DOCUMENT RESUME

ED 079 349 TM 002 948

AUTHOR	Askren, William B.; Regulinski, Thaddeus L.
TITLE	Quantifying Human Performance Reliability.
INSTITUTION	Air Force Human Resources Lab., Wright-Patterson AFB, Ohio. Advanced Systems Div.
REPORT NO	AFHRL-TR-71-22
PUB DATE	Jun 71
NOTE	17p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Data Collection; *Human Engineering; Man Machine Systems; *Mathematical Models; *Reliability; Statistical Analysis; *Task Performance; Technical Reports

ABSTRACT

Human performance reliability for tasks in the time-space continuous domain is defined and a general mathematical model presented. The human performance measurement terms time-to-error and time-to-error-correction are defined. The model and measurement terms are tested using laboratory vigilance and manual control tasks. Error and error-correction data are ordered and the underlying density functions isolated. The Weibull distribution is best fit for time-to-first-error data, and the Log-Normal distribution is best fit for time-between-errors and time-to-error-correction data. The Normal distribution is rejected in all cases. Distribution parameter values are applied to the general mathematical model, and prediction made of human performance reliability for the tasks. It is also shown that task performance reliability improves with training on the tasks. (Author)

AIR FORCE HUMAN RESOURCES LABORATORY

AFHRL-TR-71-22

JUNE 1971

QUANTIFYING HUMAN PERFORMANCE RELIABILITY

WILLIAM B. ASKREN

ADVANCED SYSTEMS DIVISION
WRIGHT-PATTERSON AFB, OHIO

AND

THADDEUS L. REGULINSKI

AIR FORCE INSTITUTE OF TECHNOLOGY
WRIGHT-PATTERSON AFB, OHIO

Approved for public release; distribution unlimited

Presented at the Psychology in the Air Force Symposium conducted at the Air Force Academy, Colorado, 20-22 April 1971

AIR FORCE SYSTEMS COMMAND

BROOKS AIR FORCE BASE, TEXAS

ED 079349

AFHRL-TR-71-22

QUANTIFYING HUMAN PERFORMANCE RELIABILITY

WILLIAM B. ASKREN

AIR FORCE HUMAN RESOURCES LABORATORY

THADDEUS L. REGULINSKI

AIR FORCE INSTITUTE OF TECHNOLOGY

TECHNICAL REPORT AFHRL-TR-71-22

JUNE 1971

AIR FORCE HUMAN RESOURCES LABORATORY
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO

QUANTIFYING HUMAN PERFORMANCE RELIABILITY

William B. Askren, Air Force Human Resources Laboratory

and

Thaddeus L. Regulinski, Air Forc. Institute of Technology

PROBLEM

The characteristics of new Air Force systems are determined early in the development cycle as a result of engineering, operational and cost analyses. If truly effective systems are to be developed, it is necessary that data describing the capabilities of the human resources of the Air Force be included in these analyses. For this to be feasible, the human resources data need to be provided in forms useful in analytical studies.

One class of human resources data relates to personnel skill. However, the means for incorporating personnel skill data in analyses of systems does not exist (Askren and Regulinski, 1969). Morever, the capability does not exist for determining the effect on man-machine systems of other human resources parameters such as training effects. Therefore, research directed at the quantification and mathematical modeling of personnel skill and training effects for application to system analytical studies is being performed. The RELIABILITY of personnel performance is being used as the measure of skill, because of its importance to system analyses, and its usefulness as an index of skill improvement as a result of training.

Classical engineering reliability analysis uses statistical deduction to translate time of equipment failure observations to a relevant model. The prediction of reliability is obtained then from the model via probability theory. In the time continuous domain this procedure requires knowledge of an analytical stochastic function, e.g., the probability density function, of failures of the equipment with respect to time for the operations involved. Also, classical reliability modeling employs the first moment of the random variable which is known variously as mean-time-to-failure, mean-time-to-first-failure, and mean-time-between-failures (Sandler, 1963). The specific objective of the research reported in this paper was to determine the feasibility of applying this classical method to the analysis of human performance, and to determine the effect that different amounts of training have on the reliability of human performance.

PROCEDURES

The research involved a number of operations. First, a general model of human performance RELIABILITY, was propounded. Then the appropriateness of the first moment of the random variable TIME, as a quantifier of human errors, was established. Next, experimental tasks were set up to generate human error data. Then, probability density functions of the errors were determined; these functions permitted use of the general model to predict human reliability for task performance. Finally, the effects of learning on the reliability of performance were determined.

RESULTS

Human Performance Reliability Model

Equipment reliability is generally modeled using time-space continuous or time continuous-space discrete stochastic models. The human performance tasks that are most analogous to equipment operation, and thus most amenable to this form of modeling are continuous operation tasks such as vigilance, monitoring, and tracking. Consequently, human performance reliability for this family of tasks was modeled.

Human performance reliability, for tasks in the time-space continuous domain, is defined as the probability that a given task will be correctly performed, subject to time constraints, and the stress constraints inherent in the nature of the task, the operator, and the environment.
This definition may also be expressed as:

$$R_h(t) = P \text{ (task performance without relevant errors under constraint of time and stress).} \quad (1)$$

This statement of human performance reliability was translated into an analytical-stochastic function through a series of derivations (Regulinski and Askren, 1969) and resulted in the equation:

$$R_h(t) = \exp\left\{-\int_0^t e\ (t)\ dt\right\} \quad (2)$$

where: R (t) = the reliability of human performance for any point in time of task operation, and,

e (t) = the error rate for the specific task.

This equation is proposed as the general model for the reliability of human performance for tasks in the time-space continuous domain.

The Random Variable Time-To-Human-Errors

In reliability engineering, the term mean-time-to-failure (MTTF) is applied to components that are not repairable and are throw-away items, such as fuses and light bulbs, whereas mean-time-to-first failure (MTTFF) and mean-time-between-failures (MTBF) are applied to equipment subject to repair. The three terms are useful in dealing with human performance reliability. MTTF translates into mean-time-to-human-initiated-failure (MTTHIF) and describes when a system function could be expected to fail as a result of an error or an accumulation of errors by one or more persons performing tasks in that function, e.g., overpressurizing a missile fuel tank, undershooting an aircraft landing, or inadvertently actuating an ejection seat.

MTTFF and MTBF translate into terms which describe errors whose effects are correctable. Thus, MTTFF transforms into mean-time-to-first-human-error (MTTFHE). This is useful in treating errors that are highly critical, such that the first occurrence of an error would be costly, or establish hazardous conditions, e.g., failing to detect a target on a radar scope or not inserting an ejection seat safety pin prior to performing maintenance work. The term MTBF converts to mean-time-between-human-errors (MTBHE). This is useful in treating errors of a less critical nature, and could be used, for example, to provide information regarding the frequency of production of defective parts, or an indication of the proficiency level of personnel.

One additional measure was determined to be necessary. This relates to the very unique characteristic of man which sets him apart from the machine. Man can correct his error. Thus, a term was needed which would describe this capability of man, and could serve as a supplement to the MTTFHE and MTBHE quantifiers. The description in this case comes from the field of maintainability engineering, in which the expression mean-time-to-restore (MTTR) is used. This indicates the time, on the average, taken to repair malfunctioning equipment. MTTR transforms into two useful human performance terms. The first is mean-time-to-first-human-error-correction (MTTFHEC), which indicates the time, on the average, for man to correct his first error. However, man, during the course of a work period may commit a number of errors, yet recover from them. Thus, a second term is necessary. This is mean-time-to-human-errors-correction (MTTHEC), and indicates the time, on the average, for man to correct all of his errors.

Experimental Tasks

Two separate experimental tasks were set up to generate error data for testing the model, and for testing the time-to-error and time-to-error-correction terms, and the effects of learning. The first utilized a vigilance task, and the MTTFHE quantifier. The second involved a manual control task, and the MTTFHE and MTBHE quantifiers. This experiment also employed the MTTFHEC and MTTHEC quantifiers. In addition, the second study was designed to test the effects of learning on the reliability of performance and the ability to correct errors.

In the first study, a vigilance task was used with subjects required to observe a circular light display and respond to a failed-light event by pressing a hand-held switch. Miss and false-alarm error data were collected. A miss error indicated that the subject did not detect the failed-light event. A false alarm denoted an error by anticipation. The subject responded as if a failed-light event occurred, when in fact the event did not occur. Fifty-one male and female subjects were used.

In the second study, a two-axis tracking task was used which simulated aircraft flight effected by random disturbances. The subject was required to operate a manual control stick which regulated instrument display needles representing pitch and roll motions of the aircraft. The subject was required to hold the two needles between limits set for each axis. Crossing the limit signalled the occurrence of an error, and also signalled the beginning of time for human error correction. Returning within the limit signalled the completion of human error correction time. Each subject had two "flight" trials separated by a rest period. Sixty-three male subjects were used.

Probability Density Functions of the Error and Error Correction Data

From the first experiment the data of times to first-miss-error, first-false-alarm-error and to combined-false-alarm-and-miss-error were analyzed to determine the relevant distributions. The Weibull distribution yielded best fit, and was significant at the .10 level using the Kolmogorov-Smirnoff test. The Weibull parameter values are given in Table 1.

A description of the density functions of the data from the second experiment is more complex. Distributions were sought for error data, and error correction data for the pitch axis, the roll axis, and for both first and second trials. The results of the analysis are summarized in Tables 2 and 3. Mean-time values are given, but other distribution parameter values are not listed for simplicity of reporting.

Prediction of Human Performance Reliability

Predictions of human performance task reliability may now be made. Using data from experiment one, the vigilance task, reliability for any time period may be predicted using the two-parameter Weibull reliability function derived from the general model:

$$R(t) = \exp\left\{-\left(\frac{t}{a}\right)^b\right\} \tag{3}$$

where a and b are respectively the scale and shape parameters. For example, if reliability of the vigilance task is defined as the probability of performance precluding both miss and false alarm errors, the reliability for t=60 seconds is predicted to be .70. This is accomplished by solving equation (3) using from Table 1 the values a=267.75 and b=.7. Inspection of Table 1 also shows that the mean-time-to-combined-errors is 315.82 seconds. Similar predictions may be made for miss errors alone, or false alarm errors alone.

A variety of task performance predictions may be made using the data from experiment two. For example, we may predict the probability of error free performance from the beginning of the task to any point in time. Using data from the pitch axis, trial #2, and the Weibull two-parameter reliability function, the probability of error free performance of 30 seconds duration of the task is predicted to be .527. Other predictions which may be made include: the probability of correcting the first error within a given time period; the probability of a given time between errors; and, the probability of a given correction time for all errors.

Effects of Learning

Data from the second experiment provide an indication of the effects of learning on the reliability of task performance. Data from trial #1 indicate untrained performance, whereas data from trial #2 indicate a degree of trained performance, since it follows trial #1 after a suitable rest period. Inspection of Tables 2 and 3 show that all mean values improve, as would be expected. For example, in Table 2, the pitch axis mean-time-to-first-error increases from 14.6 to 100.6 seconds, and the mean-time-to-first-human-error-correction decreased from 3.1 to 2.3 seconds.

The amount of learning also effects other performance predictions that could be made. For example, earlier in the paper it was predicted that the probability of error free performance in the pitch axis for the first 30 seconds of the task is .527. This was based on data from trial #2, the "trained" group. This same prediction based on data from trial #1, the "untrained" group is .139.

Finally, Tables 2 and 3 show the distributions which govern the data. Inspection of these results shows that the distributions are the same, with a single exception, for trial #1 and trial #2 data. This suggests that the nature of the human response to the task does not change with training, rather the response becomes "better." It also may be feasible to make extrapolations of performance improvement after additional learning trials, by changing the parameter values of the density function which governs the task situation.

CONCLUSIONS

It is concluded that equation (2) is a useful general model of human performance reliability in the time-space continuous domain. The human reliability function may be defined as the probability of successful task performance within temporal constraints, thus allowing predictions of task reliability for various time intervals. It is also concluded that the two terms, time-to-error and time-to-error-correction, when used together, serve to more fully describe man's performance in the system. This more complete information would permit the system engineer to determine if skilled personnel performance is compatible with the response characteristics of the equipment to be operated, and the reaction times of the operational situation.

Another conclusion relates to the types of distributions that govern the data. In both experimental studies, it was found that the Weibull distribution best fits the time-to-first-human-error-data. In the second experiment, it was found that Log Normal best fits the time-between-human-error data and both types of error-correction data. However, in neither study did the Gaussian (normal) distribution reach statistical significance. In fact, of the ten distributions examined, the Gaussian was the worst fit for the data. Therefore, the conclusion seems quite justified that human error and human error correction data are not normally distributed. Consequently, studies dealing with human performance reliability in man-machine systems should not arbitrarily assume human error data to be normally distributed, but should seek the distributions relevant to the task. Asiala, after study of portions of these data arrived at the same conclusion (Asiala, 1969). In the interim, it is proposed that human error data be modeled by either the Weibull or Log Normal functions.

It is also concluded that these results could be used to determine how much training should be given to personnel, and to perform trade-offs between training effects and system design. Given particular system requirements (Probability value P for X time without human errors, and Y time for error correction), given a particular task, and given knowledge of the density function governing human performance of the task, a determination may be made of how many training sessions are required to provide the human performance which meets these requirements. The trade-off studies would reverse the process and test the effect of various amounts of training on system effectiveness.

Finally, it is recommended that future data gathering efforts in the human performance reliability area should use standard performance quantification terms, and should use tasks and personnel representative of the types of systems to which the data are to be applied. It is recommended, of course, that the quantification method described in this paper be used in these data collection efforts.

REFERENCES

1. Asiala, C. F., Digital Simulation Model for Fighter Pilot Workload. MDC Report A0058, McDonnell-Douglas Aircraft Corporation, St. Louis, Missouri, September, 1969.

2. Askren, W. B. and Regulinski, T. L., Quantifying Human Performance for Reliability Analysis of Systems. Human Factors, Volume 11, No. 4, August 1969.

3. Regulinski, T. L. and Askren, W. B., Mathematical Modeling of Human Performance Reliability. In the proceedings of the 1969 Annual Symposium on Reliability, Chicago, Institute of Electrical and Electronics Engineers, Inc. IEEE Catalogue No. 69C8-R, Pages 5-11, January 1969.

4. Sandler, G. H., System Reliability Engineering. New York: Prentice-Hall, 1963.

TABLE 1

WEIBULL PARAMETER VALUES FOR VIGILANCE TASK ERROR DATA

Type of Error	Scale Parameter a	Shape Parameter b	Mean (seconds)
Miss	682.94	1.292	633.26
False Alarm	228.68	0.657	309.04
Combined	267.75	0.700	315.82

TABLE 2

DISTRIBUTIONS GOVERNING MEAN-TIME-TO-FIRST-HUMAN-ERROR AND MEAN-TIME-TO-FIRST-HUMAN-ERROR-CORRECTION DATA FOR 2-AXIS MANUAL CONTROL TRACKING TASK

	Time-to-First-Human-Error		Time-to-First-Human-Error-Correction	
Pitch Axis	Distribution	Mean (Seconds)	Distribution	Mean (Seconds)
Trial #1	Weibull	14.6	Exponential	3.1
Trial #2	Weibull	100.6	Log Normal	2.3
Roll Axis				
Trial #1	Weibull	23.4	Log Normal	1.8
Trial #2	Weibull	214.9	Log Normal	0.9

* All distributions listed are significant at .20 level using Kolmogorov-Smirnoff test.

TABLE 3

DISTRIBUTIONS GOVERNING MEAN-TIME-BETWEEN-HUMAN-ERRORS AND MEAN-TIME-TO-HUMAN-ERRORS-CORRECTION FOR 2-AXIS MANUAL CONTROL TRACKING TASK

	Time-Between-Human-Errors		Time-To-Human-Errors-Correction	
Pitch Axis	Distribution	Mean (Seconds)	Distribution	Mean (Seconds)
Trial #1	Log Normal	19.3	Log Normal	2.2
Trial #2	Log Normal	34.8	Log Normal	1.5
Roll Axis				
Trial #1	Log Normal	21.3	Log Normal	1.7
Trial #2	Log Normal	55.0	Log Normal	1.0

* All distributions listed are significant at .20 level using Kolmogorov-Smirnoff test.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D	
(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)	
1. ORIGINATING ACTIVITY (Corporate author) Advanced Systems Division Air Force Human Resources Laboratory Wright-Patterson AFB, Ohio 45433	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP N/A
3. REPORT TITLE Quantifying Human Performance Reliability	
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Technical Report	
5. AUTHOR(S) (First name, middle initial, last name) WILLIAM B. ASKREN THADDEUS L. REGULINSKI	
6. REPORT DATE June 1971	7a. TOTAL NO. OF PAGES: 13 7b. NO. OF REFS: 4
8a. CONTRACT OR GRANT NO. b. PROJECT NO. AFHRL 1124 c. TASK NO. AFHRL 1124-01 d. PROJECT NO. AFIT 12-69	9a. ORIGINATOR'S REPORT NUMBER(S) AFHRL-TR-71-22 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) N/A
10. DISTRIBUTION STATEMENT Approved for public release; distribution unlimited.	
11. SUPPLEMENTARY NOTES Presented at the Psychology in the Air Force Symposium conducted at the Air Force Academy, Colorado, 20-22 April 1971.	12. SPONSORING MILITARY ACTIVITY Air Force Human Resources Laboratory Air Force Systems Command Wright-Patterson AFB, Ohio 45433

13. ABSTRACT

Human performance reliability for tasks in the time-space continuous domain is defined and a general mathematical model presented. The human performance measurement terms time-to-error and time-to-error-correction are defined. The model and measurement terms are tested using laboratory vigilance and manual control tasks. Error and error-correction data are ordered and the underlying density functions isolated. The Weibull distribution is best fit for time-to-first-error data, and the Log-Normal distribution is best fit for time-between-errors and time-to-error-correction data. The Normal distribution is rejected in all cases. Distribution parameter values are applied to the general mathematical model, and prediction made of human performance reliability for the tasks. It is also shown that task performance reliability improves with training on the tasks.

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Human Performance Reliability						
Human Performance Quantification						
Man-Machine Systems						
Human Frrors						
Training Effects on Performance						
Human Factors						