IEEE Std 845-1999
(Revision of
IEEE Std 845-1988)

IEEE Guide for the Evaluation of Human-System Performance in Nuclear Power Generating Stations

Sponsor

Nuclear Power Engineering Committee
of the
IEEE Power Engineering Society

Approved 26 June 1999

IEEE-SA Standards Board

Abstract: Guidance for evaluating human-system performance related to systems, equipment, and facilities in nuclear power generating stations is provided. Specific evaluation techniques and rationale for their application within the integrated systems approach to plant design, operations, and maintenance described in IEEE Std 1023-1988 are summarized.
Keywords: design, human-system performance, integrated systems, maintenance, nuclear power generating stations, and operation

---

The Institute of Electrical and Electronics Engineers, Inc.
3 Park Avenue, New York, NY 10016-5997, USA.

Print: ISBN 0-7381-1745-5 SH94756
PDF: ISBN 0-7381-1746-3 SS94756

IEEE Standards documents are developed within the IEEE Societies and the Standards Coordinating Committees of the IEEE Standards Association (IEEE-SA) Standards Board. Members of the committees serve voluntarily and without compensation. They are not necessarily members of the Institute. The standards developed within IEEE represent a consensus of the broad expertise on the subject within the Institute as well as those activities outside of IEEE that have expressed an interest in participating in the development of the standard.

Use of an IEEE Standard is wholly voluntary. The existence of an IEEE Standard does not imply that there are no other ways to produce, test, measure, purchase, market, or provide other goods and services related to the scope of the IEEE Standard. Furthermore, the viewpoint expressed at the time a standard is approved and issued is subject to change brought about through developments in the state of the art and comments received from users of the standard. Every IEEE Standard is subjected to review at least every five years for revision or reaffirmation. When a document is more than five years old and has not been reaffirmed, it is reasonable to conclude that its contents, although still of some value, do not wholly reflect the present state of the art. Users are cautioned to check to determine that they have the latest edition of any IEEE Standard.

Comments for revision of IEEE Standards are welcome from any interested party, regardless of membership affiliation with IEEE. Suggestions for changes in documents should be in the form of a proposed change of text, together with appropriate supporting comments.

Interpretations: Occasionally questions may arise regarding the meaning of portions of standards as they relate to specific applications. When the need for interpretations is brought to the attention of IEEE, the Institute will initiate action to prepare appropriate responses. Since IEEE Standards represent a consensus of all concerned interests, it is important to ensure that any interpretation has also received the concurrence of a balance of interests. For this reason, IEEE and the members of its societies and Standards Coordinating Committees are not able to provide an instant response to interpretation requests except in those cases where the matter has previously received formal consideration.

Comments on standards and requests for interpretations should be addressed to:

Secretary, IEEE-SA Standards Board
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08855-1331
USA

> Note: Attention is called to the possibility that implementation of this standard may require use of subject matter covered by patent rights. By publication of this standard, no position is taken with respect to the existence or validity of any patent rights in connection therewith. The IEEE shall not be responsible for identifying patents for which a license may be required by an IEEE standard or for conducting inquiries into the legal validity or scope of those patents that are brought to its attention.

Introduction

(This introduction is not part of IEEE Std 845-1999, IEEE Guide for the Evaluation of Human-System Performance in Nuclear Power Generating Stations.)

This introduction provides background on the rationale used to develop this guide. This information is meant to assist in the understanding and usage of this guide.

Human factors engineering has been a part of nuclear power plant design, construction, and operation from the industry's beginning, although not under that name. (For example, see H. L. Parris, "A Review of Human Factors R&D in the Nuclear Power Industry," and Leonard C. Pugh, Sr., "In the Rear-View Mirror: Human Factors in Nuclear Power," both in *American Nuclear Society Transactions*, 1986 Annual Meeting, Volume 52, TANSAO 52 1-658, 1986.) The incident at Three Mile Island, however, brought increased prominence to the human factors field in the nuclear power industry. The U.S. Nuclear Regulatory Commission began to consider guidance and regulation concerning the human factors engineering aspects of nuclear power plant design, construction, and operation at that time. Part of the industry's response to this increasing emphasis on human factors engineering were proposals to create human factors guidelines as part of the industry standards efforts of the IEEE Nuclear Power Engineering Committee. This guide began as one of those efforts, initiated in 1980. These efforts were the subject of an IEEE-sponsored meeting held in Monterey, California, in June 1985, as documented in "1985 IEEE Third Conference on Human Factors and Power Plant," report #85CH2235-0.

This guide, IEEE Std 845-1999, was initially published in 1988. Subcommittee SC-7, Human Factors and Control Facilities, prepared this revision. The revision is intended to both update the content, and focus the purpose of the guide from the 1988 publication. Specifically, Working Group 7.4 of SC-7 has prepared this revision to make the principles and techniques contained in it more consistent with the integrated systems approach to engineering design as described in the top-level companion standard, IEEE Std 1023-1988 [B3][1]. This revision is also intended to improve the consideration of related factors to systems design, such as procedures, training, human-systems interactions, etc., in the evaluation of overall systems performance. SC-7 is also in the process of concurrently revising IEEE Std 1023-1988 [B3], which will reflect the improvements made to this standard.

This guide is intended to supply basic guidance for evaluating human-system performance related to systems, equipment, and facilities in nuclear power plants. It summarizes specific evaluation techniques and presents rationale for their application within the integrated approach to plant design, operations, and maintenance as described in IEEE Std 1023-1988 [B3]. This revision is intended for use by personnel who are familiar with the concepts of formal human factors analysis, but not necessarily human factors practitioners knowledgeable in the details of specific evaluation techniques.

In addition to preparing IEEE Std 845-1999 and revising IEEE Std 1023-1988 [B3], SC-7 has also developed the folllowing human factors-related documents: IEEE Std 1082-1997 [B4] and IEEE Std 1289-1998 [B5].

[1]The numbers in brackets correspond to those of the bibliography in Annex A.

Participants

During the initial revision of this guide, Raymond J. Christensen was chair of Working Group 7.4. At the time this guide was completed, Working Group 7.4 was chaired by James P. Bongarra, Jr. The following list represents the membership of SC-7 that attended the meetings and contributed to this document:

James P. Bongarra, Jr., *Chair*
Julius J. Persensky, *Standard Champion*

Valerie E. Barnes
Michael A. Boggi
Raymond J. Christensen
Hamilton C. Fish, Jr.
Robert B. Fuld
Bruce P. Hallbert
Scott Malcolm
John M. O'Hara
Robert Sheranko
Robert M. Waters

The following members of the balloting committee voted on this standard:

Satish K. Aggarwal
Vincent P. Bacanskas
Leo Beltracchi
James P. Bongarra, Jr.
Wes W. Bowers
Robert C. Carruth
John P. Carter
Raymond J. Christensen
Richard E. Dulski
Surinder Dureja
Leo G. Egan
Jay Forster
Joseph Fragola
Lawrence P. Gradin
John Kenneth Greene
Britton P. Grim
Robert E. Hall
Gregory K. Henry
David A. Horvath
James T. Keiper
John R. Matras
Richard B. Miller
Burt Nemroff
Ira M. Rubin
D. E. Sandwick
Marvin W. Sheets
Peter B. Stevens
John H. Taylor
James E. Thomas
John T. Ullo
John B. Waclo
Carl Weary
John L. White
G. O. Wilkinson
Paul L. Yanosy
David J. Zaprazny
Mark S. Zar
James B. Zgliczynski

When the IEEE-SA Standards Board approved this standard on 26 June 1999, it had the following membership:

Richard J. Holleman, *Chair* **Donald N. Heirman,** *Vice Chair*
Judith Gorman, *Secretary*

Satish K. Aggarwal
Dennis Bodson
Mark D. Bowman
James T. Carlo
Gary R. Engmann
Harold E. Epstein
Jay Forster*
Ruben D. Garzon
James H. Gurney
Lowell G. Johnson
Robert J. Kennelly
E. G. "Al" Kiener
Joseph L. Koepfinger*
L. Bruce McClung
Daleep C. Mohla
Robert F. Munzner
Louis-François Pau
Ronald C. Petersen
Gerald H. Peterson
John B. Posey
Gary S. Robinson
Akio Tojo
Hans E. Weinrich
Donald W. Zipse

*Member Emeritus

Also included is the following nonvoting IEEE-SA Standards Board liaison:

Robert E. Hebner

Greg Kohn
IEEE Standards Project Editor

Contents

IEEE Guide for the Evaluation of Human-System Performance in Nuclear Power Generating Stations

1. Overview

This guide is divided into three clauses. Clause 1 provides the scope and purpose of this guide. Clause 2 provides definitions that are either not found in other standards or have been changed for use with this guide. Clause 3 describes various human performance evaluation techniques that may be used to support the integrated systems approach to design, operation, and maintenance of nuclear power generating stations, as described in IEEE Std 1023-1988 [B3].[1]

1.1 Scope

This document provides guidance for evaluating human-system performance related to systems, equipment, and facilities in nuclear power generating stations. It summarizes specific evaluation techniques and presents rationale for their application within the integrated systems approach to plant design, operations, and maintenance described in IEEE Std 1023-1988 [B3].

1.2 Purpose

This document provides guidance for the selection and application of human performance evaluation techniques. It summarizes various techniques and presents recommendations for their application. Tasks to which they may be applied include the following:

a) Investigating human-system performance characteristics;
b) Testing design or operating approaches for adequacy;
c) Comparing alternative designs or configurations; or
d) Evaluating the maintainability of the system.

This guide is for use by personnel who are familiar with the concepts of formal human factors analysis, but not necessarily familiar with the details of specific techniques.

[1]The numbers in brackets correspond to those of the bibliography in Annex A.

2. Definitions

For the purposes of this guide, the following terms and definitions apply. IEEE Std 100-1996 [B2] should be referenced for terms not defined in this clause.

2.1 human-system interface (HSI): The interaction between workers and their equipment. This interaction requires information to flow in two directions. The system provides status information to the user, and the user provides control information to the system. [Used in other texts as man-machine interface (MMI), human-machine interface (HMI), human-machine system (HMS), and human-computer interface (HCI).] (For further information see IEEE Std 1289-1998 [B5].)

2.2 system development cycle: The life cycle through which a system is developed, which consists of the following:

a) Concept development;

b) Design;

c) Test and construction;

d) Operation; and

e) Maintenance (see IEEE Std 1023-1988 [B3]).

3. Evaluating human-system performance

3.1 General

To evaluate human-system performance, the evaluator needs to recognize that human performance is integral to system performance throughout design, development, testing, operation, and maintenance activities. Therefore, human performance is an integral part of system performance evaluation. Human performance is influenced by many factors. For example, environmental conditions; organizational design; training; and physiological, perceptual, and cognitive processes all influence human performance. The evaluator can apply various measures and evaluation techniques to formally evaluate the performance of people on tasks of interest.

There are several considerations that are associated with evaluating human-system performance. These include the following:

a) Selection and implementation of the measure and technique;

b) Analysis and interpretation of human performance data;

c) Measuring cognitive processes;

d) Generalizing from experimental studies; and

e) Establishing meaningful performance criteria for some tasks.

A comprehensive approach to human-system performance evaluation will require attention to these considerations. This guide includes brief discussions of selected considerations as they relate to the performance evaluation techniques recommended here. For a more detailed discussion of potential considerations, see ANSI/AIAA G-035-1992 [B1].

This guide describes human-system performance evaluation techniques that may be used to support the systems design approach described in IEEE Std 1023-1988 [B3]. These evaluation techniques include paper and pencil, observational, expert judgment, and experimental techniques. Human factors design analyses techniques (e.g., mission, function, task, and link analyses) are not included in this guide, but are described in IEEE Std 1023-1988 [B3]. Human factors expertise is desirable when selecting and applying appropriate human-system performance evaluation techniques to avoid the use of inefficient or inappropriate techniques. Descriptions of the type of data obtained from each technique, cost considerations, and other useful decision criteria are included to guide the user in incorporating human performance evaluation in system design.

3.2 Evaluation concepts

Human-system performance evaluation requires the evaluator to select appropriate measurement techniques, collect the data, and analyze and interpret the results. The selection of appropriate measurement techniques depends on the purpose of the overall evaluation and other practical constraints. Within these limits, different techniques exist that will be more or less suited to particular situations. This guide contains information for the selection and application of human-system performance evaluation techniques.

To interpret results, the evaluator should specify criteria for judging the acceptability of human-system performance. Without some form of acceptance criteria, the evaluator has performed only measurement, not evaluation. These criteria may be informal (evaluator's opinion regarding the acceptability of the performance) or formal (establishing specific criteria related to the measurement; for example, operator diagnosis within a specific time limit). Other performance criteria may also be appropriate in the overall evaluation (e.g., maintaining an adequate margin to a safety function in the operation of the plant). Ultimately, the selection of appropriate criteria should be established in the context of the overall system development process, including project goals and constraints (see IEEE Std 1023-1988 [B3]).

3.3 Characteristics of human-system performance measures

The characteristics described in 3.3.1 should be considered when selecting evaluation measures that best reflect meaningful and measurable aspects of human performance. This is not a complete list, but represents the major characteristics that should be considered when selecting techniques. For a detailed discussion of characteristics, see ANSI/AIAA G-035-1992 [B1]. The application of specific evaluation techniques is described in 3.4.

3.3.1 Characteristics associated with selecting human-system performance measures

— *Acceptability:* The degree to which evaluators at all levels agree on the use of the measure.

— *Accuracy:* Minimization of measurement error.

— *Applicability:* Not all techniques are applicable to all phases of system design and operation. Therefore, it is important in any evaluation of human-system performance to define the applicability of the specific measures. Consider the applicability of each measure with respect to its use during design and evaluation of the system.

— *Bias:* The degree to which measured results are free from systematic sources of prejudice or error.

— *Intrusiveness:* The extent to which the measure alters or interferes with the performance being measured. Avoid measures that influence the worker's performance or disrupt the activity.

— *Precision:* The level of detail of the instrument, sensor, or instrumentation.

— *Reliability:* The degree to which the measure yields consistent and reproducible findings when used in comparable circumstances.

— *Resources:* The items needed to implement the measure, such as time, budget, personnel, equipment, logistics, and the need for specialized expertise.

— *Sensitivity:* The degree to which the measure is able to discriminate meaningful variations on the dimension of interest.

— *Validity:* The degree to which an instrument or technique can be demonstrated to measure what it is intended to measure.

3.3.2 Characteristics of subjective vs. objective measures

Often, objective measures are perceived as being more meaningful than subjective measures. This is not necessarily the case.

Subjective measures yield data that are obtained from the judgments and opinions of users or experts (e.g., judgment of task difficulty).

Such measures, while vulnerable to individual bias and perspective, are typically the most practical measures available for complex or inferred behaviors (e.g., problem solving). The sources of bias in subjective measures can be controlled in most situations by attention to the characteristics described in 3.3.1.

Subjective measures, such as asking opinions on design features, may produce qualitative data. Subjective measures, such as the use of rating scales where numbers indicate the degree of response, may also produce quantitative data. Statistical techniques are available for analyzing data derived from both subjective and objective measures. Use of statistical techniques increases the formality and objectivity of performance evaluation.

Objective measures yield data that are obtained from observable human behavior (e.g., time to complete a task).

Objective measures are not as vulnerable to bias introduced by individual opinions as subjective measures. Objective measures yield data such as reaction times, error rates, display types accessed, and the number of specific control actions.

Though data from objective measures are difficult to collect from unobservable human behavior such as cognitive processes, they can help evaluators interpret complex human behaviors. When applied to observable behavior, data from objective measures can be used to draw inferences about differences between experimental groups, and between samples and the populations from which they are drawn.

3.3.3 Characteristics of using diverse measures

Given the nature of an evaluation, a single measure may not yield data to assure sufficiently valid results. In such cases, multiple (i.e., diverse) measures of the same performance should be considered. Agreement among the data obtained from multiple measures may strengthen the findings. Note that the expected benefit of adding measure(s) to an evaluation must justify the costs of their data collection and processing. However, the incremental costs of added measures may be relatively small if simple methods are used.

3.4 Evaluation techniques

Many human-system performance evaluation techniques apply to the nuclear power plant setting. This subclause describes some of the most commonly used techniques. An assessment of each technique is made relative to the attributes described in 3.3. For a detailed technical discussion of each technique, refer to the documents in Annex A.

The human-system performance evaluation techniques are divided into the following four descriptive categories, each based on the method of application:

a) Paper and pencil;
b) Observational;
c) Expert judgment; and
d) Experimental.

3.4.1 Paper and pencil techniques

The evaluation techniques in this category are different in that the observation of actual system/personnel performance is not required. The output from these techniques can be a simple accept-or-reject decision, a hierarchical ranking, or a numerical representation of merit.

3.4.1.1 Checklists

— *Description:* A checklist is a numbered set of statements presenting attributes that a piece of equipment or a system should possess to meet acceptable human factors criteria. A checklist is used to partition a complex entity into more readily understood elements. Such human engineering elements typically concern whether expected information or control functions are available, whether their design characteristics are suitable, or whether a specific performance result is obtained. If checklist items are too general, then the evaluator may make discretionary decisions, and thus the result may be less reliable. If checklist items are too specific, their application requires more effort. Evaluator discretion will determine which checklist items apply to different components of the system or equipment. Checklist items should be simple and clearly specified. Bases for the criteria should be established and retained.

— *Requirements:* A checklist is useful in at least four different ways, as follows:

1) As a memory aid, reminding the evaluator to assess a particular characteristic;
2) As a reference standard, providing criteria against which the design is verified;
3) As a documentary record of the verification process and results; or
4) As a guideline to designers.

— *Applicability:* Checklists are broadly applicable, depending on when and to what degree criteria can be confidently specified. In particular, checklists are used prior to hardware or software selection, after mock-up or prototyping of a design, and after final design as part of test and evaluation.

— *Selection criteria:* Checklists do not require special equipment or techniques for implementation, and clearly specified checklist items can reduce the amount of evaluator training needed to perform the evaluation. A main advantage of checklists is that they are relatively easy to apply, and may be useful beyond evaluation in other areas of the design process. Interevaluator variability can be reduced by a well-designed checklist. Disadvantages include the challenges of determining satisfactory underlying criteria; writing checklist items; and formatting, producing, and maintaining the checklist current with the design specifications.

3.4.1.2 Questionnaires and interviews

— *Description*: Survey techniques, such as questionnaires and interviews, can be used to assess potential users' needs, knowledge, attitudes, opinions, beliefs, perceptions, expectations, and reactions, but not system performance. Questionnaires consist of a set of written questions to which the user responds in writing. The questions may be open-ended, allowing respondents to provide as much or as little information as they choose, or constrained by the use of multiple-choice formats and rating scales.

Interviews involve face-to-face or telephone contact between evaluators and respondents. Interview techniques vary in terms of structure of the questions asked and responses required. Interviews can be conducted using a standardized set of questions with only constrained responses allowed. The interviewer may use a discussion guide that defines the general topics covered, but which does not structure how specific questions are asked or how the questions are answered.

— *Requirements*: The questions asked and answers required must be simple and easy to understand, particularly for questionnaires where respondents may be unable to ask for clarification. Questionnaire and interview items and techniques should be pretested on a sample of representative respondents. Established questionnaires and interview techniques can be used, rather than developing new questionnaires. Questions should be worded to avoid response bias. Multiple questions on the same topic can be used to test the reliability of the answers obtained. A reviewer should be familiar with interviewing techniques and data evaluation. Respondent anonymity may be an issue if sensitive issues are being tested.

— *Applicability:* Questionnaires and interviews are particularly useful during the concept development phase and in assessing the acceptability of system characteristics.

— *Selection criteria*: The primary advantages of survey techniques are as follows:

a) They are easy and relatively inexpensive to use;

b) They allow the evaluator to ask many more questions than other techniques;

c) They have face validity; and

d) They can provide a simple means of evaluating small changes in the potential users' reactions to system characteristics.

The primary disadvantages of survey techniques are that they are limited by an individual's ability to verbalize experiences or opinions (rather than system performance), and are subject to a variety of factors that can distort responses (see 3.3.2).

3.4.1.3 Operational experience review (OER)

— *Description*: OER is the examination and evaluation of specific in-house and industry operating experience related to system and human performance for systems similar to the system being reviewed. The technique entails the review of archival data sources, such as licensee event reports (LERs); significant event reports (SERs), significant operating experience reports (SOERs), plant corrective action systems, operational and maintenance logs and records, and data from interviews with experienced plant personnel.

— *Requirements:* This technique requires access to data, personnel involved with the event, experienced systems personnel, and a technique to evaluate the applicability of the data. The reviewer should be familiar with industry- and plant-specific databases, interviewing techniques, and data evaluation.

— *Applicability:* OER data used during early design work can identify potential problems as well as others who have experience with similar issues. The process is often used to evaluate (and predict) human and system performance during operation, both normal and off-normal.

— *Selection criteria:* While this technique is straightforward, it can be time consuming and tedious due to the number of data sources that can be evaluated for relevance. Historical data is secondary data and may not be valid, especially for new systems. Validity is dependent on the data available for review and its similarity to the historical system or human-system applications; the proposed (new) application under study; and the skill of the reviewer in collecting, understanding, and using the data.

3.4.2 Observational techniques

Observational techniques are distinguished from other evaluation procedures by the requirement that the evaluator(s) observe operator performance on the system or a representation of the system (testbed) to be evaluated. See 3.5 for a discussion of testbeds.

3.4.2.1 Walk-through/talk-through

— *Description:* The walk-through/talk-through technique is the most widely used observational technique. Potential users of the system being evaluated walk through one or more of the tasks required to successfully operate the new system. While walking through the tasks, users are prompted by the evaluator to describe their activities. The walk-through/talk-through can be conducted without pre-analysis, but it is more effective when preceded by a task analysis that identifies the behaviors and responses that are expected to occur during the walk-through.

— *Requirements:* The technique can be used to evaluate any operable representation of equipment, including mock-ups, training simulators, or the actual plant. An operator walk-through/talk-through can be conducted as a desktop analysis using operating procedures. Its use is not limited to any particular equipment type or configuration.

— *Applicability*: The walk-through/talk-through technique is applicable at any point in the design process when an operable representation of the necessary equipment is available. Walk-throughs/talk-throughs are typically conducted during the design phase, as soon as a prototype system (such as a mock-up) is available. Walk-throughs/talk-throughs can be used to assess whether equipment design and job procedures properly support human performance. This technique can also be effective for evaluating operating procedures on fully operational systems.

— *Selection criteria:* The walk-through/talk-through technique is practical and useful. The cost can vary, and depends on the extent of the evaluation and the nature of the facilities used. A valid walk-through/talk-through requires similarity between the evaluation and actual operations. The qualitative results of walk-throughs can be very informative, but difficult to formalize.

3.4.2.2 Verbal protocols

— *Description:* Verbal protocols are obtained by having the users/experts speak their thoughts aloud about what they are doing while performing selected tasks. This type of verbalization reflects information contained in working memory, and reflects ongoing cognitive processes employed in task implementation. Verbal protocols may be recorded to form a complete record of a user's dialogue. The tasks can be conducted as a walk-through (i.e., in non-real time) or in real time as part of experimentally controlled evaluations in a high-fidelity mock-up (see 3.4.4). The data may be treated informally as a self-explanatory record, or can be subject to highly formalized analysis.

— *Requirements:* The primary requirements for conducting a verbal protocol evaluation are the system being evaluated or a prototype of the system, a user, a framework for formal analysis, and a means of recording the user's verbalizations.

— *Applicability:* Verbal protocols can be conducted during the design phase, as soon as a prototype system (such as a mock-up) is available. Verbal protocols can be used to help ensure that the equipment design and job procedures properly support the user's cognitive model and subsequent human performance.

— *Selection criteria:* Verbal protocol data are moderately easy to obtain, and the technique can be used in conjunction with several others described in 3.4.4. The cost depends on the scope of the mock-up used and the complexity of the scenarios being evaluated. Using a dynamic full-scope simulator for a predicted plant accident transient is more costly than using part-scope simulation. The validity of the technique is dependent on the similarity between the actual operating constraints and the test setting. The qualitative results are of moderate precision and are sensitive to issues such as compatibility and understandability. The analysis of the verbal protocol data is time consuming, and a predetermined model against which to compare the user's behavior is helpful. The verbal protocol technique assumes that cognitive processes occur slowly enough to be verbalized. In some situations, verbal protocols may interfere with the primary task.

3.4.3 Expert judgment

Expert judgment relies on opinions from individuals with demonstrated competence in a particular content area. It is used when evaluating a specific technical problem. Experts are selected based on their training and experience relative to the problem of concern. An expert is a person who

a) Has background in the subject area; and
b) Is recognized by his peers or those conducting the study as qualified to address the problem.

Expert judgment, as applied in the nuclear industry, ranges from informal consensus techniques to more formal psychometric techniques (e.g., quantification of an expert's judgment using scales) that impose more structure and precision on the expert judgment process. Regardless of the expert solicitation technique chosen, a trained evaluator is needed to develop, coordinate, administer, and evaluate the results of the process.

— *Description:* The Delphi method and Nominal Group method are examples of consensus techniques based on group judgments. Group judgments are generally preferred over individual judgments because the results tend to be more accurate.

— *Requirements:* The Delphi method requires that the participants complete a survey or questionnaire in which they provide both their opinions and the basis for that opinion. An independent evaluator collects and collates the responses, then sends out a second round of surveys or questionnaires that describe the opinions and comments received in the first round. This process is repeated until a consensus is reached. The Nominal Group method is similar to the Delphi method in that judgments from multiple experts are solicited, evaluated, and clarified. The principal difference is that the Nominal Group method strives for a consensus of judgments in face-to-face meetings.

  The other group of formal methods of elicitation rely on scaling methods to enable quantitative manipulation of the variables. For example, two of the more popular scaling methods are scaling by paired comparisons and scaling by ratio estimation. These methods are also used in human reliability analyses (HRA). These and other methods for quantifying human reliability are discussed in IEEE Std 1082-1997 [B4] and are not discussed further here.

— *Applicability:* Expert judgment can be used at any point during the design or evaluation of a system, even when no equipment actually exists. Typical uses of expert judgment include the following:

  a) Providing estimates on complex, rare, or poorly understood phenomena;

  b) Forecasting future events;

  c) Learning about problem-solving or decision-making techniques and processes; and

  d) Determining the state-of-the-art/practice in a field of knowledge.

  In nuclear power plants, expert judgment plays an important role in estimating human-error probabilities in probabilistic risk assessments.

— *Selection criteria*: Some advantages of expert judgment are that experts are usually available when other data sources are not, and the cost associated with knowledge elicitation in terms of time and money is relatively small. The disadvantages are that experts do not always agree, and that levels of expertise may vary greatly.

3.4.4 Experimental techniques

— *Description:* Experimental techniques are used to answer specific questions (i.e., hypotheses) about human-system interactions. An experiment differs from other evaluation techniques in that

  a) The evaluator creates the conditions necessary for observing performance;

  b) The evaluator controls those conditions so as to vary only the crucial aspects of the situation of experimental interest; and

  c) The evaluator controls which individuals will be exposed to the conditions.

  Controlled conditions increase the choice of potentially useful performance measures. A more detailed description of these measures is available in ANSI/AIAA G-035-1992 [B1]. After results are obtained, statistical tests are used to determine whether any observed differences in performance under the different experimental conditions are significant. Personnel serving as test subjects can also be debriefed to further explore underlying processes supporting observed behavior.

— *Requirements:* Because of the complexities associated with the design, implementation, and analysis of experimental data, experimentation should only be undertaken by someone specifically trained in these techniques. The evaluator selects appropriate performance measures, determines the number of test subjects needed, and controls the testing conditions. Experimental control is designed to minimize extraneous sources of performance variation. In setting up an experiment, the evaluator needs to translate the specific questions into a set of test conditions that are observable, repeatable, and can be compared.

— *Applicability:* Experimental techniques are useful for assessing human-system interfaces. They are also useful for equipment selection and evaluation.

— *Selection criteria:* The primary advantage of experimental techniques is that they provide more objective results about system performance. However, they can only be used to assess a limited number of variables at a few discrete levels. The need for tight controls makes their use expensive and time consuming.

3.5 Testbeds

The human-system performance evaluation techniques described in 3.4 may be implemented using some type of testbed. Testbeds are any materials or equipment used to represent the system design under study. Testbeds can range from simple paper-based line drawings of system design features or human-system interfaces (HSI), through rapid prototypes of visual display unit (VDU) displays, to high-fidelity, real-time simulators of the actual system HSIs and their associated behavior. As the fidelity of the testbed increases, the confidence in the findings increases.

The evaluator must select an appropriate testbed as part of selecting measurement techniques. Early in design, high-fidelity testbeds are typically not available, leaving the evaluator to use lower fidelity, simpler testbeds. However, most nuclear power plant designs are evolutions of existing, inservice designs that are supported by full-scope simulators and trained operating staff. This allows the evaluator to perform evaluations that have the benefit of high-fidelity simulation. The range of common, practically useful testbeds include the following:

a) Line drawings of system features or HSIs;
b) Illustrations or photographs of HSIs;
c) Static mock-up of HSIs (constructed with inexpensive materials such as cardboard, fiberboard, or wood; full-scale preferred);
d) Rapid prototypes of VDU displays (these can be incorporated into a static mock-up of HSIs, such as panels, to increase overall fidelity);
e) Full-scale dynamic simulator of HSIs (e.g., control room simulator); or
f) Real plant (as-built facility; measurement techniques could be incorporated into commissioning tests).

Annex A

(informative)

Bibliography

[B1] ANSI/AIAA G-035-1992, Guide to Human Performance Measurement.

[B2] IEEE Std 100-1996, IEEE Standard Dictionary of Electrical and Electronic Terms, Sixth Edition.

[B3] IEEE Std 1023-1988, IEEE Guide for the Application of Human Factors Engineering to Systems, Equipment, and Facilities of Nuclear Power Generating Stations.

[B4] IEEE Std 1082-1997, IEEE Guide for Incorporating Human Action Reliability Analysis for Nuclear Power Generating Stations.

[B5] IEEE Std 1289-1998, IEEE Guide for the Application of Human Factors Engineering in the Design of Computer-Based Monitoring and Control Devices for Nuclear Power Generating Stations.

