SC-R-67-1150
CONF-670634--1
July 1967

Reprint

THE SANDIA HUMAN ERROR RATE BANK (SHERB)

Lynn V. Rigby, 2152

SANDIA LABORATORIES

OPERATED FOR THE U. S. ATOMIC ENERGY COMMISSION BY SANDIA CORPORATION | ALBUQUERQUE, NEW MEXICO; LIVERMORE, CALIFORNIA

DISCLAIMER

Presented at: Symposium on Man-Machine Effectiveness Analysis Techniques and Data Requirements, sponsored by the Los Angeles Chapter of the Human Factors Society, held at the University of California, Los Angeles, June 15, 1967.

Issued by Sandia Corporation,
a prime contractor to the
United States Atomic Energy Commission

CFSTI PRICES

HC $3.00; MN .65

SC-R-67-1150

THE SANDIA HUMAN ERROR RATE BANK (SHERB)

Lynn V. Rigby
Systems Reliability Division
Sandia Corporation

July 1967

ABSTRACT

For some years, Sandia has been deeply involved in the quantification of human performance, particularly in the development and use of estimates of human error to improve overall system reliability analysis. Resently, we have undertaken to consolidate the various estimates of human error into a single card file, to provide quick access to figures of interest at any particular moment. We call the figures we use "human error rates" (HER), and the file is called the "Sandia Human Error Rate Bank" (SHERB). This paper presents the background for SHERB, describes the format and procedures employed in adding to and using SHERB, and identifies some of the problems surmounted and lessons learned in the process.

THE SANDIA HUMAN ERROR RATE BANK (SHERB)

Introduction

The Sandia Human Error Rate Bank (SHERB) is not exactly an accomplished fact. It is something we have planned for a long time, and do work at occasionally, but it is still merely a small number of file cards contained in a small file box, plus a few rough notes and data not yet transferred to the cards. Nonetheless, we felt that the philosophy, methodology, and experience behind the file and the format used for the file would be of value to anyone with similar interests.

Background

Such a data bank is by no means an original idea. You are doubtlessly aware of the Index of Electronic Operability Data Store developed by the American Institutes for Research (Munger, et al, 1962). This is still the most comprehensive listing of human error available, but the literature contains many other compilations of human error rates, such as the very useful lists compiled by Dunlap and Associates (Mitchell, et al, 1966), Aerojet-General (Irwin, et al, 1964), General Electric (Stave, 1965), and Rocketdyne (Peters, et al, 1965).

Other listings and pertinent data can be found in a wide variety of sources, such as industrial engineering works, quality control reports, safety reports, and the general psychological literature. In fact, SHERB actually began some years ago as a contract with the University of New Mexico in which, in essence, Sandia asked psychology graduate students to search the literature for records of human error rates in production tasks (Hurlock & Peterson, 1963). That preliminary study led to a larger effort, again with the support of the University of New Mexico, and we soon hope to publish a 5000-item bibliography of sources of human performance, and particularly human error, data. This bibliography is now being indexed.

Concurrent with the bibliographic effort, we collected copies in various forms of some 3000 reprints of items listed in the bibliography. These reprints are now on microfilm indexed for quick access. The ultimate goal was, and still is, to convert the usable data in all those documents into a common and easily accessible data file, now called SHERB. Due to the pressures of higher priority tasks, this effort is proceeding slowly, but it is proceeding.

Why SHERB?

Before discussing the file itself, it may be well to consider the basic question, Why SHERB? The human factors group at Sandia is part of the Systems Reliability Division, and its primary purpose is to quantify human performance contributions to system reliability. In order to be meaningful, such quantification must be compatible with common reliability statistics, and the one aspect of human performance that is compatible is human error.

If human error is defined to be any variant of human performance that reduces the probability of system or mission success, then failures due to human errors can be treated in a manner very similar to component failures; that is, human errors can be predicted as a probabilistic function of the variables determining or influencing that human performance related to system performance.

The prediction techniques employed at Sandia have been described by Rook (1962, '64, and '65) and Swain (1963, '64, '66, and '67). These techniques depend primarily upon a detailed functions and task analysis; the preparation of logic tree diagrams to allow analysis of the relevant inputs, outputs, interactions, pertinent variables, and consequences;

the estimation of the probability associated with each limb of the tree diagram; and the appropriate probability statistics.

In any human task, a large number of discrete inputs, outputs, and influencing variables come into play; and the human error analyst must be able to assign occurrence and error probabilities to all of those that can effect system failure. Despite our preferences for scientific rigor, there is seldom time or funds to conduct experiments to obtain situation-specific data; so we must depend, and depend heavily, upon our ability to extrapolate from the known to the unknown, however unlike the two may be.

SHERB, past experience, and whatever can be found in a quick look at the literature constitute our pool of knowns for any given application. It is an inexact and heterogenous pool and, despite care and expertise in interpretation, our predictions can be considerably in error. But though accuracy is to be desired and sought, inaccuracy is no bar to our efforts.

Whenever we feel strongly enough about an error-likely situation to make an issue of it, we find others easy to convince that human error is so important that gross predictions are better than none. Usually, no one is really concerned with the accuracy of our figures, yet almost everyone is willing to listen if we have figures; and they are willing to accept the figures as reasonable once the basis and implications are presented. Such experience merely underscores three common expectations:

1. Scientists and engineers fully expect human performance to have a large impact on system performance; they need only to be shown how and to what degree.

2. Numbers are the fundamental structure of any decision fabric in any scientific and engineering environment.

3. The contribution of a human error analyst is primarily dependent upon how quickly he can produce relevant and acceptable estimates.

Thus, the more data we have in SHERB, the larger our pool of "knowns," the better qualified we are to make predictions, the more confidence we have in those predictions, the more work situations we can address, and the more frequently and more quickly we can contribute to a fuller and more accurate interpretation of system success or failure.

The SHERB Format

As it now stands, SHERB consists of a number of 5 x 8 inch file cards. These cards are pre-printed in the format provided in Figures 1 and 2, which show the front and back sides, respectively. Data are entered upon the cards by hand or typewriter, and the cards are filed alphabetically by task. The number of cards is small, but will increase in time; and as the file grows, more sophisticated filing and cross reference systems can be readily applied, but these are not yet necessary.

In using the file, we simply flip through the cards until we find data appropriate to the task or error we are interested in. If there is more than one card for that task or error, we must decide which set of data is most appropriate (or least inappropriate). If there is no suitable information in the file, we must develop estimates from some other basis. This usually requires some literature search, a paper analysis, and a lot of soul searching. The information on the card ordinarily fills our immediate needs, but the reference can be readily checked for further details and background.

SHERB CARD, Sandia Corporation

Mean HER:	.0040
Std. Dev.:	n/a
Lo Range:	n/a
Hi Range:	n/a
Distr. Shape:	n/a
N Opport.:	12,587
Job Area:	Criterion
Kind Data:	Criterion
N Subjects:	n/a
Kind Subjs:	n/a
Work Envir.:	n/a
Climate:	n/a

	-3	-2	-1	0	+1	+2	+3
Task Stress			Varied				
Workspace HE			"				
Equipment HE			"				
Qual Perf Aid			"				
Qual Support			"				
Reliability					x		
Validity							x
Generality				x			
Source Cred.							x

TASK: Connectors, AN/TRI-Lock* ERROR: QEST Found Defective

AREA: All, Criterion Data CRITERION: QEST

DATA BREAKDOWN:

QEST Deficiencies noted in AN & TRI-Lock Connectors	Number Occurrences	% of Errors	HER
Number of connectors inspected	12,587	--	---
Connectors w/bent pins	19	37%	.0015
Connectors w/external damage	11	22	.00087
Connectors improperly mated	9	18	.0007
Connectors w/parts omitted**	12	23	.00095
Total connector errors:	51	100%	.004***

*Based on old type tri-lock, pre scoop-proof design.

**Probably assembly errors.

***p defective connection due to one or more human errors.

REVIEWER: L. V. Rigby ORG. 2152 DATE: 1 Jun 1967

Figure 1. The Front Side of a Typical SHERB Card

As shown in Figure 1, the top of the SHERB card provides for topic descriptions of the interest area, task, type of error, and criterion for error. These blanks are filled with such representative topics as:

Area	Task	Error	Criterion
Assembly	Access	Abuse	Accident
Communication	Checkout	Interchanging	Accuracy
Design	Connection	Mismating	Completion
Inspection	Disconnection	Misreading	Consumption
Installation	Display, linear	Misuse	Cost
Maintenance	Fastening	Omission	Injury
Measurement	Fault diagnosis	Reversal	Man time
Operation	Handling	Substitution	System time

Along the left side of the card shown in Figure 1, the basic data descriptors are recorded; these include the mean human error rate, the standard deviation or comparable distribution parameter, the range, and the shape of the distribution, where these can be determined. By human error rate we mean the probability of error per opportunity for error. Such information, of course, allows some latitude in extrapolation. For instance, if the data are applicable to a situation in which other parameters seem notably higher or lower, we may choose some ordinate other than the mean as the basis for prediction. Any such choice is both the exercise and the proof of expertise, but the logic becomes tenuous to the degree that distribution parameters are unknown.

In recording the data, we use whatever significant digits are provided by the source, and leave any rounding to the instance of use, although one significant digit usually reflects the accuracy of the data. The figures are listed as decimals, for example, as 0.0021, rather than 21×10^{-4} or to some standard base such as 10^{-6}. Decimals are more easily grasped and more commonly understood, at least up to five or six decimal places.

In the "No Opportunity" blank, we fill in whatever denominator information is provided. This seems to be an inadequately understood area. In any assembly task, for instance, it is not sufficient merely to record the number of soldering errors per number of units produced. In order to be fully meaningful, the data must show the number of soldering points per unit, at least. It is also helpful to show any differences among the soldering points that might make a difference in either frequency or type of error. For instance, were all wires inserted through holes and soldered, or were some looped, wrapped, or pigtailed?

Similarly, brief topic descriptors are used to identify the job area, the kind of data, the kind and level of subjects, the working environment, and the climatic conditions the data were obtained under. The number of subjects is taken as given in the source, and representative topics in each of the other areas include:

Job Area	Kind Data	Subjects	Work Envir.	Climate
Auto driver	Accident/Incident	Analysts	Airborne	Arctic
Clerk	Deficiency report	Naive	Factory	Desert
Navigator	Feedback data	Task skilled	Field unit	High altitude
Pilot	Field test data	Tech reps	Laboratory	Indoor, Std.
Secretary	Lab experiment	Semi-skilled	Office	Under sea
Technician	Q/A inspection	Students	Space-borne	Z.I.

Such topics merely indicate the general conditions under which the data were obtained, and the next few rows identify and evaluate the major assumptions underlying the data, particularly:

The stress level the subjects were working under
The quality of workspace human engineering
The quality of equipment human engineering
The quality and representativeness of performance aids used
The quality of supply and support employed or assumed

The above are rated on a seven-point scale via checks made directly on the SHERB card, as shown in Figure 2. The values in the scale indicate the following ranges:

-3 = worse than -3σ (~ worst 0.1%)

-2 = between -2σ and -3σ (~ 2%)

-1 = between -1σ and -2σ (~ 14%)

0 = $\pm 1\sigma$ (~ 68%)

+1 = between $+1\sigma$ and $+2\sigma$ (~ 14%)

+2 = between $+2\sigma$ and $+3\sigma$ (~ 2%)

+3 = better than $+3\sigma$ (~ best 0.1%)

The use of this kind of scale is not intended to imply greater accuracy in rating; rather, it simply forces us to think in terms of a normal distribution of events. The great majority of events are "more or less average," and they receive the middle, or zero, rating. This kind of rating scale seems to be more useful and more appropriate to probability analysis than a linear scale.

Similar evaluations are made of the statistical reliability (repeatability), validity re the test or experimental situation, generalizability of the data beyond the test or experimental situation, and credibility of the source. Such notes, which are largely subjective, are merely reminders of the general limitations of the data. We may ignore these limitations, but at least we know what they were or seemed to be.

The rest of the card is essentially unstructured. The front allows condensation of any detailed breakdown of the data, as illustrated in Figure 1; and the reviewer is identified by name, organization, and date at the bottom of the card. Where others in the human factors group are familiar with the source work, we have them review and corroborate the evaluation.

The back of the card, as illustrated in Figure 2, is filled with abstracted narrative in accordance with the following instructions:

1. Task description. What task was being performed when the error was made? How frequently was this task performed? What kinds of activities intervened? What were the task inputs and outputs? And how was the task performed?

2. Error description. What was the nature of the error class or classes? What tolerance limits or requirements defined the error? And what criteria were used in the tabulation of error?

3. Situational variables. In general, what was the situation in which the task was performed and errors made? Were any key independent parameters important to definition or interpretation of errors? Were there conditions which may have systematically increased or decreased the chargeability, detectability, or recordability of errors? Were there any artifactual restrictions which may influence the generalizability of the findings? If there was any analysis or test of significance, show the procedures employed, results obtained, and conclusions drawn.

4. Source. Provide a complete bibliographic reference--authors, title, document number, publisher, city and state, date, DDC or other reference number, classification, and page reference.

All of the foregoing matters are completely dependent upon the information provided by the source. If the source does not make such matters clear, we can either estimate the apparent conditions or leave the card blank in that area. In either case, we have just that much less of an idea of how relevant

DESCRIBE TASK:	These data constitute all connector deficiencies disclosed by QEST (Quality Evaluation System Test) between Jan. 1960 and Aug. 1961, for varying numbers of different kinds of nuclear weapons.
DESCRIBE ERROR:	Errors recorded are all defects which would limit the operability of the connection. Except where shown, these errors are most likely attributable to the last installation action.
DESCRIBE SITUATION:	The data listed are criterion data in that QEST exhaustively and systematically reveals all deficiencies in the equipment inspected. These, then, were the actual and total number of connector problems disclosed in that time period. Classified details are provided in the source document.
KEY VARIABLES RESTRICTIONS:	
SOURCE:	

Figure 2. The Back Side of a Typical SHERB Card

the data are to any potential application. Of course, these are the kinds of information which are, or should be, provided by even reasonably thorough research reporting.

Data Sources and Interest Areas

The data incorporated into SHERB comes from many sources. Most of it is extracted directly from the literature, particularly works already mentioned. Some of it is derived from Sandia development and field tests, some from special Sandia studies (unpublished), and some of it consists of estimates that we have had to develop at one time or another and keep on file for later use. A summary of the major kinds of data encountered, and estimates of their relative merits, is provided in Table I.

With the present paucity of such data, we really don't do much in the way of selection. If we can find it, we will use it, at least until better is available. But the information must be convertable to the probability of error per opportunity for error; data which do not have good denominator information are essentially useless, except to indicate failure events or modes. We are, of course, primarily concerned with four broad species of human error:

1. Assembly errors are human errors committed in component and equipment production, which somehow pass acceptance procedures and remain undetected until they cause problems in the field. These include both things like soldering errors, which eventually cause failures outright, and defects which may contribute to other errors, such as an off-center handle or control, etc. Incidentally, we are beginning to believe that undetected

TABLE I

Evaluation of Human Error Data Available

Kind of Data	Availability	HER Coverage	HER Reliability	HER Validity
Q/A In-Plant Inspections*	Good	Poor	Poor	Poor
Individual opinion, no analysis	Good	Good	Poor	Poor
Acceptance test data*	Fair	Poor	Fair	Fair
Individual analytic estimate	Poor	Good	Fair	Fair
Accident/Incident data summary*	Good	Poor	Fair	Fair
In Work Deficiency Reports*	Poor	Poor	Fair	Good
Field Feedback Data*	Fair	Poor	Fair	Good
Accident/Incident data, raw*	Poor	Poor	Good	Good
Field Test Data*	Fair	Poor	Fair	Good
Mean of Scaled Opinion	Poor	Good	Good	Good
Experiment in Work Situation	Poor	Good	Good	Good
Quality Evaluation System Test	Good	Fair	Good	Good
Laboratory Experiment	Good	Good	Good	Good

*Assuming good denominator information, which is usually lacking.

assembly error is the primary source of unreliability, particularly in equipment composed of highly reliable components.

2. Installation errors are human errors committed in the installation or integration of a unit into a larger equipment or facility complex. Like assembly errors, installation errors may have long lasting effects on total system reliability, particularly if we include the integration of operational procedures.

3. Operator errors are human errors committed in the operation of the equipment and associated transport, handling or support equipment. The effects of such errors are directly related to both equipment reliability and mission success or failure.

4. Maintenance errors are human errors committed in the performance of equipment maintenance, which directly influence equipment reliability and thereby indirectly influence mission success or failure. Maintenance can also directly influence mission success.

Taken in aggregate, the above account for a large portion of total system failure. Just how much is a matter of growing concern, and this concern, we hope, will be accompanied by increasing attention to systematic prediction and measurement of human error. Our own experience indicates that the percentage of system failures caused by human error is at least as high as the 50 to 60 percent suggested by the classic studies of Shapero (1960) and Zeller (1955) and can be as high as 80 to 90 percent in some cases.

Unfortunately, accidents and mission failures resulting from human errors that do not result in equipment failures are not reported with the same regularity and accuracy as equipment failures. And even the reporting of equipment failures omits much good human error data. Our greatest need is still for good feedback data to tell us not only what the real problems are, but what the actual error rates are. If we know the error rates, we can plan around them or try to reduce them and evaluate the effectiveness of whichever course is taken.

We do have unpublished, classified data showing that mission failure due to human error is four times as frequent as that due to component failure in weapon drop tests. We also have a rough idea as to how the various species of human error are generally related to the total life cycle of equipment, and these are diagrammed in Figure 3.

The effects of assembly and installation errors, of course, tend to decrease with time as faulty units are detected and replaced in equipment checkout, maintenance, and retrofit programs. There is usually a slow startup of operations and some initial learning effect in both operator and maintenance errors; then, the operator error rates tend to stabilize, but maintenance errors tend to increase with increases of component failures during the wearout phase of components. This is a rough notion, but it may give you something to think about, for it has implications for the question: What are we predicting to? And it has some relevance to the meaning of error rate data collected at different phases of the life cycle.

Second only to the lack of field feedback data, the major problem in human error analysis is the variety and unevenness of the data available. Of necessity, we must often use data at its face value, but the data vary widely in terminology, manner of development, and level of reporting. Any efforts at standardization of these matters will greatly aid the progress of prediction techniques.

Along these lines, we prefer to call our figures "human error rates," because this is a straightforward, unequivocal, and generally acceptable

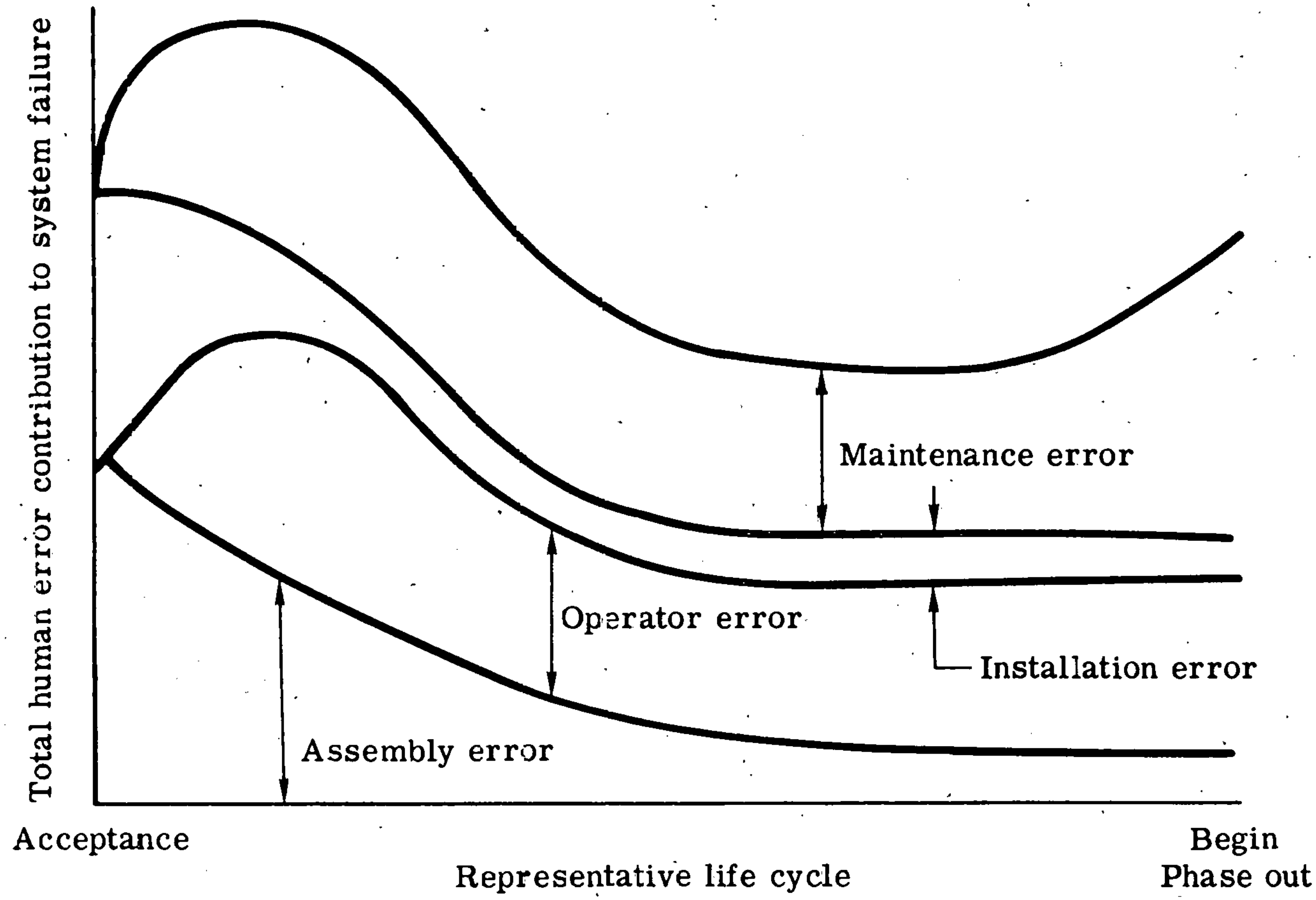

Figure 3. Proportional Contribution of the Different Species of Human Error to System Failure.

concept; it describes exactly the kind of information we can use most effectively; and the acronym, HER, is guaranteed to get attention. More euphemistic terms such as "human reliability," "zero defects," or "human success probability" mean different things to other specialists, such as flight surgeons, quality inspectors, and personnel people.

Most people seem to be ready to accept the fact of human error, and this fact can be dealt with more effectively if dealt with openly. Too, if it is called "human error," it is more likely to be dealt with by behavioral scientists, as it should be. It is both useful and important, however, to distinguish, as Rook (1965) does, between situation-caused errors (SCE) and human-caused errors (HCE). Emphasis on SCE, especially when setting up error collection programs, helps remove the unfortunate and inappropriate onus attached to the words "human error."

Concluding Notes

SHERB, then, is a small file as yet; more an idea than an actuality. But it is growing, and it is a very useful and necessary adjunct to human error prediction, for the accuracy of such predictions and the effort required to develop them depend heavily upon the availability and accessibility of reasonably solid and generalizable data, upon the "knowns" of human performance.

When the file is more presentable, perhaps it can be published in full. In the meantime, we would be interested in exchanging such information with those of you who are developing comparable files of your own. And for those of you who are not developing such files, may we suggest that you consider it. You will be surprised at how useful it will become. Obviously, the data currently available leave much to be desired. Merely complaining about this will accomplish little. Rather, it is the responsibility of every human factors specialist to specify what he needs, to determine how it should be collected, and to state clearly the value of having it. As soon as the human factors community acts in concert in this fashion, we will have good human error rate date; and there does not seem to be any aspect of human or man-machine performance that cannot be meaningfully interpreted in terms of human error.

REFERENCES

Hurlock, R. E. and Peterson, G. M., A survey of the literature on human error. Albuquerque, N. M.: University of New Mexico, January 1963.

Irwin, I. A., Levitz, J. J. and Freed, A. M., Human reliability in the performance of maintenance. Sacramento, Calif.: Aerojet-General Corp. Report LRP 317/TDR-63-218, May 1964.

Mitchell, M. B., Smith, R. L. and Verdi, A. P., Development of a technique for establishing personnel performance standards (TEPPS): Phase III - final report. Santa Monica, Calif.: Dunlap & Assoc., Inc., July 1966.

Munger, S. J., Smith, R. W. and Payne, D., An index of electronic equipment operability: data store. Pittsburgh, Pa.: American Institutes for Research Report AIR-C43-1/62-RP(1), January 1962.

Peters, G. A., Hall, F. S. and Kuplent, C., Human reliability data, Canoga Park, Calif.: Rocketdyne Report IDEP 347. 90. 00. 00-G1-03, June 1965.

Rook, L. W., Reduction of human error in industrial production. Albuquerque, N. M.: Sandia Corporation Technical Memorandum SCTM 93-62(14), June 1962.

Rook, L. W., Evaluation of system performance from rank-order data. Human Factors, 1964, 6, 533-536.

Rook, L. W., Motivation and human error. Albuquerque, N. M.: Sandia Corporation Technical Memorandum SC-TM-65-135, September 1965.

Shapero, A., Cooper, J. I., Rappaport, J., Schaeffer, R. H. and Bates, C., Jr., Human engineering testing and malfunction data collection in weapon system test program. Wright-Patterson Air Force Base, Ohio: Wright Air Development Center Technical Report WADC TR 60-36, February 1960.

Stave, A. M., The quantification of human reliability, a feasibility demonstration, Philadelphia, Pa.: General Electric Spacecraft Department Report TIS 65SD216, March 1965.

REFERENCES (cont)

Swain, A. D., A method for performing a human factors reliability analysis. Albuquerque, N. M.: Sandia Corporation Monograph SCR-685, August 1963.

Swain, A. D., THERP. Albuquerque, N. M.: Sandia Corporation Reprint SC-R-1338, August 1964.

Swain, A. D., Field calibrated simulation. Albuquerque, N. M.: Sandia Corporation Reprint SC-R-67-1045, February 1967.

Swain, A. D., Some limitations in using the simple multiplicative model in behavior quantification. Reliability of human performance in work: a symposium of the 1966 annual convention of the American Psychological Association. Wright-Patterson Air Force Base, Ohio: Aerospace Medical Research Laboratories Technical Report AMRL-TR-67-88, in press.

Zeller, A. F., Human limitations and aircraft design, Air Force-Industry Conference, 1955.