Report LRP 317/TDR-63-218 May 1964

HUMAN RELIABILITY IN THE PERFORMANCE OF MAINTENANCE

Report LRP 317/TDR-63-218

May 1964

Contracts AF 04(694)-424, S.A. 24
and AF 04(647)-521

Prepared by

Irl A. Irwin
Joel J. Levitz
Alvyn M. Freed

AEROJET-GENERAL CORPORATION
Liquid Rocket Operations
Sacramento 9, California

Prepared for

BALLISTIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
Air Force Unit Post Office
Los Angeles 45, California

AEROJET-GENERAL CORPORATION 329T
A SUBSIDIARY OF THE GENERAL TIRE & RUBBER COMPANY

ABSTRACT

A method for estimating the reliability of maintenance performance is developed and applied to tasks involved in scheduled maintenance for Titan II engines. The approach involves the combined use of ratings and empirically derived reliability figures. A modification of the design engineer's redundancy formula is developed for estimating the increase in human reliability achieved when two mechanics work together in the performance of a single maintenance task.

This study demonstrates that highly consistent ratings of task-element-reliability can be obtained from groups of qualified raters.

Plans for validating the human reliability estimates obtained during Category II testing at Vandenberg Air Force Base are described. Suggestions for further research and application of the findings are given.

TABLE OF CONTENTS

APPENDIX

TABLE LIST

FIGURE LIST

I. INTRODUCTION

The objective of this study was to develop a means of predicting personnel effectiveness during scheduled checkout and maintenance activities performed on the Titan II Propulsion System. It was recognized that the inherent capability designed into the Titan II Propulsion System could not be realized fully unless due consideration was given to the performance of the personnel who service and maintain the system.

In the development of aerospace systems, the United States Air Force considers the human component as the personnel subsystem. The personnel-subsystem concept enables the Air Force to pursue a development and test program for the human component similar to that for the hardware components. In the past, reliability predictions for personnel performance were not employed in the development of weapon systems, and decisions regarding these activities were based on generalized ratings such as "good," "average," "fair," or "poor." In their efforts to provide a highly effective man/machine system, design engineers, technical writers, and training personnel were hampered by the vague nature of such performance descriptions.

II. SUMMARY

Quantitative reliability information was available for hardware components so that decisions and system integration could be made in a technically sound way. Designers, for example, were able to choose gages, valves, controls, and similar components by examining their reliability as well as other pertinent characteristics. Efforts were made to develop techniques for predicting and evaluating the reliability of the human component in a manner similar to that which proved so useful in hardware research and development.

The availability of quantitative human reliability information should aid decision making in each of the following areas:

1. Designing the propulsion system and supporting equipment for ease of operation and maintenance.

II, Summary (cont.)

2. Providing clear and effective Checklists and Technical Orders.

3. Providing effective inputs to the training of Titan II Missile Engine Mechanics.

While actions in these three areas will make a substantial contribution toward achieving the highest possible human reliability under operational conditions; this reliability is clearly a function of additional factors, including such critical determinants as attitudes, feelings, and motivations of Air Force mechanics.

This report describes the development of a method for estimating the reliability of maintenance performance. The reader should bear in mind that the method has not been validated using Titan II operational performance data. Until the results of present validation efforts are available, the performance reliability estimates should be cautiously applied as they were derived from subjective judgements and limited empirical data. Because the empirical data were obtained relatively early in the research and development cycle, it is highly probable that the operational-reliability figures will be higher than those reported. Even at that early date, the task failure rate of 14.4% was close to the 10 to 15% figure typical of the defense electronics industry as estimated by the American Institute for Research (Ref 1).

A. METHOD

A promising method for estimating the reliability of maintenance performance has been developed and applied to Titan II Propulsion System scheduled maintenance activities. The method involves the following steps:

1. Specify the maintenance tasks to be performed, e.g., service oil sump of turbopump gear box, perform functional check of thrust-chamber valves, perform leak check of fuel system, etc.

II, A, Method (cont.)

2. Identify the task elements that must be performed to accomplish the total tasks, e.g., verify switch position, connect flexible hose, read time (Brush Recorder), install lockwire, etc.

3. From judges familiar with maintenance tasks and typical Air Force mechanics, obtain rating for likelihood of error in performing the task elements. In the present study, the 33 rating judges were 18 Aerojet-General Personnel Subsystem and Reliability Engineers, 6 Aerojet-General technicians, and 9 Vandenberg AFB missile-engine mechanics. More consistent ratings were obtained from the engineers than from either the technicians or the mechanics. The degree of inter-rate agreement among the 18 engineers was very high. Only ratings from the engineers were used in the subsequent analyses.

4. Obtain empirically based reliability estimates for at least some of the task elements. In the present study, 29 empirically based reliability estimates were derived from the Payne-Altman Index of Electronic Equipment Operability (Ref 1). The Payne-Altman figures are extrapolations to field conditions based on the results of laboratory studies available in the experimental literature. Using results of the Titan II Personnel Subsystem Test and Evaluation Studies, conducted in a simulated silo in Sacramento, a new conversion factor as developed and applied to the index data to make experimental findings more compatible with Titan II operating conditions.

5. Prepare a scatter diagram for the task elements for which both ratings and empirically based reliability estimates are available. Fit a regression equation to the data and derive reliability estimates for the task elements for which only ratings are available (Table 9). In the present study, the correlation of the task-element ratings and modified reliability estimates of the Index of Electronic Equipment Operability was .457, indicating a significant relationship. In deriving reliability estimates, a logarithmic curve was applied to the data because it provided a better fit rather than a linear regression line and was consistent with previous findings in the study of rating phenomena.

II, A, Method (cont.)

6. If a second mechanic will be available to assist the first one in the performance of the task, adjust the task element reliability estimates to take account of this redundancy.

7. Working from a detailed set of procedural instructions, determine the task elements involved in performing each maintenance task. Record the appropriate reliability estimate for each task element.

8. Determine the Task-performance reliability by computing the product of the separate task-element reliabilities.

The method has logical validity, and its results are supported by the limited empirical evidence presently available. Task-element reliabilities derived by this method are similar to some empirically based reliability figures acquired from records of aircraft maintenance. While such evidence is recognized as limited in quantity, its confirmatory nature strengthens confidence in the method described and justifies a further investigation and a possible increase in range of application. The final test of the method will involve a comparison of the present predictions with the human reliability measures obtained during Category II testing at Vandenberg Air Force Base. Plans for such a validation study are described. The report is concluded with suggestions for further research.

B. RESULTS

The results of the study are presented in Table 1 and consists of the human reliability estimates for the 64 scheduled maintenance tasks of the Titan II engine. The lower figure in Table 1 is the reliability estimate for the task when individually performed. The higher figure is the reliability for the task when two mechanics are present in the silo and redundancy is considered. Tasks having both relatively low predicted human reliability and high criticality should receive priority attention. Tasks that are among the lowest third in redundant reliability

II, B, Results (cont.)

of all schedule maintenance tasks and having a criticality rating of 3 are starred (*) in Table 1. To aid in interpreting the significance of the reliability estimates, criticality ratings from Type II Basic Data are included in Table 1. The criticality code is defined below.

It should be noted that the figures in Table 1 are estimates of the reliability of the personnel subsystem and are not estimates of the reliability of the propulsion subsystem developed by Aerojet-General. The reliability of the propulsion subsystem is considerably higher than the estimated reliability of the personnel subsystem.

In addition, the figures should not be interpreted to mean that any indication of the lack of personnel reliability entered in Table 1 will result in weapon-system failure or extreme safety hazard. Many of the performance errors will be less critical, resulting only in increased maintenance time or excessive consumption of spare parts.

Criticality Code	Definition
2	Tasks critical for subsystem operation that may result in some system degradation if not correctly performed. Such tasks fall into the category of things that affect equipment that is nice to have or that the system would be more effective if the task had been performed correctly but the mission can succeed by using alternate modes, manual concontrol.
3	Tasks that are critical to system operation must be performed correctly. If not performed correctly, they may prevent the system from working or reduce operational effectiveness to an unacceptable level.

TABLE 1

ESTIMATED RELIABILITY OF THE PERSONNEL SUBSYSTEM PERFORMANCE OF TITAN II MAINTENANCE TASKS

		Human Reliability		
Performance Task	Stage	Individual	Redundant	Criticality
Prepare for Leak Check of Thrust Chamber and Lines Below Thrust-Chamber Valves	II	.6712	.7322	2
Prepare for Leak Check of Thrust Chamber and Lines Below Thrust-Chamber Valves	I	.6736	.7422	2
*Prepare for Functional Check of Thrust-Chamber Valves	II	.6856	.7561	3
*Perform Functional Check of Thrust-Chamber Valves	I	.6933	.7658	3
Prepare for Leak Check of Subassembly Hot-Gas System and Turbine Seal	I	.7014	.7657	2
*Leak-Check Installation of Pressurization Kit	II	.7526	.8228	3
*Leak-Check Installation of Pressurization Kit	I	.7531	.8219	3
*Perform Leak Check of Hot-Gas System and Turbine Seal	II	.7644	.8523	3
*Perform Leak Check of Turbopump Oxidizer Gearbox Seal	II	.7678	.8240	3
*Perform Leak Check of Turbopump Fuel Pump Seal	II	.7680	.8194	3
*Perform Leak Check of Turbopump Fuel Gearbox Seal	II	.7761	.8242	3
*Perform Electrical Check	I	.7838	.8361	3
*Perform Leak Check of Turbopump Oxidizer Pump Seal	II	.7906	.8403	3
*Perform Leak Check of Thrust Chamber and Lines Below Thrust-Chamber Valves	II	.8228	.8699	3
*Perform Leak/Functional Check of Thrust-Chamber Pressure Switch	I	.8234	.8809	3
*Perform Leak Check of Subassembly Hot-Gas System and Turbine Seal	I	.8271	.8734	3

TABLE 1 (cont.)

		Human Reliability		
	Stage	Individual	Redundant	Criticality
Prepare for Leak Check of Hot-Gas System and Turbine Seal	II	.8276	.8696	2
*Prepare for Leak Check of Subassembly Turbopump Fuel Pump Seal	I	.8337	.8852	3
*Perform Leak Check of Thrust Chamber and Lines Below Thrust-Chamber Valves	I	.8402	.8820	3
*Perform Leak Check of Subassembly Turbopump Fuel Pump Seal	I	.8450	.8762	3
*Prepare for Leak Check of Subassembly Turbopump Oxidizer Pump Seal	I	.8452	.8920	2
Prepare for Leak Check of Subassembly Fuel System	I	.8457	.8928	2
*Perform Leak Check of Subassembly Fuel System	I	.8465	.8814	3
Perform Leak Check of Subassembly Turbopump Fuel Gearbox Seal	I	.8513	.8941	3
Perform Leak Check of Subassembly Turbopump Oxidizer Gearbox Seal	I	.8513	.8941	3
Perform Electrical Check	II	.8553	.8918	3
Perform Leak Check of Subassembly Turbopump Oxidizer Pump Seal	II	.8650	.9146	3
Prepare for Leak Check of Subassembly Turbopump Oxidizer Gearbox Seal	I	.8750	.9163	2
Prepare for Leak Check of Turbopump Fuel Pump Seal	II	.8740	.9163	2
Prepare for Leak Check of Subassembly Turbopump Fuel Gearbox Seal	I	.8751	.9098	2
Prepare for Leak Check of Subassembly Oxidizer System	I	.8775	.9211	2
Prepare for Leak Check of Turbopump Fuel Gearbox Seal	II	.8785	.9137	2
Prepare for Leak Check of Turbopump Oxidizer Pump Seal	II	.8801	.9134	2

TABLE 1 (cont.)

		Human Reliability		
	Stage	Individual	Redundant	Criticality
Prepare for Thrust-Chamber Valves Functional Check	I	.8811	.9069	2
Prepare for Thrust-Chamber Valves Functional Check	I	.8841	.9085	3
Perform Leak Check of Subassembly Oxidizer System	I	.8899	.9206	3
Perform Leak Check of Oxidizer Autogenous System Between Interface and Superheat Burst Diaphragm	I	.8914	.9207	3
Prepare for Leak Check of Turbopump Oxidizer Gearbox Seal	II	.8916	.9253	2
Perform Leak Check of Fuel System	II	.8962	.9264	3
Perform Engine Electrical-Conduit Leak Check	II	.8993	.9303	3
Prepare for Electrical-Conduit Leak Check	II	.9032	.9312	2
Prepare for Electrical-Conduit Leak Check	I	.9130	.9388	2
Prepare for Leak Check of Oxidizer Autogenous System Between Interface and Superheater Burst Diaphragm	I	.9051	.9357	2
Perform Engine Electrical-Conduit Leak Check	I	.9083	.9318	3
Prepare for Thrust-Chamber Switch Leak/Functional Check	I	.9084	.9350	3
Prepare for Leak Check of Fuel System	II	.9085	.9361	2
Prepare for Leak Check of Oxidizer System	II	.9095	.9376	2
Install Gearbox Pressurization Kit	II	.9198	.9455	3
Perform Leak Check of Oxidizer System	II	.9200	.9436	3
Prepare for Electrical Check	I	.9210	.9449	2
Prepare for Pressure-Decay Check of Turbopump Gearbox	II	.9212	.9444	2
Install Gearbox Pressurization Kit	I	.9244	.9469	3
Perform Pressure-Decay Check of Turbopump Gearbox	I	.9267	.9492	3

TABLE 1 (cont.)

	Stage	Human Reliability Individual	Human Reliability Redundant	Human Reliability Criticality
Prepare for Pressure-Decay Check of Turbopump Gearbox	I	.9327	.9535	2
Prepare for Electrical Check	II	.9337	.9523	3
Perform Pressure-Decay Check of Turbopump Gearbox	II	.9350	.9553	3
Perform Visual Inspection of Areas Worked on in Test Above, to Ensure Hardware Integrity and That all Tools Have Been Removed from Area	I	.9478	.9603	3
Service Oil Sump of Turbopump Gearbox	II	.9544	.9686	3
Prepare for Turbopump Torque Check	II	.9575	.9727	3
Prepare for Installation of Gearbox Pressurization Kit	I	.9583	.9728	3
Perform Visual Inspection of Areas Worked on in Test Above, to Ensure Hardware Integrity and That all Tools Have Been Removed from Area	II	.9590	.9794	3
Perform Turbopump Torque Check of Sub-Subassembly	I	.9601	.9751	3
Prepare for Installation of Gearbox Pressurization Kit	II	.9609	.9747	3
Prepare for Subassembly Turbopump Torque Check	I	.9834	.9917	2

II, Summary (cont.)

C. SUGGESTED APPLICATIONS

If the results of the current validation study at Vandenberg Air Force Base are favorable, design engineers for future systems may use the reliability figures derived from the study to improve the selection of components. The total reliability of a system is dependent on the interaction of its components with personnel. Designers may prefer to select a component of somewhat lower hardware reliability which has higher human reliability in operation and maintenance, so that the man/machine interaction will result in improved system reliability.

Because this study indicates that the number of task elements critically determines the reliability with which a maintenance task will be performed, the number of steps required to operate and maintain equipment should be held to a minimum.

The development of a personnel redundancy formula provides a convenient way of estimating the improvement in task reliability when more than on mechanic is assigned to a single maintenance task. Optimum reliability can be approached if procedural writers give careful consideration when specifying the degree and manner of interaction of mechanics performing a task.

To determine optimum preventive maintenance schedules, the results of the present study could be used with the information derived from the effects of silo storage on the reliability of hardware. Such schedules, by incorporation quantitative information on human reliability in mathematical models of system effectiveness, would give realistic consideration to the personnel subsystem.

Thus far the evidence indicates the possibility of serious problems for planners of future systems of extremely high reliability requirements. Traditionally, the reliability of systems is estimated without taking the reliability of the personnel subsystem into account. As Williams (Ref 6) points out, the product reliability is very different when the reliability of human functions is used in computing system reliability. The present study offers an approach to acquiring the necessary quantification of human behavior that is essential (if the human element is to be accounted for) in estimating system reliability.

One final caution is in order. Majesty (Ref 9) points out that while a human reliability metric is highly desirable and will immeasurably aid engineers in their approach to designing increasingly reliable systems, human feeling, motivations, changes, aberrant reactions under stres, etc. must also be considered. Man should not be viewed simply as a component of fixed or predetermined reliability.

III. DEVELOPMENT OF THE METHOD

A. PREVIOUS EFFORTS

This section contains a fairly extensive discussion of previous efforts in determining human reliability. It is intended for readers professionally concerned with this problem. Readers, who are professionally interested in other fields and in a detailed description of the method finally selected, may wish to turn directly to Section III, B.

In recent years, there have been numerous attempts to predict and measure human reliability (Ref 2,3,4,5, and 6). The topic received considerable attention in the program on personnel subsystem reliability held at the National Aerospace Systems Reliability Symposium in Salt Lake City, Utah, 16 through 18 April 1962 (Ref 7,8,9, and 10).

Recently, Task Group 2 of the Electronics Industry Association's Military Subcommittee on Human Factors in Electronics was established to achieve two goals: (1) to design a system to collect and process human-error rate data, and (2) to prepare the necessary mathematical models and procedures for predicting the degradation that the human element introduces to system or subsystem effectiveness.

Five general approaches to determining human reliability can be identified:

1. Analysis of Field Experience
2. Extrapolation from the Experimental Literature
3. Conducting Special Studies in a Simulated Environment
4. Conducting Special Studies in the Operational Environment
5. Judging and Rating Reliability

These five general approaches to determine human reliability are discussed as follows:

III, A, Previous Efforts (cont.)

1. Analysis of Field Experience

The first and most common approach is the analysis of field experience. During normal maintenance operations, empirical data, derived from accumulating records, may be used to establish error rates. Although this approach appears attractive, certain disadvantages become apparent on investigation. Most reporting systems are designed to report hardware malfunctions but give inadequate attention to description of human errors. In addition, such reports depend heavily on the accuracy and thoroughness of the reporting personnel. Often, the person making an error fails to notice it or is unwilling to report it. Consequences of error may be discovered later, but the analysis of cause depends on inferences about probable actions leading to the event. Such limitations are so serious that unless special precautions are taken, field reports are misleading in establishing degree of human error.

For example, Shapero, et al. (Ref 11), studied 3829 malfunction reports from seven different missile test programs and 419 "unscheduled hold" reports from two missile-systems test efforts:

a. Of the 3829 malfunction reports, 39% were classified as human-initiated.

b. Of the 419 unscheduled holds, 20% were classified as human-initiated.

c. Most important, the analysis of two systems in which the Ballistic Missile Division Failure and Code List was used revealed 322 human-initiated malfunctions in one system and 193 in the other. Prior to the analysis, the reporting personnel (who filled out the original failure data report) classified only three human-error incidents in one system and none in the other.

Field reports are seriously limited because they cover only undesirable events. Data on successes or total number of attempts are difficult to obtain. To establish error rate or human reliability, little is gained by knowing that a percentage of malfunctions or failures is due to human error or that so many human errors occurred this month compared to the number reported last month, or a year earlier. To determine the reliability with which certain tasks are performed by mechanics, it is necessary to know the number of times a task was attempted and the number of attempts resulting in success. A final limitation is that such data is available only for existing systems. To predict human reliability for a system in the early stages of research and development, performance data will be limited to existing systems only. It may be possible to extrapolate from data on an earlier, similar system, but additional problems will be encountered.

In many ways, the best illustration of the analysis of field experience is the recent study by Rook (Ref 4) for the Sandia Corporation. In the field of nuclear ordnance, emphasis is on high reliability and a "no-field-test" philosophy. Critical importance is placed on the reduction of human error in production. Rook was able to collect error rate data for certain classes of human error in production, such as "Two wires which can be transposed are transposed" and "Soldering operation results in insufficient solder." These rates were obtained from an analysis of over 23,000 production defects detected in assembly operations of electronic equipment. Rook suggests an extremely promising classification scheme that will aid in generating design and procedure change proposals. A quantitative model was also developed for evaluating the contribution of human error to the degradation of product quality.

2. Extrapolation from Experimental Literature

Recently an ingenious approach to predicting the reliability of operator performance was developed by Payne and Altman (Ref 1 and 3). They reasoned that during 15 to 20 years of experimentation, most significant factors affecting performance were identified and studied. They surveyed several thousand research reports and finally selected 164 reports meeting their requirements.

To make the findings of experimental studies useful in predicting the operability of new equipment, it was necessary to develop a common conceptual framework. The framework involved 23 components such as circular scales, lights, cable connections, object positioning, etc. For each component, the associated criteria affecting the operator's performance were identified. For example, the criteria affecting performance on the component "lights" were size, brightness, type/function, number, and presentation.

Data were abstracted from the literature in this framework. Rational and empirical approaches were used to reduce and integrate the data and render the laboratory data more compatible with field conditions. The final result is a Data Store that provides reliability estimates for the criteria of each performance. These data can be combined by using the product rule to obtain reliability estimates at the behavior step, mission phase, and total mission levels.

3. Conducting Special Studies in a Simulated Environment

Another approach for determining human reliability is to conduct special studies in a simulated environment. Applied experimental research would make the subject's task as similar as possible to the anticipated field situation. Performance data from the experimental situation are used to predict performance in the field situation. One problem is the difficulty of simulating certain situations at reasonable cost. Also, in the absence of field data, it is difficult to assess the adequacy of the simulation. Knowing that the situation is a laboratory rather than a field situation will affect the subject's performance thus adding to the difficulties. However, the most serious limitation of this approach for purposes of estimating human reliability is the very large number of observations needed to establish reasonably accurate reliability estimates. Payne and Altman (Ref 1) estimate that mean behavior step unreliability in typical field situations is .0026. Rook (Ref 4) obtained error-rate figures varying from .07 to .00003. To establish such rates, hundreds and preferably thousands of observations would be necessary.

4. Conducting Special Studies in an Operational Environment

Limitations of data generated from routine field experience, problems associated with assessing the relevance of laboratory data, and adequacy of simulation in the laboratory approach have led to special studies in the operational environment. A recent study of this nature is reported by Meister (Ref 12). The study involved analysis of 702 job operations during Atlas testing at Vandenberg Air Force Base. The study was conducted at the Operational System Test Facility (OSTF), which is a facility for testing the adequacy of the total missile system, including personnel functions. This facility is a complete operational missile site in which missiles are received, checked out, and fired.

The 702 job operations were observed over a period of nine months by personnel trained to observe these operations. Observations made on a one-observer to one-operator basis overcame many disadvantages inherent in the operator-reported errors of field experience analysis. Unfortunately, for the purposes of the present study, Meister does not report the rate at which errors occur but rather the percent of errors falling in certain error categories. For example, his findings suggest that the largest proportion of the observed errors are "system" rather than "operator" errors. System errors include such things as nonavailability of personnel, improper personnel utilization, nonavailability of equipment, nonavailability or inadequate technical data, and inadequate system organization.

5. Judging and Rating Reliability

To illustrate how the probability analysis techniques utilized for equipment can also be applied to the analysis of human factors in a system, Williams (Ref 6) used judges' ratings of the probability of success for various operator actions such as "Reading frequency from a moving scale-type dial," and "Recording frequency as received over telephone." Ratings on a variety of tasks can be obtained quickly and inexpensively. The ratings should have some validity if the raters are thoroughly familiar with kinds of tasks being rated and with the type of personnel performing them. Raters were asked to predict absolute reliability directly (.97, .99, etc.). Williams' approach required judges to estimate figures with which they had little experience (human reliability in performing tasks). The present study utilized a combination of a modified rating approach and empirically based reliabilities.

B. METHOD OF OBTAINING RATINGS

This section describes the method utilized to obtain likelihood-of-error ratings in the performance of maintenance of the Titan II propulsion System.

1. Specify Maintenance Tasks that must be Performed

To determine the performance reliability for each silo maintenance task, the activity tasks must first be identified. In the present study, Type II and Type III basic data, prepared in accordance with AFBM Exhbiit 60-26A, was used for this purpose. Two broad areas of maintenance activities were considered. First, scheduled or preventive maintenance, and second, unscheduled or corrective maintenance. To date, the following scheduled maintenance phases have been investigated:

a. Engine servicing
b. Engine visual inspection
c. Engine checkout.

III, B, Method of Obtaining Ratings (cont.)

These activities were further factored into specific tasks such as:

a. Service oil sump of turbopump gearbox.

b. Perform visual inspection of areas worked to ensure hardware integrity, and that all tools have been removed from the area.

c. Prepare for electrical check.

d. Perform electrical check.

(The full set of tasks is presented in Table 1)

2. Identify the Task Elements Necessary to Accomplish the Total Tasks

Each task requires specific acts that are smaller units of human behavior than are the tasks themselves. Because they occur repeatedly in a variety of tasks, these smaller units, or task elements, lend themselves readily to an **analysis** of performance reliability. For example, the activity "Engine Checkout" includes the task "Prepare for Leak Check of Turbopump Fuel Pump Seal," which contains 13 elements (Table 2).

These elements, although related directly to the task of "Prepare for Leak Check of Turbopump Fuel Pump Seal," are repeated in other tasks such as "Prepare for Leak Check of Turbopump Fuel Gearbox Seal." Furthermore, a closer analysis of each element reveals a discrete action that is common not only to the propulsion system but to a variety of mechanical systems. Therefore, establishing performance reliability at this level broadens the applicability of the results. An analysis was made of the activity tasks, and 60 discrete task elements (Position Hand Valves, Install "O" Ring, Remove Lockwire, etc.) were selected for reliability determination (Table 9).

TABLE 2

TASK ELEMENTS IN TASK, "PREPARE FOR LEAK CHECK OF TURBOPUMP FUEL PUMP SEAL"

TASK ELEMENT

1. Verify that all valves, regulators, and switches are in proper position, and that the functional test set is otherwise prepared for use.
2. Connect a valve and hose assembly to the "Fuel System Leak Test" port on the functional test set.
3. Remove the protective closure from the quick-disconnect coupling on the turbopump fuel discharge line.
4. Connect the valve and hose assembly to the quick disconnect coupling.
5. Disconnect the fuel pump seal cavity drain tube assembly "B" nut from the pump seal cavity drain fitting.
6. Remove the cavity drain tube assembly from the engine.
7. Install the reducing adapter on the cavity drain fitting.
8. Connect a flexible hose assembly to the reducing adapter.
9. Remove the pressure caps from the flowmeter inlet and outlet ports.
10. Connect the flexible hose from the cavity drain to the flowmeter inlet port.
11. Disconnect the pressure sequence valve overboard drain tube assembly "B" nut from the overboard drain check valve.
12. Remove the overboard drain tube assembly from the engine.
13. Install a pressure cap on the overboard drain check valve.

III, B, Method of Obtaining Ratings (cont.)

3. Determine the Performance Reliability for Each Task Element

Once the task elements have been identified, the reliability of the performance of each by a missile engine mechanic must be rated. The present approach differs from that of Rook (Ref 4), who determined the rate at which specific types of errors occur. For example, in "Soldering a wire," Rook predicts the rates at which excess solder, insufficient solder, or a hole in the solder will occur. In the present study, the aim is to predict the reliability with which the total task of soldering a wire will be accomplished. The effort to predict reliability rather than associated error rates was based on the following considerations: all possible errors are not known, whereas required task elements are clearly specified in Type III basic data, technical orders, checklists, and other procedural documents. Because there are fewer task elements than possible errors, the judging task of the raters is simplified. Communicating reliability estimates of the personnel subsystem in familiar terms to engineers working on hardware subsystems increases the liklihood of their using human reliability information in problems involving man-machine interactions. The approach has the disadvantage of leaving the nature of the errors unspecified.

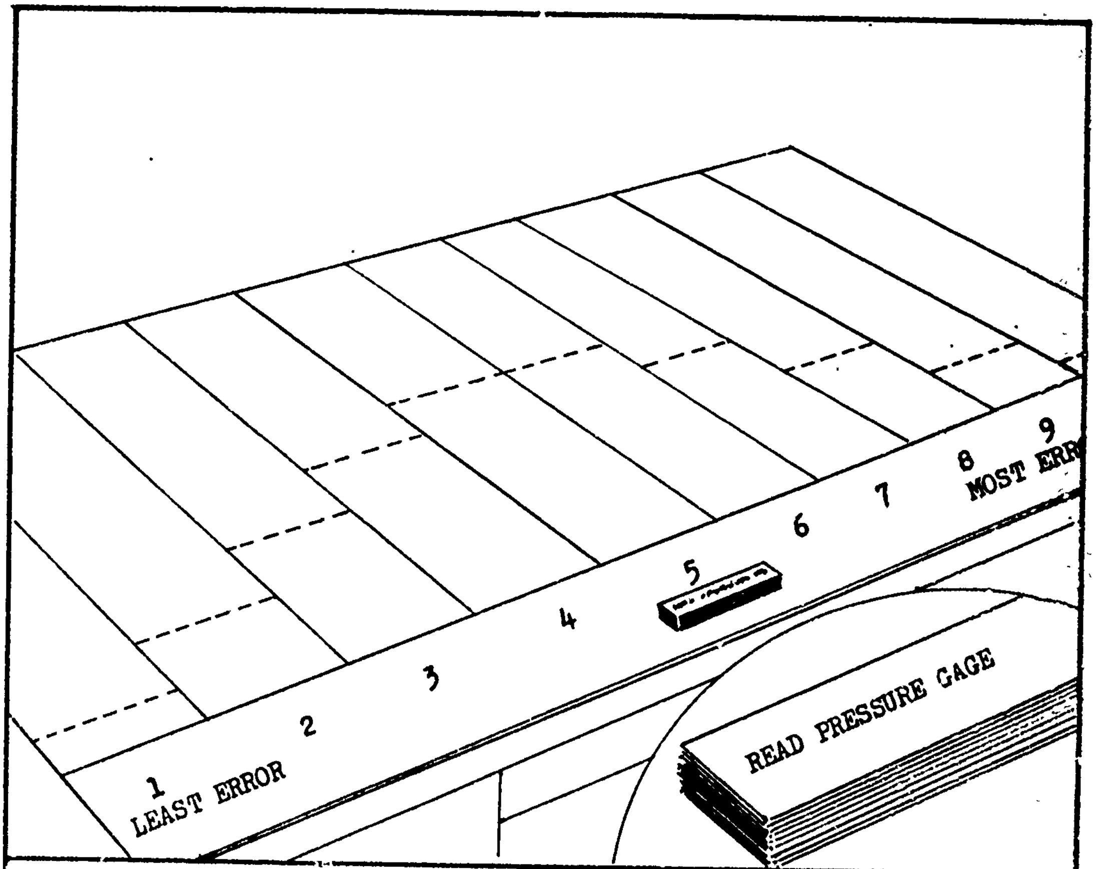
4. The Preliminary Rating Study

To test the feasibility of a rating approach in determining performance reliability, task elements were sorted into a forced normal distribution (Figure 1). Judges, selected for their knowledgeability and experience, were requested to sort 60 task elements printed on 1-1/4-in. by 5-in. cards into a set of tied ranks in accordance with prescribed directions (Appendix). Five personnel-subsystem engineers, three Air Force missile-engine mechanics, and one Aerojet engine mechanic were appointed judges. Criteria for their selection included knowledge of Titan II maintenance tasks and Air Force engine-mechanic performance. To prevent the judges from putting many cards in only one or a few categories and to ensure their use of the full continuum, a forced normal distribution, such as used by Stephenson (Ref 13), was selected for the preliminary study. On the other hand, the forced-choice approach has the disadvantage of violating the assumption of independence of ratings. If a rater places a task in Category I (least error), he cannot place another task in that category even though he believes it belongs there.

a. Measurement Reliability of Ratings

Reliability, as the term is used by psychologists, refers to the accuracy and consistency of measurement. It should not be confused with the engineering use for referring to the probability that a component or system will perform successfully. To avoid possible confusion, the former will be referred to throughout this report as "measurement reliability." The probability that the personnel subsystem will perform successfully is referred to as "human reliability" or "reliability of task performance." The probability that a hardware component or system will perform successfully is referred to as "hardware reliability."

Using the intraclass correlation formula recommended for such problems by Ebel (Ref 14), the measurement reliability of the nine sets of ratings was estimated. The results of this analysis are presented in Table 3. Because the intended use of the obtained ratings will be based on average ratings of task elements from all nine raters, the measurement reliability of those averages are the most interesting. In this case, the measurement reliability was .88, indicating a high degree of agreement among raters concerning the likelihood of error for our 60 task elements. Of course, this did not establish the validity of such ratings, only the inter-rater measurement reliability. No completely valid measures of the human reliability of Titan II missile-engine mechanic task performance were available. However, some empirically-based measures of task-element reliability were available through the Index of Electronic Equipment Operability Data Store (Ref 3). From the Data Store (Table 4), it was possible to obtain human reliability estimates for 29 of the 60 task elements.



Insert illustrates typical task-element sorting card.
Dotted area represents forced normal sort presented to raters during preliminary study.

Sorting Arrangement

Figure 1

TABLE 3

MEASUREMENT RELIABILITY OF RATINGS, PRELIMINARY STUDY

Sum of squared ratings		=	18,450
Product of sum and mean	$2970 \times \frac{2970}{540}$	=	16,335
Sum of squares			
For raters	$\frac{980,100}{60} - 16,335$	=	0.0
For task elements	$\frac{156,840}{9} - 16,335$	=	1,092
For total	$18,450 - 16,335$	=	2,115
For error	$2,115 - 1,092$	=	1,023
Mean square			
For task elements	1092/59	=	18.5
For error	1023/472	=	2.2
Reliability of ratings	$\frac{18.5 - 2.2}{18.5 + (9-1)2.5}$	=	.42
Reliability of average ratings	$\frac{18.5 - 2.2}{18.5}$	=	.88

TABLE 4

TASK ELEMENT RELIABILITIES FROM DATA STORE

Task Element	Data Store Reliability
Read Electrical or Flow Meter	.9860
Read Time (Brush Recorder)	.9873
Read Pressure Gauge	.9897
Tighten Nuts, Bolts, and Plugs	.9960
Position "Zero-In" Knob	.9962
Connect Electrical Cable (threaded)	.9968
Install Lock Wire	.9968
Install O-Ring	.9971
Position Multiple Position Electrical Switch	.9972
Install Marman Clamp	.9978
Install Gasket	.9978
Position Two Position Electrical Switch	.9979
Install Reducing Adapter	.9980
Install Nuts, Plugs and Bolts	.9980
Install Drain Tube	.9980
Install Union	.9980
Remove Pressure Cap	.9981
Disconnect Flexible Hose	.9981
Remove Nuts, Plugs and Bolts	.9981
Install Torque Wrench Adapter	.9981
Remove Reducing Adapter	.9981
Remove Union	.9981
Remove Drain Tube	.9981
Connect Flexible Hose	.9984
Remove Protective Closure (friction fit)	.9990
Install Protective Cover (friction fit)	.9990
Remove Torque Wrench Adapter	.9991
Install Funnel or Hose In Can	.9991
Remove Funnel From Oil Can	.9991

b. Results

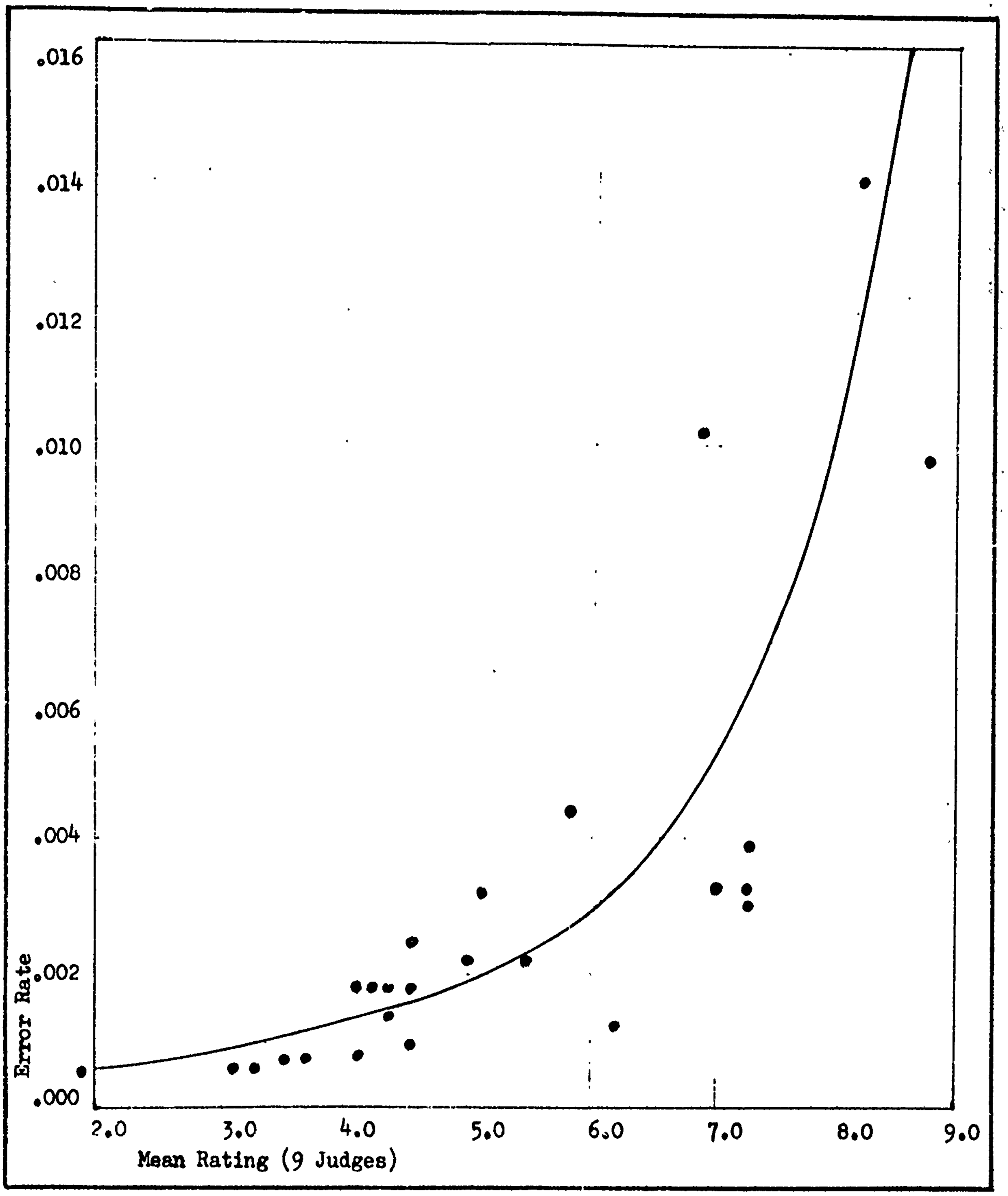
Figure 2 presents the scatter plot relating ratings and Data Store figures for the 29 tasks. The curvilinear relationship is clear. We had expected a linear relationship and wondered if the obtained relationship might have been due to an artifact. Some raters indicated they wanted to place more of the cards toward the least-error end of the continuum. The distribution of the Data Store figures was noticeably skewed in that direction. Forcing raters to normalize what would other wise have a skewed distribution of ratings would result in the sort of curvilinear relationship obtained in Figure 2.

Because of this and the lack of independence of forced ratings, it was decided to abandon the use of forced ratings and to use a modified Thurstone Method of Equal Appearing Intervals procedure in the main study (Ref 15).

5. Main Rating Study

Based on the experiences gained in the preliminary study, the decision was made to eliminate the use of the forced choice but to retain the remainder of the rating technique. Judges again were selected on the basis of their availability and knowledge of Air Force technician and Titan II maintenance tasks. Identical task elements and the cards used in the first study were employed. A modification was made in the instructions as follows:

> "You have been selected for this task of Titan II because of your familiarity with the maintenance of Titan II engines, plus your knowledge of the capability of Air Force personnel to perform maintenance activities. Please use your knowledge and experience in estimating how reliable average Air Force missile engine mechanics would perform a set of selected jobs. In some we could



Scatter Plot of Ratings and Data-Store Error Rates, Preliminary Study

Figure 2

> expect almost perfect performance. In others they will make errors. For example, in one study more errors occurred when soldering than when removing wire insulation. In another, more errors occurred in disconnecting locking-type cables than nonlocking-type cables. There are 60 cards in front of you, each with a different task printed on it. These are the types of tasks that a missile engine mechanic will perform in the silo during maintenance.
>
> On the table before you are 10 columns. At your extreme left (Column 1) place the items you believe will be performed with the least error and on the extreme right (Column 10) place the items that will be performed with the most error. You may place as many cards as you wish in any of the columns. As you move from left to right, each column represents an equal increase in the degree of error. If at any time you wish to change the location of any cards, feel free to do so."

The sorting technique was administered mainly by Aerojet personnel subsystem and reliability human-factors personnel. In some cases, the task was delegated to representatives in the test areas at Sacramento or at Vandenberg Air Force Base.

Eighteen Aerojet personnel subsystem and reliability engineers, six Aerojet technicians, and nine Air Force missile-engine mechanics were used as judges in the main rating study. The engineers were thoroughly familiar with maintenance of the Titan II Propulsion System. They were knowledgeable of the engine and, from extensive observation, were familiar with the performance of typical Air Force mechanics. The Aerojet technicians were familiar with the tasks involved, having themselves performed the tasks as part of their work in the Aerojet test area or, in one case, as a technician in the Personnel Subsystem Test and Evaluation Program. The Air Force mechanics from Vandenberg Air Force Base were familiar with the tasks and were participating in Category II testing of Titan II during the time they served as raters in the present study.

The columns into which the judges sorted the task elements were numbered left to right from 1 to 10 (Figure 1). An item placed in a column received the value of that column. When sorting was completed, the cards were turned over and the number of the card and column placement was recorded. Thus, each card received 33 independent ratings by 33 qualified judges.

a. Measurement Reliability of Ratings

Because the raters had somewhat different backgrounds, the reliability of ratings was investigated by sub-groups of raters as well as for the entire group of 33 raters. The intra-class correlation for estimating the reliability of ratings (Ref 14) was used as in the preliminary study, and the results are presented in Tables 5, 6, 7, and 8.

It can be seen that the most reliable ratings were obtained from the Aerojet engineers (.295 at the individual rating level as compared with .263 for the Aerojet technicians and .130 for the Vandenberg AFB missile-engine mechanics). Because the number of raters differs from group to group, the individual rating level is the most appropriate level for determining the most reliable source of ratings. The measurement reliability of average ratings based on 18 raters would be higher than the reliability of average ratings based on 6 raters even though the ratings were fully equivalent in other respects.

Inspection of the distribution of ratings reveals the reason for the relatively low measurement reliability of the technicians and Air Force mechanics. Whereas 14 of the 18 Aerojet personnel subsystem and reliability engineers used the full range of available ratings (from 1 to 10), only one of the

TABLE 5

MEASUREMENT RELIABILITY OF RATINGS--33 RATERS

Sum of Squared Ratings		=	36,211
Product of Sum and Mean	$\frac{(7003)(7003)}{1980}$	=	24,769
Sum of Squares			
For Raters	$\frac{1{,}671{,}853}{60} - 24{,}769$	=	3,095
For Tasks	$\frac{886{,}651}{33} - 24{,}769$	=	2,099
For Total	36,211 - 24,769	=	11,442
For Error	11,442 - 3,095 - 2,099	=	6,248
Mean Square			
For Tasks	2,099 ÷ 59	=	35.5763
For Error	6248 ÷ 1888	=	3.3093
Reliability of Ratings	$\frac{35.5763 - 3.3093}{35.5763 + 32\,(3.093)}$	=	.2281
Reliability of Average Ratings	$\frac{35.2542 - 3.3194}{35.2542}$	=	.9153

TABLE 6

MEASUREMENT RELIABILITY OF RATINGS--18 AGC ENGINEERS

Sum of Squared Ratings		=	28,680
Product of Sum and Mean	$\frac{(4900)(4900)}{1080}$	=	22,231
Sum of Squares			
For Raters	$\frac{1,352,316}{60} - 22,231$	=	307
For Tasks	$\frac{437,120}{18} - 22,231$	=	2,053
For Error	$6449 - 2053 - 307$	=	4,089
Mean Square			
For Tasks	$2053 \div 59$	=	34.7966
For Error	$4089 \div 1003$	=	4.0767
Reliability of Ratings	$\frac{34.7966 - 4.0767}{34.7966 + 17\,(4.0767)}$	=	.2950
Reliability of Average Ratings	$\frac{34.7966 - 4.0767}{34.7966}$	=	.8828

TABLE 7

MEASUREMENT RELIABILITY OF RATINGS--6 AGC TECHNICIANS

Sum of Squared Ratings		=	4,205
Product of Sum and Mean	$\frac{(1003)(1003)}{360}$	=	2,794
Sum of Squares			
For Raters	$\frac{186,081}{60} - 2,794$	=	307
For Tasks	$\frac{19,317}{6} - 2,794$	=	426
For Total	$4,205 - 2,794$	=	1,411
For Error	$1,411 - 307 - 426$	=	678
Mean Square			
For Tasks	$426 \div 59$	=	7.2203
For Error	$678 \div 295$	=	2.2983
Reliability of Ratings	$\frac{7.2203 - 2.2983}{7.2203 + (6 - 1)\, 2.2983}$	=	.2630
Reliability of Average Ratings	$\frac{7.2203 - 2.2983}{7.2203}$	=	.6816

TABLE 8

MEASUREMENT RELIABILITY OF RATINGS--9 VANDENBERG AFB MISSILE-ENGINE MECHANICS

Sum of Squared Ratings		=	3,326
Product of Sum and Mean	$\frac{(1100)(1100)}{540}$	=	2,241
Sum of Squares			
For Raters	$\frac{151,456}{60} - 2,241$	=	283
For Tasks	$\frac{21,808}{9} - 2,241$	=	182
For Total	3,326 - 2,241	=	1,085
For Error	1,085 - 283 - 182	=	620
Mean Square			
For Tasks	182 59	=	3.0847
For Error	620 472	=	1.3135
Reliability of Ratings	$\frac{3.0847 - 1.3135}{3.0847 + (9 - 1) 1.3135}$	=	.1303
Reliability of Average Ratings	$\frac{3.0847 - 1.3135}{3.0847}$	=	.5741

technicians and one of the Air Force missile-engine mechanics did so. Three of the Air Force missile-engine mechanics and one of the Aerojet technicians assigned no ratings higher than 3. All engineers used ratings from 1 to 9. Those using a highly restricted range did not understand the instructions or they perceived errors as much less likely than did the other raters. One possibility considered was to eliminate some of the raters. Thurstone (Ref 15), using an eleven-point continuum and 100 to 125 items to be rated,eliminated raters who placed 30 or more statements in a single category. However, the validity of this approach has been questioned by Hovland and Sherif (Ref 16), who found in one study, that if raters were eliminated according to this criterion, more than two-thirds of their judges would have been eliminated and genuine differences between raters obscured.

In a sense, the Aerojet technicians and the Air Force mechanics were rating themselves, whereas the Aerojet engineers were rating others. Generally, the error of leniency, which is a problem in performance ratings, is particularly pronounced in self-ratings. Perhaps the present finding is simply another example of the tendency of humans to be unwilling or unable to report their own errors. Perhaps job incumbents are not good raters for this purpose.

A final possibility is that the difference in ratings is related to the circumstances under which the ratings were requested. The Aerojet engineers were asked to perform the ratings by a colleague. Because the requester and the raters were working in the same department, the engineers had a fairly adequate understanding of the purpose of the ratings. In the other two cases, the ratings were requested by an "outsider," whose purpose was less likely to be fully understood. In the absence of a fully defensible procedure for rejecting "careless" judges, and because ratings from the engineers were the most reliable (.88) and were obtained from fully qualified judges, who were not ego-involved in the ratings, it was decided to base the remaining analyses on those ratings.

6. Results

The means and standard deviations of the ratings assigned to each task element are presented in Table 9. The task elements are arranged in order of rank, from most to least error. Generally, reading, inspecting, and installing tasks were judged more likely to produce error than removal tasks. Raters showed the greatest agreement (smallest standard deviation) in rating these tasks: (1) Install Funnel or Hose in Can, (2) Loosen Nuts, Bolts, and Plugs (3) Remove Funnel from Oil Can, (4) Remove Union, (5) Remove Drain Tube, and (6) Position Two-Position Electrical Switch.

Raters showed the least agreement (largest standard deviation) in rating these tasks: (1) Read Electrical or Flow Meter, (2) Inspect for Bellows Distortion, (3) Lubricate Bolt or Plug, (4) Inspect for QC Seals, (5) Close Hand Valves, and (6) Tighten Nuts, Bolts, and Plugs.

To determine the relationship of the present results to the preliminary study, or the results that would have been obtained using Aerojet technicians or Air Force missile-engine mechanics as raters, Pearson product-moment correlations were computed between the mean ratings received for the task elements under the four conditions. These results are presented in Table 10. There is substantial agreement among the four groups of raters. Similar though not identical results would have been obtained if another source of ratings had been used instead of the one finally selected.

Figure 3 shows the mean frequency with which the task elements were assigned to categories by the 18 engineers. The distribution is skewed unlike the forced normal distribution used in the preliminary study.

TABLE 9

TASK-ELEMENT RELIABILITY ESTIMATES

Task Element	Rating Mean	Rating S.D.	Individual Reliability	Percent of Redundancy	Redundant Reliability
Read technical instructions	8.3	2.2	.9901	80	.9979
Read time (brush recorder)	8.2	2.1	.9904	30	.9933
Read electrical or flow meter	7.0	2.8	.9928	30	.9949
Inspect for loose bolts and clamps	6.4	1.9	.9938	10	.9944
Position multiple-position electrical switch	6.3	2.4	.9940	20	.9952
Mark position of component	6.2	2.1	.9941	40	.9964
Install lockwire	6.0	2.3	.9944	20	.9955
Inspect for bellows distortion	6.0	2.7	.9944	20	.9955
Install Marman clamp	6.0	1.8	.9944	50	.9972
Install gasket	6.0	2.1	.9945	30	.9961
Inspect for rust and corrosion	5.9	2.1	.9946	20	.9957
Install "O"-ring	5.7	2.2	.9948	30	.9964
Record reading	5.7	2.3	.9949	20	.9959
Inspect for dents, cracks, and scratches	5.6	2.4	.9950	20	.9960
Read pressure gauge	5.4	2.2	.9952	20	.9962
Inspect for frayed shielding	5.4	2.3	.9952	20	.9962
Inspect for QC seals	5.3	2.6	.9953	20	.9962
Tighten nuts, bolts, and plugs	5.3	2.6	.9953	20	.9962
Apply gasket cement	5.3	2.3	.9954	20	.9963
Connect electrical cable (threaded)	5.2	2.2	.9955	30	.9968
Inspect for air bubbles (leak check)	5.0	2.2	.9957	30	.9970
Install reducing adapter	4.9	1.6	.9958	30	.9975
Install initiator simulator	4.9	2.5	.9958	50	.9975
Connect flexible hose	4.9	2.4	.9958	40	.9975

TABLE 9 (cont.)

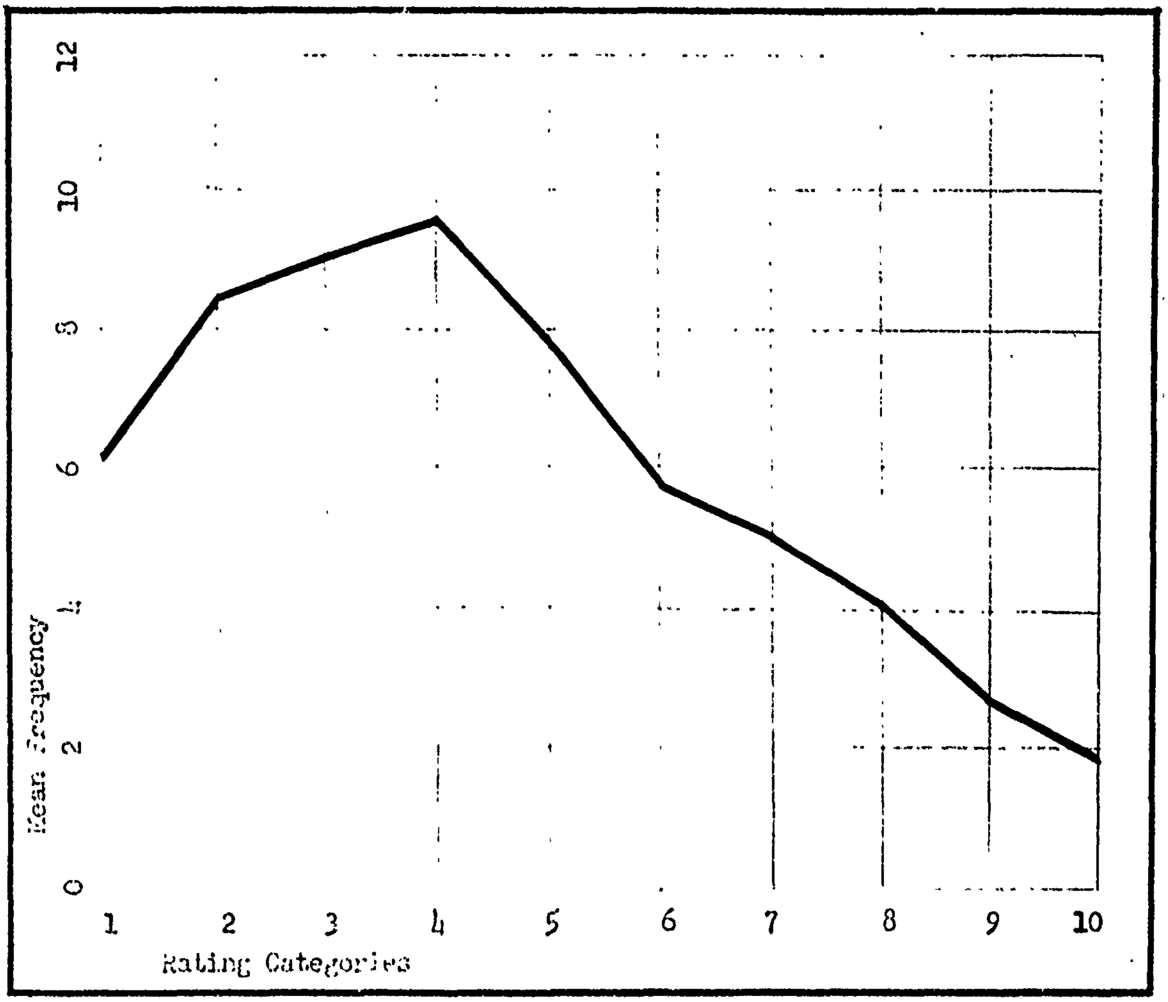
TASK-ELEMENT RELIABILITY ESTIMATES

Task Element	Rating		Individual Reliability	Percent of Redundancy	Redundant Reliability
	Mean	S.D.			
Position "zero-in" knob	4.8	1.6	.9959	10	.9963
Lubricate bolt or plug	4.7	2.7	.9960	10	.9964
Position hand valves	4.6	1.6	.9962	10	.9966
Install nuts, plugs, and bolts	4.6	1.7	.9962	30	.9973
Install union	4.5	1.8	.9962	40	.9977
Lubricate "O"-ring	4.5	2.5	.9962	10	.9966
Rotate gearbox train	4.4	2.0	.9963	20	.9970
Fill sump with oil	4.3	1.6	.9964	30	.9975
Disconnect flexible hose	4.2	2.0	.9965	30	.9975
Lubricate torque-wrench adapter	4.2	2.2	.9965	10	.9968
Remove initiator simulator	4.1	1.9	.9966	50	.9983
Install protective cover (friction fit)	4.1	2.2	.9966	20	.9973
Read time (watch)	4.1	2.1	.9966	30	.9976
Verify switch position	4.1	1.9	.9966	20	.9973
Inspect for lockwire	4.1	2.1	.9966	10	.9969
Close hand valves	4.0	2.6	.9966	30	.9976
Install drain tube	4.0	2.1	.9966	40	.9980
Install torque-wrench adapter	3.9	1.7	.9967	10	.9970
Open hand valves	3.8	2.6	.9968	30	.9978
Position two-position electrical switch	3.8	1.5	.9968	30	.9978
Spray leak detector	3.7	2.0	.9969	20	.9975
Verify component removed or installed	3.5	2.4	.9971	40	.9983
Remove nuts, plugs, and bolts	3.5	1.7	.9971	30	.9980
Install pressure cap	3.4	1.6	.9971	30	.9980

TABLE 9 (cont.)

TASK-ELEMENT RELIABILITY ESTIMATES

Task Element	Rating Mean	Rating S.D.	Individual Reliability	Percent of Redundancy	Redundant Reliability
Remove protective closure (friction fit)	3.2	1.6	.9973	30	.9981
Remove torque-wrench adapter	3.0	1.6	.9974	30	.9982
Remove reducing adapter	3.0	1.7	.9974	30	.9982
Remove Marman clamp	3.0	1.7	.9974	60	.9990
Remove pressure cap	2.8	1.8	.9975	30	.9982
Loosen nuts, bolts, and plugs	2.8	1.3	.9975	30	.9982
Remove union	2.7	1.4	.9976	30	.9983
Remove lockwire	2.7	1.5	.9976	20	.9981
Remove drain tube	2.6	1.4	.9976	40	.9986
Verify light illuminated or extinguished	2.2	1.6	.9979	30	.9985
Install funnel or hose in can	2.0	0.8	.9980	20	.9984
Remove funnel from oil can	1.9	1.4	.9980	20	.9984



Mean Frequency of Ratings

Figure 3

III, Development of the Method (cont.)

C. EMPIRICALLY BASED PERFORMANCE-RELIABILITY ESTIMATES

Before the ratings could be integrated into system reliability and effectiveness models, it was necessary to convert them into reliability or error-rate figures. If a task received a mean rating of 6.3, it was necessary to know what that represented in terms of probability of success in 10,000 attempts. Also, the validity of the ratings could be better judged if some empirically based figures were available for comparison.

1. Approaches Considered

Each of the following four empirically based approaches described in the introduction was considered: (a) analysis of field experience, (b) extrapolation from the experimental literature, (c) special studies in a simulated environment, and (d) special studies in the operational environment. Aerojet-General utilizes a trouble report form called the Quality/Reliability Report (QRR). These reports are screened to identify failures resulting from human errors. On completion of the investigation of a human-error incident, a Trouble Investigation Report (TIR) is issued to cognizant departments documenting the corrective preventive action initiated. QRR's and TIR's provided a basis for identifying the number of human errors of various types. However, information on the number of task attempts was not available. Because reliability is the ratio of the number of successes to the number of attempts, it can not be computed without information on the number of attempts.

An unsuccessful effort was made to develop a method for estimating the number of times maintenance tasks were performed. Such information could have been derived from engine logs if sufficient time and personnel had been available; however, this was not the case. Because the Titan II was in an early stage of weapon-system development, only limited field data were available, and these data were confined to research and development operations, which may have differed in important ways from maintenance performed in the silo under environmental field conditions. This approach was finally abandoned.

TABLE 10

INTERCORRELATIONS AMONG MEAN RATINGS OF TASK ELEMENTS BY DIFFERENT GROUPS OF RATERS

Number	Type	18 AGC Engineers	6 AGC Technicians	9 AF Missile Engine Mechanics
9	Preliminary Study Subjects	.85	.61	.58
18	AGC Engineers	--	.74	.61
6	AGC Technicians	.74	--	.58
9	AF Missile Engine Mechanics	.61	.58	--

III, C, Empirically Based Performance-Reliability Estimates (cont.)

The Index of Electronic Equipment Operability: Data Store (Ref 3) provided a ready means for obtaining empirically based reliability estimates for 29 of the 60 task elements. As explained in the introduction, the Data Store figures are based on an extrapolation from the experimental literature. Because this was the main approach utilized in the present study, it will be described more fully in succeeding sections.

A third approach considered was that of conducting special studies in a simulated environment. As part of Aerojet's Category I Personnel Subsystem Test and Evaluation (PSTE) Program for Titan II, first- and second-stage engines have been installed in engine demonstrators to simulate an operational missile and launch silo at Aerojet's Sacramento plant.

The test results, using Aerojet technicians (with backgrounds similar to Air Force missile-engine mechanics) as test subjects performing scheduled maintenance activities in the simulated silo, are available in a series of technical operating reports published in accordance with AFBM Exhibit 60-20A and AFBM Exhibit 58-1. The activities were carefully observed by human-factors psychologists and propulsion-system maintenance engineers, thus overcoming the limitations involved in self-reported errors. However, each observed maintenance task was performed only a few times by a limited group of subjects. Hundreds or thousands of additional performance trials would have to be scheduled to provide dependable estimates of performance reliability. To do this at the task level would have been extremely costly and required many additional test subjects.

Another possibility would be to simulate the performance situation at the task element rather than at the full task level. Instead of observing subjects performing leak checks of the fuel system, observe them reading time on a Brush Recorder, installing lock wire, etc. The time and personnel requirements would be considerably reduced by working at the elemental rather than the full task level. Each trial would take minutes instead of hours to complete. A disadvantage of this approach would be the artificial nature of the task situation.

Subjects may not be motivated to perform in the same way in laboratory situations as under more realistic circumstances. Also, the reliability with which a task element is repeatedly performed might differ significantly from the reliability with which the same element is performed when it is only one element in a sequence of 40 elements. Logically, one might expect that errors of omission would be more likely under the latter condition. Although information on the possibilities and limitations inherent in the task-element-simulation approach is needed, it was decided not to attempt such a study as part of the present effort.

The fourth approach considered was based on special studies in the operational environment. An attempt was made to obtain copies of Meister's reports (Ref 12) describing the analysis of maintenance operations at an Operational System Test Facility at Vandenberg Air Force Base. The plan was to translate Atlas maintenance activities into the 60 task-element framework and obtain empirically based reliability figures (based on the nine-month period of observations at Vandenberg Air Force Base during which observations were conducted on a one observer and one operator basis). Unfortunately, copies of the detailed reports will not be received in time to use them in the present study.

As an alternative, such information will be obtained during Category II testing of the Titan II at Vandenberg Air Force Base. Because the breakdown of task elements was based on Type III data for Titan II, the translation problem was eliminated. Also, because Aerojet-General is developing the Titan II propulsion system, ready access is available to expert knowledge regarding Titan maintenance. This is not possible working with Atlas data. The plan for obtaining performance reliability data during Category II testing will be described more fully in a later section.

2. The Index of Electronic-Equipment Operability

The major source of empirically based reliability estimates in the present study is the Payne-Altman *Index of Electronic-Equipment Operability*.

III, C, Empirically Based Performance-Reliability Estimates (cont.)

Interested readers should consult the documents describing the development and use of the Index (Reference 1 and 3). To illustrate the use of the index in estimating reliability (the index also provides estimates of time required to perform tasks), the following example from the present study is given:

One of the 60 task elements is "Read electrical or flow meter." To estimate reliability for this element, one enters the Data Store under the input component, Semicircular Scale. Reliability figures for each of the following parameters are determined:

Parameter	Reliability
Size c. 1-2 in.	.9993
Scale style b. Quantitative information (1) Moving pointer	.9982
Parallax (not applicable)	
Scale arc length b. 50-100	.9950
Scale interval spacing c. 1/10 in, less than 1/2 in.	.9955
Scale brightness c. Easily perceptible from normal position	.9998
Number of graduation marks per unit of required resolution b. Every 5th unit	.9992
Proportion of graduation marks numbered b. 1:5	.9995

III, C, Empirically Based Performance-Reliability Estimates (cont.)

Parameter	Reliability
Scale increase b. Right to left	.9996
Exposure (viewing) time' d. Indefinite	.9997

The reliability of reading an electrical or flow meter is the product of these separate parameter values, in this instance .9860.

In developing the index, Payne and Altman discovered that error estimates obtained from the experimental literature seemed gross overestimates of operational errors. They attributed this to the tendency of experimenters to make tasks unusually difficult, or to count near errors in order to have measurable error without running a large number of trials. Whatever the cause of the difference, to derive operational meaning from experimental data, it was necessary to adjust all the experimental results to make them more compatible with field operation. Payne and Altman did so in the following manner: Based on this experience (with a variety of equipment) of the American Institute for Research, they estimated that the "mean mission failure rate" (roughly analogous to the task failure rate in this study) was 13 per cent. That is, it may be said that about 13 per cent of the time, operator error will fail or seriously degrade mission effectiveness. Because no field studies had been conducted, which provided reliability estimates at what is called the task element level, it was necessary to derive a task-element reliability from the task reliability estimate. This was done by first determining the mean number of steps in a task. A variety of operating manuals for electronic equipment were examined. The number of steps required to operate each item of equipment was determined, and the mean number of steps computed. The mean number of steps was approximately 50. When the mean mission (task) unreliability estimate of .13 was divided by the mean number of steps, a mean mission step (task element) unreliability estimate of .0026 was obtained.

III, C, Empirically Based Performance-Reliability Estimates (cont.)

This step (task element) unreliability was then compared with an estimate of mean unreliability per experimental trial. This was determined from the (abstracts) data available from experimental literature. From the data available, it was found that mean unreliability per trial was .31935. Thus, there were two estimates of mean-step unreliability, one based on actual field operation and one based on laboratory experimentation. Assuming that experimental trials are roughly equivalent to individual steps of operation, the ratio of these means is a reasonable conversion factor for laboratory results. Accordingly, Payne and Altman corrected all experimental results by a factor of $\frac{.0026}{.31935}$, or .008145.

a. Modified Conversion Factor

For the purpose of this study, it was decided that it may be possible to derive a more appropriate conversion factor by utilizing the results of the Personnel Subsystem Test and Evaluation (PSTE) studies conducted at the simulated silo. All studies conducted during 1962 were examined. During this period, 974 tests were conducted.

The following partial list of tasks will indicate the level of description involved:

(1) Prepare for leak check of oxidizer system

(2) Perform leak check of oxidizer system

(3) Prepare for leak check of turbopump oxidizer pump seal

(4) Perform leak check of turbopump oxidizer pump seal

(5) Prepare for leak check of turbopump oxidizer gearbox seal.

(6) Perform leak check of turbopump assembly oxidizer pump gearbox seal

III, C, Empirically Based Performance-Reliability Estimates (cont.)

Human errors that concern the present study occured on 140 (or 14.4%) of the 974 tests. Thus, the task-level reliability was .8560. To determine the mean task-element level reliability, the mean number of task elements per task was determined. This was 42. Because it was planned to use the product rule in estimating task-level reliability from task-element reliabilities, the procedure was reversed in the present instance. The task-element reliability was determined. When raised to the 42nd power (representing the 42 task-element steps), it would give a task reliability of .8560. This was .9963. The estimate of task-element reliability of .9963 was similar to the Payne-Altman estimate of mean-step reliability under field conditions of .9974 (mean unreliability = .0026).

Because figures obtained using the <u>Index of Electronic Equipment Operability:</u> <u>Data Store</u> (Reference 4) would represent slight over estimates of the reliability to be expected from Titan II missile-engine mechanics, each task-element-level estimate was reduced by the following conversion factor:

$$\frac{.9963}{.9974} = .9988$$

In effect, this meant reducing each index-derived task-element figure by .0011.

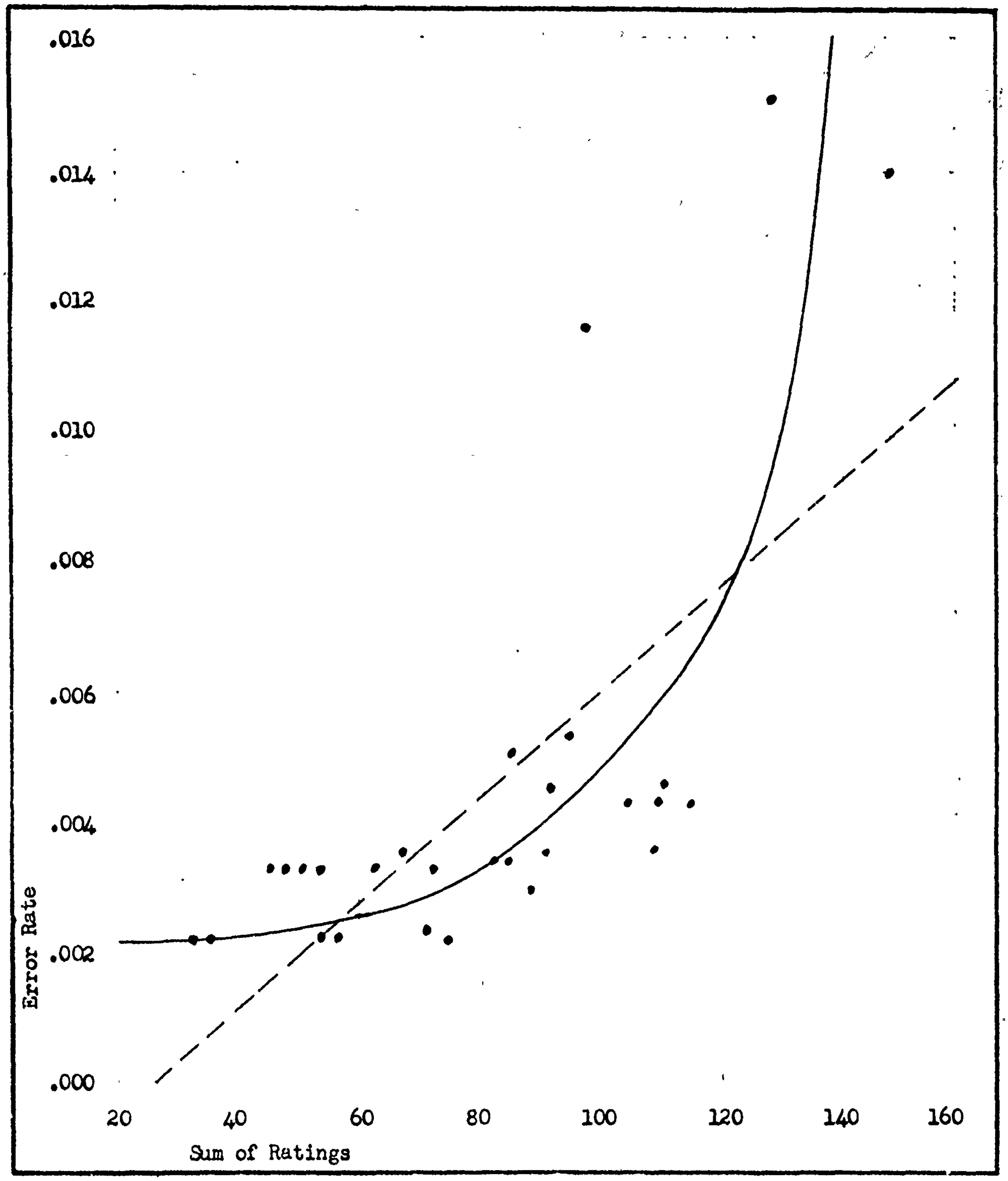
IV. THE RELATIONSHIP BETWEEN THE RATINGS AND THE EMPIRICALLY BASED RELIABILITY ESTIMATES

The 29 task elements for which both ratings and modified Index of Electronic Equipment Operability (Ref 3) figures were available were plotted on a scatter diagram (Figure 4). The Pearson product-moment correlation between these two variables is .457, which is significant between the .05 and .01 levels of confidence. Inspection of Figure 4 suggested that a curvilinear relationship would fit the data better than a linear one. The data were replotted on semi-logarithmic paper with the results shown in Figure 5.

A logarithmic curve was fitted to these data. The equation for the logarithmic curve in Figure 4 is $\log E - 2.9174 - .006122R$, where E = error rate and R = pooled rating of likelihood of error. This equation was used to derive reliability estimates for each of the 60 ratings. These estimates are presented in Table 9.

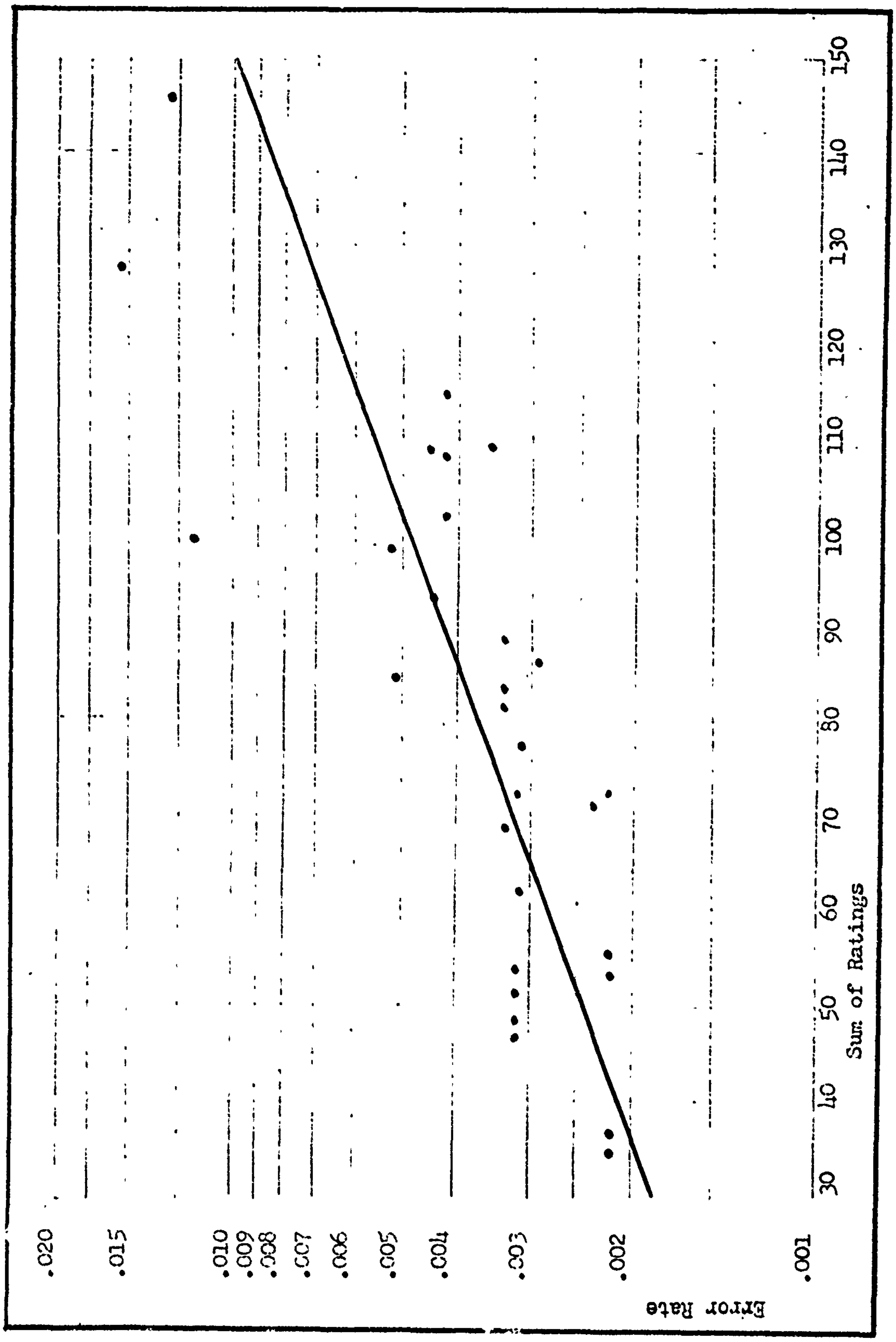
The logarithmic curve was selected in preference to a linear function for two reasons. First, the fit was better. Both linear and logarithmic regression equations were computed relating empirically derived error rates and pooled ratings. The sum of squared deviations from the linear regression line was .00016419; the sum of the squared deviations from the logarithmic regression line was .00015242, a reduction of 7.2 percent. Second, the logarithmic relationship has a long history in the study of psychophysics, the study of human perception of physical magnitudes.

The Weber-Fechner law (Ref 17) ($S = k \log R$) states that sensation (perception of brightness, loudness, weight, etc.) is a function of the log of the magnitude of the physical stimulus. While not a universal "law," this relationship holds approximately, except for extreme values of the stimulus. The present results are consistent with the Weber-Fechner law. Sensation (perception of the likelihood of error by raters) is a function of the log of the physical probability of error.



Scatter Plot of Ratings and Data-Store Error Rates

Figure 4



Scatter Plot of Ratings and Data-Store Error Rates (Semilogarithmic Paper)

Figure 5

V. TAKING PERSONNEL REDUNDANCY INTO ACCOUNT IN ESTIMATING TASK RELIABILITY

Previous studies of human reliability have been limited to the behavior of individuals operating independently. Because the Air Force directs that a technician must be accompanied by at least one other individual when entering the silo, the direct application of data acquired by such investigations is somewhat inappropriate for estimating the reliability of Titan II missile-engine mechanics. Because the second individual is likely to be a missile engine mechanic also, his presence should be considered in the manner as one considers redundant equipment.

The principle of redundancy is commonly applied by design engineers. Dewing (Ref 18) defines redundancy as "the existence of more than one means for accomplishing a given task where all means must fail before there is an overall system failure." "Functional redundancy," he states, "applies to systems where two or more means are working at the same time." The latter definition is appropriate for the purpose of this study. If, for example, one operating power generator failed, the standby generator would immediately begin operation, and required electrical power would continue to be transmitted. If the reliability of a single generator is estimated, then the redundancy reliability of the two can be calculated as follows:

$$R_r = 1 - (1-R_1)^n$$

where R_r is the redundancy reliability

R_1 is the reliability of one power generator, and

n is the total number of generators.

If the reliability of one generator is .90 then the equation would be as follows:

$$\begin{aligned} R_r &= 1 - (1-.90)^2 \\ &= 1 - (.10)^2 \\ &= 1 - .01 \\ &= .9900 \end{aligned}$$

V, Taking Personnel Redundancy into Account in Estimating Task Reliability (cont.)

In estimating human reliabilities, a modification of the hardware redundancy assumption is appropriate. Unlike the generator, a second individual may not always be "available" to "back up" the first individual. To assume, however, that he would never be available, would be equally inappropriate. Therefore, When two men are working together to perform a maintenance task, their redundancy reliability may be expressed as follows:

$$R = \frac{1 - (1-R_1)^n \quad (T_1) \; + \; R_1(T_2)}{T_1 \; + \; T_2}$$

where T_1 is the time (in %) that the second man is in a position to react to a potential error of the first man (redundancy time), and T_2 is the remaining time (100% - T_1).

If the same figures are used for human reliability as were used previously in the example of generator redundancy, changing only the percentage of redundancy time, the results should be .9000 but .9900.

$$R_r = \frac{1 - (1-.90)^2 \quad (40) \; + \; .90(60)}{100}$$

$$R_r = \frac{1 - (.10)^2 \quad (40) \; + \; .90(60)}{100}$$

$$R_r = \frac{.99(40) \; + \; .90(60)}{100}$$

$$R_r = \frac{39.6 \; + \; 54.0}{100}$$

$$R_r = .9360$$

For the purpose of the above example, 40 and 60% were selected arbitrarily. However, when applied to specific-problem estimating, it is essential that time percentages be estimated by knowledgeable judges.

In the present study, ratings of percentage of redundancy time (by 10% intervals) were made independently by two engineering psychologists. On 83.3 percent of the 60 task elements, they disagreed by no more than one interval, on eight they differed by two intervals, and on two, by three intervals. The Pearson product-moment correlation between the ratings was .54. In applying the redundancy formula in the present study, the average of the independent ratings used was rounded to the nearest ten-percent interval. Table 9 presents the percentage of redundancy figures used.

VI. CONVERTING TASK-ELEMENT RELIABILITIES INTO TASK RELIABILITIES

The estimated reliability for each Titan II scheduled maintenance task is presented in Table 1. These figures were derived by taking the reliability estimates for task elements from Table 9 and applying them to Type III basic data. The reliability for the task is the product of the reliability of the task elements. To illustrate the method involved, one of the shorter tasks is presented in detail in Table 11. The circled reliabilities in Table 11 are reliabilities for Type III basic data task "Prepare for Turbopump Check." Because two mechanics will normally be assigned to a single task, the most appropriate estimate of the reliability of this task is .9727. This figure was computed by determining the product of all the task-element reliabilities in Table 11. Each of the "Redundant" reliability figures was derived from the corresponding "Individual" figure, using the redundancy formula described in Section V.

There are assumptions, which may not be fully warranted, regarding the use of the product rule. For a discussion of this, see Section X, C, More Elaborate Treatment of Dependency.

TABLE 11

DERIVATION OF TASK RELIABILITY FROM TASK-ELEMENT RELIABILITIES

WEAPON SYSTEM FUNCTION — Subsystems Checks

ACTIVITY — Stage II Engine Checkout

TASK		TASK ELEMENT	INDIVIDUAL RELIABILITY	PER CENT REDUNDANCY	REDUNDANT RELIABILITY
Prepare for Turbopump Torque Check	1.	Read technical instructions	.9901	80	.9979
	2.	Remove the lock wire from the four access cover bolts on the bottom of the roll control duct elbow.	$.9976^4$	20	$.9981^4$
	3.	Remove the four access cover bolts and washers.	$.9971^4$	30	$.9980^4$
	4.	Remove the access cover and gasket from the roll control duct elbow.	$.9973^2$	30	$.9981^2$
	5.	Lubricate the torque wrench adapter threads with approved lubricant.	.9965	10	.9968
	6.	Install the torque-wrench adapter in the threaded boss on the turbopump turbine rotor.	.9967	10	.9970
		Task Reliability (Product of Steps 1 to 6)	.9575		.9727

VII. A COMPARISON OF RELIABILITY ESTIMATES WITH AIR FORCE MAINTENANCE RECORDS

A highly promising source for obtaining empirical data for validation of reliability estimates is Air Force operational maintenance data. The commander of an Airborne Early Warning and Control Wing (AEW&C) of the Air Defense Command cooperated with the Aerojet research team by providing information concerning maintenance-inspection activities in support of RC-121 radar aircraft. A Quality Control Officer of the above organization provided records of performance of thousands of inspections performed by aircraft mechanics. These records were carefully checked by quality control personnel. The tasks were compared for their similarity to the list of 60 task elements described in Table 9. From the list of tasks submitted by the Air Force unit, the following three were selected as being directly related to the present study: (1) "Inspection of exhaust clamps," (2) "Inspection of ignition system," and (3) "Inspect for dents, cracks, and damage of the power-recovery turbine."

Inspection of the exhaust clamps entails the visual inspection and physical manipulation of the engine exhaust system to ensure the security of the bolts and clamps (adjustment is necessary if they are insecure or loose). The part is then reinspected by the quality-control inspector. This task is made up of elements very similar to "Inspect for loose bolts and clamps." Reliability for the Air Force task was .9998 compared to the estimated reliability of .9944 of the present study. The Air Force task similar to the Aerojet element, "Inspect for lockwire," was the "Inspect ignition system." In addition to inspecting for loose wires, ignition leads, etc., the presence of lockwire is checked. Once again, the empirical error rate is similar to the estimated rate (109/14,202 or .0076) compared with an estimate of 1-.9944 or .0056 in the present study. Because other errors entered into the computation, the reliability for "lockwire" might be higher, but no lower, than .9924. The task "Inspect for dents, cracks, and damage of the recovery turbine" is similar to our "Inspect for dents, cracks, and scratches." The Aerojet reliability estimate is .9950, while the Air Force maintenance-inspection reliability, based on 14 errors in 784 attempts, was .9921.

Such data seem to justify confidence in the present approach and support the idea that the technique of deriving estimates of reliability based on ratings of human performance is both useful and accurate. The evident potential of such sources, as the Air Force, to supply empirical data in support of reliability estimates is encouraging. It has been difficult in past to identify sources of such data.

Because the amount of data available was limited to only 3 of the 60 elements, providing congruent rather than predictive validity, plans were made for validating the reliability estimates using Titan II test data.

VIII. PLAN FOR VALIDATION OF RELIABILITY ESTIMATES

The final test of the method developed herein is a comparison of predictions with error data collected in the Titan II silo during the performance of Air Force maintenance. There is, at present, a program at Vandenberg Air Force Base that will provide the needed data. In the program, Category II Personnel Subsystem Test and Evaluation, the performance of missile-engine mechanics is observed and reported. For a period of about one year, each scheduled maintenance task will have been observed approximately four to twelve times. Individual Summary Forms (ISF's) are written at the conclusion of each performance indicating the observer's comments on the quality of performance (Figure 6).

There are two particularly appropriate ingredients in the program that lend it to a validation of the method. First, the ISF reports cover "no deficiencies" as well as "deficiencies". This permits the use of a ratio that can be expressed as a reliability estimate. That is, if an airman opened 30 valves without error, his reliability in the performance of opening valves could be expressed as 30/32 or .9375. Secondly, the ISF data collected during this program are based on observations of performance rather than on inspections following performance.

The reliability estimates were derived from observations of performance, so it is appropriate that they be validated by data derived from observation.

Inspections reveal error that remains in a system after the task has been completed, while "over the shoulder" observations reveal errors that occur during performance. These errors are consequential, though corrected, because they may result in an increase in maintenance time, personal injury, and use of spare parts.

The Individual Summary Forms provide information in a manner that requires minimal conversion for the purposes of data reduction (Figure 6).

In the sample provided, the task was "Perform Stage I Thrust-Chamber Pressure Hot-Gas System Leak Check and Thrust-Chamber Pressure-Switch Functional Check." By reviewing the procedures in Technical Order (T.O.) 21-SM68B-CL-2-1, the number of required lubrications can be counted and recorded. This number becomes the "total number of attemps." The deficiency indicates two errors: the omission of lubricant, and the application of inappropriate lubricant. Therefore, the "number of errors" is two. Each time the task is accomplished at Vandenber Air Force Base, both the number of attempts and the number of errors are counted. At the conclusion of the Category II program, a reliability will be computed of each element for comparison with the estimates.

IX. SUGGESTIONS FOR FURTHER RESEARCH

A. DETERMINING PREVENTIVE MAINTENANCE SCHEDULES

Results of the present study could be used to help resolve the conflict between those who advocate frequent and extensive checkout and maintenance of Titan II engines at operational sites and those who advocate a "hands off" policy once the missile is in the silo. The answer to the problem depends on the trade-off between degree of reliability loss of a ready system through time and the reliability gain or loss resulting from more frequent checkout and maintenance.

FIGURE 6

INDIVIDUAL SUMMARY FORM

No. ____________

Title Perform Stage I Thrust-Chamber Pressure (P_c) Hot-Gas System Leak Check and TCPS Functional Check
Date Observed 1 May & 2 May 1963
Phase/Section 2
T.O. No. & Rev. 21-SM68B-CL-2-1 (37)
Function/Para/Fig. No. 10 (36)
Observer(s) John Doe
Test Location VAFB

Deficiency Classification	Observer	Data Eval. Group
Critical		
Major		
Minor		
Retest Required		
None		

AFSC:	Time:	Hold	
Specified	Specified		
443XLE (2)	2:00	2:20	Fixing High-Press. Leaks in V-99.1 Item 1101.3/ TTU-188/E
Used	Used		
443XLE (2)	5:35		

DEFICIENCY CATEGORY:

Performance

END ITEMS OF EQUIPMENT

Rocket Engine LR87-AJ-5

S/N 9510034 S/A1 & S/A2

DEFICIENCY DESCRIPTION:

Leakage was detected by soap check on 1/4" hard-line "B" nut from TCPS to the injector boss union. The union was removed from the injector boss. Upon reinstallation the technicians did not use thread lubrication on the union end into the injector. They used incompatible lube on the "B" nut connection at the union "B" nut end.

OBSERVER RECOMMENDATIONS:

Provide all compatible thread lubricants in the work area and observe thread-lubrication practives recommended by the engine manufacturer.

Because only limited data are presently available concerning effects of storage on assembled Titan II propulsion subsystems, consideration should be given to the use of adaptive checkout and replacement policies in which the interval between checkout and replacement is adjusted to new information as it accumulates (Ref 19).

Results of studies of the effects of silo storage on the reliability of the Titan II engines should be considered in conjunction with the results of the present study to establish an optimal preventive maintenance schedule. Mathematical models similar to those provided by Kamins (Ref 20) are needed to establish schedules that will maximize readiness and minimize cost.

B. MORE EXTENSIVE COVERAGE OF MAINTENANCE TASKS

The present report covers scheduled maintenance and checkout tasks only. The approach could be extended to cover unscheduled maintenance, trouble shooting, and depot and factory-level tasks. One advantage of the task element approach is that only a few additional task elements are needed to cover many additional tasks of a similar nature. For example, a preliminary inspection suggests that only 15 new tasks elements are needed to cover most unscheduled maintenance tasks. These are as follows:

1. Verify retaining screw in place
2. Inspect for cleanliness
3. Clean surface
4. Align pin holes
5. Install pin
6. Install retaining ring
7. Remove retaining ring
8. Remove pin
9. Cut electrical leads
10. Attach pull wire to leads

11. Pull leads through conduit
12. Install crimp connector
13. Adjust nut
14. Install identifying tag
15. Remove identifying tag

C. MORE ELABORATE TREATMENT OF DEPENDENCY

The use of the product rule assumes that the performance of each element is independent of the performance of other elements in the task. The effect of dependence has been to some extent accounted for in an "average" manner by applying the rule to element performance. There is, however, a need to know more than "average" interaction. In some cases, the interaction of preceding and succeeding elements results in significant changes in either direction from the "average" error rate. For example, in the illustration in Table 10, an access cover cannot be removed until the lockwire, bolts, and washers have been removed. If the mechanic forgets to remove the lockwire, he is likely to recall it during succeeding steps. Interaction may also effect an increase in error. A mechanic may be required to remove a protective cover from a port before connecting a cable to that port. Should the wrong port have been uncovered it is possible that the cable will be connected in error. A comprehensive study of interaction is needed to facilitate accurate determinations of its effect on the reliability of performance.

D. MORE EXTENSIVE TREATMENT OF CONSEQUENCES OF ERRORS

The present approach attempts to account for errors that occur during maintenance. Those errors discovered and corrected by the missile-engine mechanic or by an independent inspection scheduled after the original task performance are discounted. Also, the treatment of the criticality of errors in the present report is recognizably at a gross level. The present approach was taken deliberately in the belief that all errors are undesirable. Even though corrected at a later time, an error may have caused personal injury, increase in maintenance time, or consumption of spare parts. Though all errors are consequential, it would be unrealistic not to recognize that some errors are more serious than others. Future refinements of the method should take these factors into account.

One such factor that may lead to a more comprehensive understanding of the nature of error is the distinction between acts of omission and commission. At the inception of the study, the raters were instructed to make judgements of likelihood of error. Discussions with these judges, after the completion of their task, revealed their desire for more specific descriptions of error. The statements most commonly heard were regarding "things which are done wrong" as opposed to "things which one forgets to do."

The purpose of this study was not to dissect error but to develop a means of predicting the rate of its occurrence. However, there is value in further research of errors of omission and commission because the consequences of these may vary. If, for example, a mechanic forgets to open a valve, there may be no more consequence than a loss of time. He must correct his error to continue the task requiring the valve to be opened. If, on the other hand, the mechanic had opened the wrong valve, then he may have permitted fluids to enter an area, thus incurring degradation. It can also be argued that certain cases of omission may be of greater consequence when they occur because they are not brought to the attention of the mechanic by a succeeding task. Once again, if a valve was erroneously left closed, it is possible that the initiation of an action, which required the valve be opened, could induce severe damage to personnel or equipment as well as abort the task.

Another distinction contributing to a better understanding of the nature of error is that made between "intentional" and "unintentional" error. Intentional error, as suggested by Rook (Ref 4), is "one of conscious awareness by the operator .... the operator intends to perform the act correctly, but erroneously performs it out of limits." Unintentional error is where "there is no element of intent in the performance of the act - it just happens."

IX, D, More Extensive Treatment of Consequences of Errors (cont.)

If, for example, a mechanic solders a wire to an incorrect connecting point, he has committed an intentional error. However, if his soldering iron inadvertently brushed against a wire removing its insulation, he would have committed an unintentional error.

It should be concluded, therefore, that more must be known about the nature of error so that both the reduction of its rate and the prediction of its consequences can be attempted with success..

E. DEVELOPMENT OF METHODS FOR RAPIDLY ESTIMATING TASK RELIABILITY

Occasions arise when a simple approximation of the reliability for a task is needed. In the present study, close approximation to the present results could be obtained by taking the mean task-element reliability of .9963 and raising it to a power corresponding to the number of task elements in a task. Rough reliability estimates for tasks (for which Type III Basic Data are available) could be generated very quickly by estimating the mean-element reliability and determining the number of task elements in each task.

Further research could explore the adequacy of this and other simple approximations.

REFERENCES

1. Payne, D. and Altman, J. W. An Index of Electronic Equipment Operability. Report of Development. American Institute for Research Report, AIR-C-43-1/62 FR. Contract DA-36-039-SC-8055.

2. Esch, L. J. Personnel Subsystem Reliability Analysis. Astropower, Inc, Electronics Products Division, Costa Mesa California, (October 1961).

3. Munger, S. J., Smith, R. W., Payne, D. An Index of Electronic Equipment Operability: Data Store. American Institute for Research Report, AIR-C-43-1/62-RP1. Contract DA-36-039-SC-80555.

4. Rook, L. W. Reduction of Huma Error in Industrial Production, Sandia Corporation Report, SCTM 93-62(14), June 1962 (U).

5. Smith, B. J. Achieving 0.9 Plus Human Reliability in Checkout and Maintenance, Institute of Aerospace Sciences, Proceedings on Aerospace Support and Operations, December 4 to 6, Orlando, Florida, 1961.

6. Williams, H. I. Reliability Evaluation of the Human Component in Man-Machine Systems. Electrical Manufacturing, 4, pp 78-82 (1958).

7. Brady, J. S. Application of Personnel Performance Metric. National Aerospace Systems Reliability Symposium, Salt Lake City, Utah (April 1962).

8. Meister, D. Prediction and Measurement of Human Reliability Proceedings. National Aerospace Systems Reliability Symposium, Salt Lake City, Utah (April 1962).

9. Majesty, M. S. Capt., USAF. Personnel Subsystem Reliability for Aerospace Systems, National Aerospace Systems Reliability Symposium, Salt Lake City, Utah (April 1962).

10. Rabideau, G. F. Prediction of Personnel Subsystem Reliability Early in the Sys-Development Cycle. National Aerospace Systems Reliability Symposium, Salt Lake City, Utah (April 1962).

11. Shapero, A., Cooper, J. I., Rappaport, M., Schaffer, K. H., and Bates, Jr. C. Human Engineering Testing and Malfunction Data Collection in Weapon System Test Programs. WADD Technical Report 60 to 36, Wright Air Development DIV (Contract AF 33(616)-5688, February 1960 (Stanford Research Institute Aerospace Medical Laboratory).

12. Meister, D. Individual and System Error in Complex Systems. Paper Read at American Psychology Association, St. Louis, September 1962.

13. Stephenson, W. A Statistical Approach to Typology. The Study of Trait Universes. J. clin. Psychol., Jan 1950, pp. 26 to 37.

14. Ebel, R. L. Estimation of the Reliability of Ratings. Psychometrika, 16, 4 December 1951, pp. 407 to 424.

REFERENCES (cont.)

15. Thurstone, E. L., and Chave, E. J. The Measurement of Attitude. Chicago: University of Chicago Press, 1929.

16. Hovland, C. I. and Sherif, M. Judgemental Phenomena and Scales of Attitude Measurement: Item Displacement in Thurstone Scales, J. abnorm soc Psychol., 47, pp. 822 to 832 (1952).

17. Boring, E. G. History of Experimental Psychology. 2nd Ed., New York: Appleton-Centry-Crofts, 1950.

18. Dewing, D. L. Reliability Handbook. Aerojet-General Corp., RCS 60-2, March 25, 1960.

19. Jorgenson, D. W. and McCall, J. J., Optimal Replacement Policies for a Ballistic Missile, The Rand Corporation Memorandum, RM-3101-PR-ASTIA AD 276721, May, 1962

20. Kamins, M., Determining Checkout Intervals for Systems Subject to Random Failures. The Rand Corporation Memorandum, RM-2578, (ASTIA AD 247383), 15 June 1963.

GLOSSARY

CATEGORY II TESTING (SYSTEM DEVELOPMENT TEST AND EVALUATION) -- Development testing and evaluation of integrated subsystems through the mating process that progresses into a complete system. Conducted to determine the functional capability of subsystems and redesign requirements. A joint contractor Air Force effort during which the Air Force effort becomes predominant.

HARDWARE RELIABILITY -- The probability that a hardware component or system will perform successfully.

HUMAN RELIABILITY -- The probability that the personnel subsystem will perform successfully. Individual reliability refers to human reliability when the task is performed by a single person, redundant reliability when performed by two persons.

MEASUREMENT RELIABILITY -- The consistency and accuracy with which something is measured.

METHOD OF EQUAL APPEARING INTERVALS -- A scaling method adapted by L. L. Thurstone for the measurement of attitudes. In this method, judges sort a large and representative pool of statements about an object into groups separated by equal steps of intervals. The average of their judgements defines the scale value of a statement.

NORMAL DISTRIBUTION -- A unimodal symmetrical, bell-shaped frequency distribution.

PEARSON PRODUCT MOMENT CORRELATION COEFFICIENT -- A statistic for describing the degree to which two (or more) variables are so related that a change in one is accompanied by a corresponding change in the other. Perfect correspondence is expressed as +1.00, perfect inverse correspondence as -1.00, and complete independence of the variables is expressed as 0.00.

PERSONNEL SUBSYSTEM -- That major functional part of a system that, through effective implementation of the various elements, provides the human performance necessary to operate, maintain, and control the system in its intended operational environment.

REDUNDANCY -- A design factor, for the purpose of increasing the probability of success, that introduces more than one means for accomplishing the given task.

REGRESSION EQUATION -- A formula for computing the most probable value of one variable from the known value of other variables.

SKEWED DISTRIBUTION -- A nonsymmetrical distribution, where the items with the largest frequencies cluster close to one end of the curve, as opposed to a normal distribution.

TASK -- A group of related task elements performed within a work cycle.

Glossary (cont.)

TASK ELEMENT -- A single action (perception, decision, or response) that a person is required to perform in the completion of a task.

TYPE II BASIC DATA (Operation/Maintenance Activities Analysis) -- This analysis translates weapon-system functions in terms of equipment and personnel required in the performance of each identified weapon-system activity. The data are given in tabular form listing derived equipment and personnel performance characteristics against system function. Prepared in accordance with AFBM Exhibit 60-26A.

TYPE III BASIC DATA (Performance Standards (Proficiency) Analysis)--A detailed description at the task-element level of the performance of tasks identified in the Type II analysis. Prepared in accordance with AFBM Exhibit 60-26A.

WEBER-FECHNER LAW -- A psychophysical law showing that the difference in a stimulus just barely noticeable is a constant proportional part of the original stimulus. Thus, human perception of brightness, loudness, weight, etc. is a function of the log of the magnitude of the physical stimulus.

APPENDIX

INSTRUCTIONS TO RATERS USED IN PRELIMINARY STUDY

"It is generally recognized that man is not a perfect machine. In the performance of some tasks he makes many mistakes, while in others he can be expected to make very few. You are being asked to assist us in determining just how well or how poorly man performs various tasks. Others have investigated this question, and examples from their studies may give you some idea of what we are seeking to find."

"For example, in a study of electronic equipment, it was shown that error occurred more often in a soldering operation than in removing insulation from a wire. In another study it was shown that mechanics made more errors in disconnecting locking-type cables than in disconnecting nonlocking type cables."

"On the table before you are ten categories of task errors. Category I represent the least error, and Category X represents the most error. There are 60 cards in front of you, each with a different task printed on it. These are the types of tasks that a missile engine mechanic will perform in the silo during scheduled maintenance. Note that only one card may be placed in Category I, three in Category II, six in Category III and so forth. Assume that missile-engine mechanics had performed each of the tasks 1,000 times. Place the card in the category you believe represents the relative degree of error that probably occurred. If, for example, you think that a certain task could have had so little error, then this task would be put in Category I. If, on the other hand, you believe a task could have been performed with considerable error, for example, 80 times out of the 1,000 attempts, and that no other task could have resulted in as much error, then this task would be placed in Category X. If, as your participation progresses, you wish to change any of your judgements, feel free to do so."

BEST
AVAILABLE COPY

Aerojet-General Corporation, Sacramento, California, REPORT ON HUMAN RELIABILITY IN THE PERFORMANCE OF MAINTENANCE. 13 May 1964, 64 pages, 17 illustrations, (Report LRP 317/TDR-63-218) (Contracts AF 04(647)-521, S.A. 24 and AF 04(694)-424.

Unclassified Report

This report shows the development of a method for estimating the reliability of maintenance performance and demonstrates its application to tasks involved in scheduled maintenance for Titan II engines. A major section involves the combined use of ratings and empirically derived reliability figures. A modification of the design engineer's redundancy formula is developed for estimating the increase in human reliability achieved when two mechanics work together in the performance of a single maintenance task. This study demonstrates that highly consistent ratings of task-element-reliability can be obtained from groups of qualified raters. Plans are described for validating the human reliability estimates obtained during Category II testing at Vandenberg Air Force Base, and suggestions for further research and application of the findings are given.

I. Aerojet-General Corporation
II. Ballistic Systems Division
Air Force Systems Command
III. Report on Human Reliability in the Performance of Maintenance
IV. Contracts
AF 04(647)-521, S.A. 24
AF 04(694)-424

Aerojet-General Corporation, Sacramento, California, REPORT ON HUMAN RELIABILITY IN THE PERFORMANCE OF MAINTENANCE. 13 May 1964, 64 pages, 17 illustrations, (Report LRP 317/TDR-63-218) (Contracts AF 04(647)-521, S.A. 24 and AF 04(694)-424.

Unclassified Report

This report shows the development of a method for estimating the reliability of maintenance performance and demonstrates its application to tasks involved in scheduled maintenance for Titan II engines. A major section involves the combined use of ratings and empirically derived reliability figures. A modification of the design engineer's redundancy formula is developed for estimating the increase in human reliability achieved when two mechanics work together in the performance of a single maintenance task. This study demonstrates that highly consistent ratings of task-element-reliability can be obtained from groups of qualified raters. Plans are described for validating the human reliability estimates obtained during Category II testing at Vandenberg Air Force Base, and suggestions for further research and application of the findings are given.

I. Aerojet-General Corporation
II. Ballistic Systems Division
III. Report on Human Reliability in the Performance of Maintenance
IV. Contracts
AF 04(647)-521, S.A. 24
AF 04(694)-424

Aerojet-General Corporation, Sacramento, California, REPORT ON HUMAN RELIABILITY IN THE PERFORMANCE OF MAINTENANCE. 13 May 1964, 64 pages, 17 illustrations, (Report LRP 317/TDR-63-218) (Contracts AF 04(647)-521, S.A. 24 and AF 04(694)-424.

Unclassified Report

This report shows the development of a method for estimating the reliability of maintenance performance and demonstrates its application to tasks involved in scheduled maintenance for Titan II engines. A major section involves the combined use of ratings and empirically derived reliability figures. A modification of the design engineer's redundancy formula is developed for estimating the increase in human reliability achieved when two mechanics work together in the performance of a single maintenance task. This study demonstrates that highly consistent ratings of task-element-reliability can be obtained from groups of qualified raters. Plans are described for validating the human reliability estimates obtained during Category II testing at Vandenberg Air Force Base, and suggestions for further research and application of the findings are given.

I. Aerojet-General Corporation
II. Ballistic Systems Division
III. Report on Human Reliability in the Performance of Maintenance
IV. Contracts
AF 04(647)-521, S.A. 24
AF 04(694)-424

Aerojet-General Corporation, Sacramento, California, REPORT ON HUMAN RELIABILITY IN THE PERFORMANCE OF MAINTENANCE. 13 May 1964, 64 pages, 17 illustrations, (Report LRP 317/TDR-63-218) (Contracts AF 04(647)-521, S.A. 24 and AF 04(694)-424.

Unclassified Report

This report shows the development of a method for estimating the reliability of maintenance performance and demonstrates its application to tasks involved in scheduled maintenance for Titan II engines. A major section involves the combined use of ratings and empirically derived reliability figures. A modification of the design engineer's redundancy formula is developed for estimating the increase in human reliability achieved when two mechanics work together in the performance of a single maintenance task. This study demonstrates that highly consistent ratings of task-element-reliability can be obtained from groups of qualified raters. Plans are described for validating the human reliability estimates obtained during Category II testing at Vandenberg Air Force Base, and suggestions for further research and application of the findings are given.

I. Aerojet-General Corporation
II. Ballistic Systems Division
III. Report on Human Reliability in the Performance of Maintenance
IV. Contracts
AF 04(647)-521, S.A. 24
AF 04(694)-424