a computer application of Monte Carlo techniques. This methodology is most important, since it provides a mechanism for adding greatly to human performance data without establishing a requirement to collect a large volume of data under complex experimental conditions. In short, this mechanism makes a human performance reliability assessment feasible from a cost standpoint. Swain (1969) has done extensive research and has described the promise of this technique. The benefits of this methodology include being oriented towards a specific system, not requiring a large volume of empirical data and, ultimately, the generation of a large data bank that can be applied to other programs.
D. Establishing a Set of Minimum Requirements for a Human Performance Reliability Assessment

The models described in Sections A & B above provide an adequate mathematical framework for a reliability and maintainability assessment of the hardware elements of a machine gun system. Since any model is limited in scope to the input information, the problem becomes one of providing accurate and valid failure rates for the tasks performed by the operators and maintenance crews. Even a cursory review of the literature suggests that this is a major effort.

A vehicle exists for providing these data. Contract Data Item DI-H-1334 was developed by HEL and is being used on certain current Army programs. This data item provides for a set of minimum standards in reporting Human Factors Engineering (HFE) tests. The data item was developed to provide objective data on human performance early in the system design to avoid the proliferation of critical system performance failures due to unattainable human performance requirements. The information required by this data item is sufficiently detailed and structured to impose constraints on the design of HFE tests and the collection of data. Because of these features, this data item can also be used to collect data on human performance reliability. Data reported in the format and level of detail of DI-H-1334 will be of value not only in answering the immediate question of predicting human performance reliability on a system, but also will provide standardized data for a human performance data bank.
Having established that the Bazovsky model is adequate for including human error information, and that an adequate vehicle for collecting this information is also available, it is now necessary to identify a set of requirements which must be met to establish a human performance reliability program. These requirements include the following:

a) Standardize task descriptions through a task taxonomy
b) Develop classification schemes for errors
c) Develop scaling techniques for describing the "error likeliness" of human performance
d) Collect empirical data on task performance
e) Develop error rate curves for tasks using gaming techniques such as Monte Carlo
f) Apply human error data to system reliability models
g) Develop a human error data bank for future reference.

A task taxonomy is a standardized procedure for the classification of behavior. If a scaling process is going to be developed, this classification process is necessary to establish relationships between tasks. Examples of task taxonomies can be found in Miller (1962) and Gagne (1962).

The task at hand is to evaluate all classifications and select a single scheme which meets the Human Performance Reliability Program requirements.

Error classifications have been proposed by Swain (1969), Blanchard & Harris (1967) and Altman (1964). Great care must be taken to use a classification scheme which meets the needs of the system. For example, when considering a system reliability model for a mechanical system, it is clear
that only those errors which will result in a mechanical failure need be considered. A principal requirement for an error classification system is a failure mode and effects technique which will provide a full description of all errors.

To develop interval scales for measuring error likeliness (probability of error), these rankings are repeated until a scale is developed where the distance between each point on the scale appears to be at equal intervals from the other points. This scale will provide an interval measurement of the rated probability of error.

The data collection methods described in DI-H-1334 will provide a source for empirical data. One of the purposes of using the rating/scaling approach is that the requirement for an instant data store does not exist. Maximum use of experimental data will be possible using the rating system and rate curves.

If sufficient controls are imposed on the collection of human performance data, error rate curves can be developed using empirical data from the HFE test reports (DI-H-1334). The reasoning is that if an accurate estimate of one error can be made on an equal interval scale, then all other error rates can be derived. The new scale will be a true error rate scale which can be refined into a ratio scale using computer simulation techniques. The refined error rates can be inserted into an equipment reliability model such as the Bazovsky model to provide an estimate of system reliability. As error data become available on a larger number of systems, a data bank for human reliability prediction will become feasible.
III. APPROACH TO ESTABLISHING A HUMAN RELIABILITY PROGRAM ON ARMY WEAPON DEVELOPMENT PROGRAMS

Any program established for the assessment of human performance reliability on Army weapon development programs must satisfy two objectives. The first and most critical is that the data resulting from the assessment have immediate impact on weapon system design issues. The second and most far-reaching is the establishment of a human reliability data bank which will affect future systems design activities by pointing up design problems in the concept development stage.

There are also restrictions that must be placed upon a human reliability assessment program. The first restriction is concerned with data. The requirement for empirical data must be kept to a minimum to avoid proliferating data and raising development costs. Another restriction is that all human performance reliability predictions be placed within the general context of system reliability and effectiveness assessment. By maintaining this perspective, it is more probable that design changes proposed for improving the human component of reliability can be expressed as trade-offs rather than modifications to design requirements.

The proposed approach to the assessment of human reliability will meet the objectives and satisfy the restrictions indicated above. Since the proposed approach is stated in terms of system reliability models (i.e., the Bazovsky models), the reliability estimates can be compared directly to the system specification requirements. The proliferation of data is avoided by using an existing Contract Data Item (DI-H-1334) for the collection of reliability data. The use of this data will permit standards for data
to be set and thus increase the applicability of data from one program to another. The result of standardizing data will facilitate the construction of a human performance data bank.

The approach to assessing human reliability is shown in a simplified form in Figure 5. The block diagram shown does not contain iterative loops but rather describes a straight-through problem.

The system reliability model can be developed from the system functional diagrams and from the work breakdown structure. The standards for task analysis should include a task taxonomy, i.e. standard language, as discussed above. (In addition to the above, these standards include performance requirements and skill levels needed for performance of a task.) Operations tasks can be identified from the sequence of operations and the interface points between the operator and the equipment. Maintenance tasks can be obtained from the maintainability analysis conducted to estimate mean-time-to-repair.

Developing an interval scale of error likeliness is achieved by means of rating the tasks. This function can best be performed by assembling Army technicians who have performed similar tasks on other systems. The list of tasks would be given to the technicians with instructions to rate the tasks in terms of the "likelihood" of errors expected. Once a ranking is developed, the raters are asked to estimate whether one task is twice as error prone as another. Each list of tasks is rated a number of times to validate each rating. The result of this rating is a scale of likeliness of error where each interval is a fixed distance from the next. This rating is most critical to the success of the entire operation, and it is proposed that task
Figure 5. Block Diagram of Proposed Human Reliability Activity
ratings be validated during the early stages of the human performance reliability effort. This validation could be performed by having engineers and human factors specialists assist the technicians.

The HFE studies required by the contract will be conducted and error data obtained. If these tests do not supply sufficient data, then simulation studies can be run to establish empirical error rates for some of the tasks. The empirically derived error rates are then applied to the interval scale of error likeliness. Since the size of the intervals is known and the position of each task on the scale is known, it is possible using gaming techniques to establish probability density functions for the error rates of each task. These probability density functions can be applied directly to the system reliability model.
IV. SUMMARY

The purpose of this study was to investigate the capacity of a system reliability model to assess adequately the contribution of human performance to system failure. Two models developed by Bazovsky and Associates were reviewed to determine if they accounted for human performance. The first model addressed reliability, the second maintainability. The reliability model made no specific provisions to account for human performance errors. However, if a methodology existed to provide human performance failure data, the model would be able to encompass such data. The maintainability model was based on human performance, but made no specific provisions to handle failures introduced by human performance. This condition is not particularly unusual, since maintainability models do not normally make such provisions.

One of the most important prerequisites of a program to assess human performance reliability is a standardized vehicle to collect and process human performance data. A recently developed Contract Data Item Description (DI-H-1334) was evaluated to determine if it would serve as that vehicle. DI-H-1334 was developed to provide guidance on the preparation of Human Factors Engineering Test Reports. This data item will also serve the purpose of collecting human performance reliability data because it requires the controls and standards necessary to assure the quality of human performance data. It should further be noted that use of this data item avoids the proliferation of data by raising the standards of quality for human performance data.
Relevant research in the area of human performance reliability was reviewed to determine the state of the art and to define a minimum set of requirements for assessing human performance reliability. The following steps were proposed to establish such a program:

a) Standardize task descriptions
b) Develop a classification scheme for human errors
c) Develop scaling techniques for describing the likeliness of error
d) Collect empirical data on task performance
e) Develop error rate curves based upon error scales and empirical data
f) Apply derived error rates to system reliability models
g) Develop a human error data bank for future reference.

Assessment of system reliability is possible only if all aspects of the system are described in the model. The failure to account for human performance in system performance will result in optimistic estimates of system reliability particularly in the area of allocating failure rates to system elements. The framework exists for assessing human performance reliability; the requirement to apply this type of reliability data to system reliability models needs to be included in Army Weapon Development programs.
REFERENCES


Knoweles, W., Automation and personnel requirements for guided missile ground support functions, USAF, Wright-Patterson AFB, Ohio, WADC TR-59-240, 1959.


Rook, L. W., Predicting human error rates with the simple multiplicative model. As reported in Meister (1964) op. cit.

