CONF-8506100-3



# COLLECTION AND ANALYSIS OF TRAINING SIMULATOR DATA*

CONF-8506100--3

TI85 015043

P. A. Krois
P. M. Haas

Oak Ridge National Laboratory
Oak Ridge, Tennessee

Presentation Made At:

1985 IEEE Third Conference on
Human Factors and Power Plants
Monterey, California
June 23-27, 1985



*Research sponsored by the U.S. Nuclear Regulatory Commission under DOE Interagency Agreement 40-550-75 with Martin Marietta Energy Systems, Inc. under Contract No. DE-AC05-84OR21400 with the U. S. Department of Energy.

**ABSTRACT**

The purposes of this paper are: (1) to review the objectives, approach, and results of a series of research experiments performed on nuclear power plant training simulators in support of regulatory and research programs of the U. S. Nuclear Regulatory Commission (NRC), and (2) to identify general research issues that may lead to an improved research methodology using the training simulator as a field setting. Research products consist of a refined field research methodology, a data store on operator performance, and specific results pertinent to NRC regulatory positions. Issues and potential advances in operator performance measurement are discussed.

## INTRODUCTION

The purposes of this paper are: (1) to review the objectives, approach, and results of a series of research experiments performed on nuclear power plant (NPP) training simulators in support of regulatory and research programs of the U.S. Nuclear Regulatory Commission (NRC), and (2) to identify general research issues that may lead to an improved research methodology using the training simulator as a field setting. The experiments and related studies collected data on operator performance from both full-scope NPP control room simulators and actual operating experiences obtained from plant records. Th NRC programs address such human factors issues as operator qualifications, licensing, training, human reliability, procedures, and the man-machine interface.

In the context of this paper, two phases of these training simulator experiments are distinguished. Part of the difference between the two phases corresponds to the high level research objectives for which the experiments were designed and conducted, and part of the difference parallels the modifications and refinements to the experimental methodology over time. The first phase was concerned with the development of criteria for safety-related operator actions (SROA). The second and current phase of training simulator experiments has multiple objectives pertinent to NRC regulatory positions.

The objectives of the SROA experiments were to collect and assess operator performance data in support of the ANSI-N660 Standard (Ref. 1). These data support criteria for designers and regulators in the allocation of safety actions to manual operator control or automatic action. Part of the operator performance criterion in the N660 Standard involves time requirements for the activation of safety systems and considers the severity and frequency of event conditions. The rationale behind this standard was that rare, severe events will require more available time to maintain a given degree of operator reliability in that the operator will likely need that additional time in order to overcome the shock of the highly stressful event before initiating manual safety actions.

Time tests are used to determine if the time available is sufficient for the operator to take action. In order to determine the time available, the designer establishes the time necessary for completion of the safety function and subtracts out the equipment and process delay time required by the particular safety system. This determines the maximum permissible delay in activating the safety system. The designer also subtracts out the interval between initiation of the event and the activation of the first control room alarm. It is on the basis of the remaining time available that the designer determines the extent to which control of the safety system should be automated or allocated to the operator for manual control.

A preliminary study of available data supporting the N660 Standard found a paucity of data on operator response times from actual plant casualties (Ref. 2). The study suggested that operator performance criteria should heavily rely on operator data collected from training simulators with the assumption that sufficient data on actual operating experiences exist to provide a calibration of the simulator data. This preliminary study, then, provided the impetus for undertaking the collection of operator performance data over the course of several SROA experiments in boiling water reactor (BWR) and pressurized water reactor (PWR) control room simulators.

The initial PWR simulator exercises were undertaken to provide empirical response time data, to begin a characterization of operator skilled performance, including the effects of likely performance shaping factors (PSFs), and to gain experience and demonstrate the use of NPP training simulators to provide the necessary data base on operator performance (Ref. 3). Additional data were collected through initial BWR simulator exercises providing operator response times and error rate information (Ref. 4). These initial data were used in a simulator to field data calibration so that simulator results could be more confidently extrapolated to field conditions (Ref. 5). A second set of PWR simulator exercises was conducted to collect additional data on operator response times, error rates, and PSFs (Ref. 6). Other studies in the SROA program were an examination of current practices using simulators in operator training and requalification (Ref. 7), a study of practices, conclusions, and recommendations on the specification and verification of NPP training simulator response characteristics (Refs. 8 and 9), and development of a control room task analysis approach for PWRs and BWRs (Refs. 10 and 11). The final study in the SROA program was the development of a computer simulation methodology using simulator, field, and task analytic data to provide quantitative estimates of operator performance reliability of safety actions (Ref. 12).

The scope of the second phase of the training simulator experiments moved beyond the development of the SROA criteria. The objectives of these latest and current simulator experiments include the continued development of a research methodology using the training simulator as a field setting for data collection, the development of a data store on operator performance, and the preparation of specific results pertinent to NRC regulatory positions. A data base on operator performance during abnormal or emergency events was seen to potentially provide increased understanding of such operational safety issues as control room staffing requirements, the relationships of acquired experience and education level with operator performance, the effectiveness of symptom- vs. event-based procedures, and the effectiveness of operator aids during emergency events.

The primary objective of the first experiment in this second phase was to assess the effects of the experience level of the senior reactor operator (SRO) and the presence of a shift technical advisor (STA) on the performance of BWR control room operators (Ref. 13). Additional objectives were assessments of the subjective workload of operators during emergencies, of content-referenced performance measures, and of the effects of crew practice in working with the STA over the course of several emergency events. A second experiment examined the influences of SRO experience and presence of an STA on the performance of PWR control room operators (Ref. 14). Subjective measures of mental workload and experts' performance ratings were collected and the latter were compared with the content-referenced performance measures.

In general, over the course of these experiments and studies, results and experiences have led to several modifications to the manner in which measures of operator performance are defined, the types of operator performance data collected, and the design, administration, and assessment of simulator experiments as a research methodology. In the following sections, the current approach to the field research methodology is described including "quasi-controlled" experimental conditions and measures of operator performance. This is followed by a brief review of experimental results in terms of a refined research methodology and an empirical data base on operator performance. Research issues are then identified promising further improvements to the research methodology and also address such points as operator cognitive behavior and the relationship of operator performance to overall system performance.

## CURRENT FIELD RESEARCH METHODOLOGY

Through experience, the sophistication and rigor of the training simulator experiments as a field research methodology have steadily improved. The feasibility of the experiments was strongly influenced by the voluntary participation of utilities. Because of practical considerations of cost and the limited availability of simulator time, and the accessability of qualified operators to serve as research subjects, the utilities routinely allowed the experiments to be overlapped with regularly scheduled training simulator exercises. Utility participation necessitated minimum interference with the training program, which resulted in less experimental control compared to a laboratory experimental setting. Through discussions with training staff, sufficient controls were established to permit reproducibility of the experimental conditions. For cases in which complete control was not possible, key variables by which the conditions varied were identified and recorded for later analysis. These exercises, then, may be referred to as employing quasi-experimental designs (Ref. 15).

There are several points in the current methodology that contribute to a systematic assessment of research issues. In this section, the development of and constraints to the design of field experiments using training simulators are reviewed. This is followed with a discussion of a systems/task analysis that provide a framework for defining measures of operator performance and which subsequently determines the types of operator performance data that are collected. Data collection techniques are then described consisting of automatic and manual recordings of objective and subjective data, as well as videotaping of the simulator exercises. The collection of field data from plant records on actual operating experiences is also examined to support calibration of operator data collected from the simulator experiments.

### Experimental Designs for a Field Setting

Some of the practices involved with the experiments using the training simulator as a field setting include the design of the experiments, the assignment of operators as subjects into design categories, and the selection of operating sequences.

The designs of the experiments have become more sophisticated in relation to the research objectives. In the SROA experiments, control room crews were presented with a set of emergency events. For all experiments, operators were naive as to the events presented, were instructed only to respond as they would in an operating plant, and were asked to refrain from discussing the events with other operators. In the latest simulator experiments, hypotheses have been developed related to multiple research objectives leading to multi-factor experimental designs for testing those hypotheses. These latest experiments have used a three factor design, with two between-subjects factors: (1) the supervisor's months of experience as a supervisor, and (2) the assistance or absence of an STA, along with one within-subjects factor of scenarios (four emergency events). The high and low experience categories for the supervisor were relative to the qualifications of the SROs available to participate in the exercises.

The subjects for the experiments were licensed operators and STAs. A training group usually consisted of three to six operators, with equal or nearly equal numbers of SROs, reactor operators (ROs), and STAs. These operators were selected for assignment to training on the basis of the plant's operating schedule. The simulators were referenced to the operators' plants, although in one experiment operators were used coming from two PWR plants of similar design (Ref. 14). Operators from both plants had trained in the plant-referenced simulator for Plant A, which was operational, and Plant B was sufficiently near completion for the issuance of cold licenses to the operators. The engineering design of both plants is reportedly similar with very few variations in control panel layout between the two plants.

The sizes of the crews were influenced by the number of available SROs, ROs, and STAs. In the SROA experiment, crew size consisted of four or five operators. In the first training simulator experiment eight 2-operator (one SRO and one RO) and eight 3-operator (one SRO, one RO, and one STA) crews were used based on the number of available qualified operators. In the second training simulator experiment, twenty 3-operator (one SRO and two ROs) crews were used with the STA as an additional fourth crew member in half the crews.

The experimental groups were formed on the basis of the supervisors' experience, that is, their months of experience as an SRO. A pure test of the effects of supervisor experience has been limited by constraints imposed by the actual experience levels of available SROs, as well as the indeterminant influences on SRO performance from the SROs' experience as an RO, from the experience of SROs when used in RO positions because of an insufficient number of ROs, and from the experience levels of ROs. Operator experience data was supplied by the plant ahead of time so that the scheduling of operators for simulator exercises was compatible with assigning operators into the best possible categories in the experimental design.

The selection of operating sequences has been responsive to varying criteria. In the initial SROA experiments, the criteria included the following factors (Ref. 4):

1. Applicability to the N660 Standard.
2. Safety impact or consequences.
3. Generic to BWR plants.
4. Range of complexity of event diagnosis, of accident scenarios, and of required operator actions.
5 Ease of identifying measurable operator actions.
6. Adaptability to simulation.
7. Training value to utilities.
8. Sufficient frequency of occurrence in operating plants for field data collection.

In a subsequent SROA experiment, an additional requirement involved a request from the plant operations department that special evolutions be performed during the requalification training program. In the latest training simulator experiments, operating sequences were selected in consultation with the utility so as to represent a range of difficulty and complexity. As part of one experiment (Ref. 14), a severe accident precursor sequence consisting of a main steam line break (MSLB) with loss of the reactor water storage tank was specially developed by the utility's instructors to uniquely challenge the crew and invite increased participation by the STA. The order of presentation for most of the emergency events was systematically counterbalanced across crews, while in some experiments certain sequences such as MSLB were always presented last due to the time factor.

In sum, accrued experiences have shown reasonable successes in implementing experimental designs in association with plant training programs. However, constraints inherent to field research, notably the availability of operators with sufficiently divergent experiences as licensed operators and as SROs, lead to some limitations to the interpretation of research results. The utilities and operators have generally been very cooperative and supportive by voluntarily furnishing operators, training personnel, and simulator time.

**Systems/Task Analysis**

The recent simulator experiments have used a systems/task analysis developed as a system engineering approach to defining operator performance requirements. This analysis is altogether different from the approach used in the SROA experiments which was concerned primarily with response times and error rates. A fundamental problem with developing performance measures is distinguishing individual from crew performance. The purposes of the systems/task analysis, which is undertaken for each operating sequence, are to determine the tasks required, how they are performed, and which safety functions are affected. Five task performance measures are then derived on which to score individual operator performance for each task in the operating sequence. Functional performance measures are also used which indicate how well the crew as a whole controlled selected critical parameters associated with important safety functions.

An additional analysis consists of an assessment of the subjective workload experienced by the operators. This analysis uses the Subjective Workload Assessment Technique (SWAT) to measure mental workload (Ref. 16).

Briefly, and as noted above, the primary measures of operator performance used in the SROA experiments were response time and error rate. Response time was defined as the interval between the appearance of the first cue to the operators that something abnormal had happened and the first correct action they made in response to the malfunction. Critical task elements (CTEs) for each operating sequence were identified from plant procedures, and performance of a CTE by any operator was judged to indicate that at least an initial assessment of the situation had occurred. Errors of omission were examined based on switch manipulations called for in the procedures, as well as informational errors which were defined as errors associated with actions not having a direct impact on operations (e.g., test status light panels and select nuclear instruments).

The examination of response time in the SROA experiments considered the timing of the CTEs regardless of which operator performed the switch manipulation. Such a measure is representative only of the crew's overall performance. In the latest training simulator experiments, increased emphasis has been placed on better defining operator performance measures that distinguish individual performance apart from that of the crew. This emphasis comes from the fact that operators are licensed as individuals and not as a formal crew. As an additional point, in order to assess the effects from the SRO's experience as a supervisor on performance, it was necessary to measure the performances of the operators as individuals. Crew measures represent the collective performances of the SRO and ROs.

Operator performance measures were defined using a systems/task analysis. Components of this analysis for each operating sequence involved a task analysis of operator actions and a functional analysis for identifying important plant parameters. The result of each analysis was a set of performance measures as described below.

The task analysis procedure used was developed in the NRC crew task analysis project (Ref. 17). Inputs to the analysis were operating procedures, videotapes of a rehearsed real-time run in the simulator, and expert judgment. Task data forms were completed providing a detailed record of the tasks and task elements comprising each operating sequence. An additional in-depth operational analysis identified criteria for task success and documentable constraints on performance (e.g., required sequences of action) which were recorded on a task performance measures worksheet. Altogether, these task elements, success criteria, and constraints constitute a task-specific set of performance criteria and were listed on a task performance criteria form. This form also describes cues for task initiation and how task initiation may be observed.

It is noted that task elements involving the monitoring of system parameters, while a large part of the operator's job, could not be measured directly because an observer could not be certain that an operator had not checked a particular display. For this reason, some monitoring task elements were listed as "not observable." Other monitoring task elements were scored on the basis of whether the operator failed to comply with system limitations which could result from, among other factors, poor monitoring. From a systems point of view, exceeding a system limit is an error, regardless of how it happened.

Performance of each task in the operating sequence was described by five measures:

1 - Whether the task was initiated (scored as 1 or 0).
2 - The number of task elements performed correctly (counts).
3 - The number of preconditions or limits complied with (counts).
4 - Whether task success criteria were met (scored as 1 or 0).
5 - Time elapsed from the appearance of the cue to initiate the task until the task was completed (seconds).

To determine overall performance for an operating sequence, the performance scores for individual tasks were summed across all tasks and converted to percentages, e.g., the percentage of the number of tasks initiated to the number required for the sequence as a whole. The averaging of task times across tasks was accomplished by standardizing task times and using deviation scores to indicate relative performance.

The functional analysis consisted of identifying parameters in the emergency operating procedures associated with those critical safety functions that were challenged in the operating sequences. For example, in the current training simulator experiment at a PWR site (Ref. 14), three functions were identified by experts consisting of maintaining both the subcooling margin and pressurizer level for core cooling, and maintaining adequate steam generator levels for heat removal. From these three functions, six functional measures were derived:

6 - Time (seconds) subcooling margin less than 40 degrees.
7 - Root-mean-square (RMS) error in subcooling margin for the time subcooling margin less than 40 degrees.
8 - Time (seconds) pressurizer level less than 20%.
9 - RMS error (below the limit of the 20-80% operating band) for pressurizer level.
10 - Time (seconds) level in no one of the four steam generators greater than 79% (wide range).
11 - RMS error in steam generator for the time the highest steam generator level less than 79%.

These six functional measures were combined with the five task performance measures to provide an 11-variable metric of crew performance for each operating sequence. The task performance measures were scored on the individual operator level of analysis and then averaged together yielding crew scores. The functional measures were scored on the crew level of analysis to indicate how well operators worked together to control plant parameters.

An analysis was added in the latest training simulator experiments to determine the effect of assistance by an STA on the subjective mental workload of the SRO and to provide a meaningful index of scenario difficulty. The SWAT scale combines separate operator judgments of time pressure/ load, mental effort or concentration required, and perceived stress (Ref. 16). Separate ratings on these dimensions are calibrated on an individual basis and converted to a single estimate of overall workload. The scaled workload estimates are standardized to range from 0 to 100. This analysis allowed comparisons of subjective workload with measured performance across the operating sequences.

The systems/task analysis, then, consisted of several different analyses providing objective measures of operator and crew performance. A subjective measure of the operator's mental workload was also used. These measures would seem to have some degree of orthogonality and thereby provide a broad-based assessment of operator performance.

## Data Collection Techniques

Data collection techniques include an automated recording system, checklists completed by observers, videotapes of operators responding to the operating sequences, questionnaires completed by operators and STAs, and an instructor rating form. Most of these techniques support the scoring of operator performance on the measures described in the preceding section, and the last technique was developed to provide an assessment of an altogether different approach in the measurement of operator performance.

An automated recording system was a primary source of data on operator control actions and simulator systems status. This performance measurement system (PMS) is a computer software system consisting of an on-line data collection segment and a series of off-line data interpretation programs (Ref. 18). The on-line assembly language data collection program is concurrently executed with the plant simulation program in the simulator's computer. The recorded simulator data consist of four types of inputs and outputs:

1. Digital inputs - discrete inputs from the control room to the simulation program, e.g., repositioning of a two-position switch on one of the control panels.
2. Digital outputs - discrete outputs from the simulation program to the control room, e.g., signals changing illumination of annunciator tiles or indicator lamps.
3. Analog inputs - continuous inputs from the control room to the simulation program, e.g., repositioning of a control knob on one of the control panels.

4. Analog outputs - continuous outputs from the simulation program to the control room, e.g., value of a meter reading on one of the control panels

Special off-line scoring programs were written to provide scoring on the individual task performance measures and crew functional performance measures.

During the experimental sessions, operators assigned to the training group but who were not directly participating in the experiment acted as observers and completed a task performance identification checklist (TPIC). The TPIC was developed for manually recording the occurrence of task-critical communications and monitoring, which operator performed each task with the time when task elements occurred, actions taken outside the camera's field-of-view, and unusual occurrences during the simulator runs.

All operating sequences were videotaped using two cameras located to give the best possible view of operators' switch manipulations. Operators wore wireless microphones and their communications were dubbed onto the videotape. At a point later in time, a new TPIC was completed using the videotapes and observers' TPICs. Additional procedures were developed to account for deviations from the analyzed scenario and scoring from incomplete records. For example, some deviations result when operator actions performed early in the scenario dynamically modify the nature of tasks performed later in the scenario. Incomplete records result from simulator equipment malfunctions and mistakes in operating the recording equipment.

Several questionnaires have been used to obtain data from the operators and STAs. A biographical questionnaire requests such information as position, age, education, commercial NPP operator experience, total commercial NPP experience, and nuclear Navy experience. In the SROA experiments, an event questionnaire was completed by the operators after each exercise and provided ratings on: (1) difficulty of event diagnosis, (2) difficulty of response, (3) adequacy of procedures, (4) control board design, and (5) plant indications. A five-point scale ranged from "no problem at all" to "a significant problem." In the more recent training simulator experiments, a formal procedure was followed in the use of SWAT providing two types of data. One part of the SWAT procedure, completed prior to the simulator exercises, results in a customized scaling of the way each operator perceives subjective workload, i.e., the manner in which each individual subjectively combines perceptions of time, mental, and stress loads into a single scale. Another part of the SWAT procedure follows completion of each operating sequence and involves each operator rating the sequence for its time, mental, and stress loads.

In one training simulator experiment, another approach in the measurement of operator performance underwent an initial assessment (Ref. 14). The premises of this approach are that experts are good raters in discriminating and ranking performances, and that the primary high level concern is with total system performance rather than the performance of operators, of systems, or of other elements comprising the overall system. The general approach consists of: (1) collecting expert ratings of crew performance demonstrations, (2) analyzing data collected from the experts with the PMS data collected from the simulator exercises using special mathematical, control analytic techniques, (3) identifying candidate system performance measures that rate performance like experts, and (4) confirming the candidate measures on additional simulator exercises. An initial assessment of the first step in this approach was included as part of the current experiment through the development of an instructor rating form (IRF). The scope of the IRF consisted of numeric ratings of crew and individual operator performances, as well as three open-ended questions asking the expert to identify the factors or evaluation categories believed to be important in evaluating the particular operating sequence, the specific actions taken by the crew that were especially good and significantly influenced the performance evaluation, and the specific actions that could have been better performed. It is noted that construction of the IRF was intended to assess an open-ended written technique to collecting expert subjective data in a field setting, and that these data supported another performance measurement research effort beyond the simulator experiment itself.

A fundamental issue in performance assessment is that there is some high level of performance for which operators are trained and are expected to maintain along with the ability of experts to sensitively distinguish superior from less-than-superior performance. On the one hand, one may hypothesize that the performance standards associated with operator licensing drive actual operator performance upward to an asymptote on some hypothetical performance curve, such that there is an intrinsic restriction of range in the performance data. On the other hand, experts may make distinctions across operator performances because they dynamically synthesize an array of observed information in comparison with their own mental image of how operators should respond to an operating sequence.

Based on experience with the IRF's open-ended format, another technique was developed using rating scales. The scope of this effort consisted of the identification by two experts of a set of dimensions of operator performance, which when combined with system data obtained from PMS records, would provide a comprehensive description of operator actions vis a vis system performance. A rating form was drafted consisting of one question for expert evaluation of overall crew performance, and ten questions and rating scales for describing operator performance on such dimensions as communications, knowledge of immediate actions, efficient use of time available, monitoring parameters and alarms, performing within job assignments, diagnosis of plant conditions, and awareness of changes in plant state. These experts then reviewed videotapes of fifteen BWR crews responding to a loss of feedwater operating sequence, and completed their ratings after studying each videotape. These data are currently being used in an analysis with PMS data to identify candidate performance measures.

**Field Data from Plant Records**

The purpose of collecting field data from plant records of actual operating experiences is to provide a calibration of simulator results so that they can be more confidently extrapolated to field conditions. Field data were obtained for those operating sequences included in the initial PWR and BWR SROA experiments (Refs. 3 and 4). The following discussion includes techniques for data collection, and problems and limitations with field data (Ref. 5).

The field data collection procedure in the SROA program consisted of pre-visit, on-site, and post-visit activities. The pre-visit activities included screening of Licensee Event Reports (LERs) to identify events useful to the study and identifying sites where these events occurred. The on-site activities involved screening and collecting detailed plant records. The post-visit activities were reconstructing the event, identifying operator actions, and tabulating response times.

Two problems seem to limit the use of field data. First, there seems to be a scarcity of useable data, and this is due to several factors. These factors include misleading LER titles, a lack of sufficient chronological data (event and action times), and differences across plants as to the types, accuracy, and specificity of data retained. Second, the precision of data on operator response times seemed to vary widely. In plant logs that are completed manually, actions are typically recorded rounded to the nearest minute, and are often entered some time after the event such that the operator may not have perfect recall of response times.

It is noted that a field data study is underway in association with the most recent PWR training simulator experiment (Ref. 14). However, field data are only being collected at the plant for which the reference simulator is being used.

## PRODUCTS AND EXPERIMENTAL RESULTS

Over the course of the SROA and training simulator experiments and studies, several products and experimental results have emerged. These include a refined research methodology using the NPP training simulator for a field setting, a data store on operator performance, and specific results pertinent to NRC regulatory positions. Additional results pertain to the relationship of expert ratings with objective measures of performance. The products and results discussed in the following sections are presented at a fairly general level of detail, and specific reports are cited for the reader requiring more detailed information.

### Research Methodology for Field Setting

A product resulting from this series of SROA and training simulator experiments is a sophisticated, refined research methodology for collecting operator performance data using the tr ining simulator as a field setting. Important elements of this methodology are the interface of the utility and operators with the research investigators, the objective and subjective measures of operator and crew performance, and specific tools used in data collection (Refs. 13 and 14).

The interface of the research investigators with utility, training, and operator personnel has been carefully nurtured over the course of the experiments. These experiments have involved primarily two utilities with their respective simulator facilities, and several other utilities provided field data in the SROA program. The utilities have generously provided gratis the use of the training simulators and simulator time, as well as allowing the overlapping of experimental data collection with ongoing operator training.

The development of objective and subjective measures of operator performance provides a comprehensive and detailed set of variables for describing and analyzing operator actions. The system/task analysis supports a system engineering approach to the definition of performance measures. The task performance measures allow the scoring of actions of each individual operator, and the functional measures are used to score the performance of the crew, as an operating team, in controlling those plant parameters that are important to an operating sequence. The subjective measures of operator performance entail sources and types of data different from objective measures. Subjective operator measures have been used to gain additional understanding of mental workload relative to control room staffing and difficulty of operating sequences, and generally provide a technique for assessing the operator's cognitive behavior. The development of subjective measures for obtaining experts' judgments of operator performance lends itself to the enrichment of the operator performance data base. While work is still required to refine these subjective measures, they are an important part of the general research methodology.

There are several tools used in the collection of training simulator data. The definition of performance measures, of course, influences the selection of the best tool for obtaining the necessary data. The scoring of operator performance on the objective task performance measures and functional measures is supported by the PMS records, videotapes, off-line computer programs, and various forms used by the investigators. The PMS is a powerful tool with regards to the accuracy, timing, and fine level of detail with which operator switch actions and plant digital and analog data are recorded (Ref. 18). Videotaping the simulator exercises provides a permanent type of record of operator performance during the operating sequences and can also be used in subsequent analyses. A tool for the measurement of operator mental workload is the SWAT, which is a recognized technique in mental workload assessment (Ref. 16). Procedures for collection and analysis of data using the SWAT scales have been standardized.

In general, then, the research methodology entailed in the training simulator experiments represents an up-to-date approach for the collection of operator performance data in a field setting. As a research methodology, it allows an empirical assessment supporting regulatory and operational issues.

**Data Store on Operator Performance**

Over the course of the SROA and training simulator experiments and studies, a growing data store on operator performance has been developed. The richness of these data are reflected by objective and subjective performance data collected during the PWR and BWR operating sequences, and the collection of field data for calibration of simulator results. Summaries of these data have been included in technical reports, and raw data have been stored on PMS computer tapes, rating forms, and videotapes.

From the SROA experiments, data exist on operator response times and error rates. These data were collected from 34 PWR crews on a total of 21 operating sequences (Refs. 3 and 6), and from 24 BWR crews on 10 operating sequences (Ref. 4). Operator data on operating experience were also collected to relate operator experience with performance. Field data on actual operating experiences have been collected for calibration of SROA simulator data (Ref. 5).

From the more recent training simulator experiments, data exist on the task performance measures and functional measures. These data were collected from 16 BWR crews on 4 operating sequences (Ref. 13), and from 20 PWR crews on 4 operating sequences (Ref. 14). Field data corresponding to the latter operating sequences are currently being collected and should be reported shortly.

**Results Pertinent to NRC Regulatory Positions**

The scope of results from the series of training simulator experiments and studies involved several NRC regulatory positions. In the SROA program, the concern was collecting empirical operator performance data in support of the N660 Standard. In the more recent training simulator experiments, research hypotheses concerned the qualifications of the SRO in terms of the level of supervisory experience, and control room staffing relative to the addition of an STA to support monitoring and diagnosis of operating sequences.

From the SROA experiments, specific results included that operator response times fit a log-normal distribution, and that the overall error of omission rate was about 5% (Refs. 3, 4, and 6). Reaction times did not seem to be related to the severity of the casualty. The effects of operator experience on response times and error rates were not consistent across experiments, and there was no apparent relationship between error rate and response times. Based on their ratings, operators reported that the operating sequences differed in difficulty. The collection of field data on actual operating experiences for calibrating the simulator data resulted in some suggestive, though certainly not conclusive, findings. These findings included that response times for experienced BWR operators were six to seven times shorter, and less variable, in the simulator compared to the field data. For relatively inexperienced PWR operators, response times in the simulator may be as great or greater than typical response times in the field. However, this latter result was modified by subsequent analysis showing more comparable response times in field data collected from the plant whose reference simulator was used in the experiment.

The collection of these SROA data provided a basis for the quantitative prediction of operator performance through a coupling with task analytic data. An Operator Personnel Performance Simulation (OPPS) computer model was developed to calculate a time-reliability distribution predicting time to correctly complete safety-related operator actions, and to provide data for the objective evaluation of quantitative NPP design criteria (Ref. 12). The OPPS model, while representing only an initial developmental effort for predicting operator reliability, demonstrates the application and extrapolation of SROA data.

Specific results from the two latest training simulator experiments concerned the influence on performance from the experience level of the SRO and the presence of an STA (Refs. 13 and 14). In general, no significant differences in overall performance were found between crews led by SROs having high and low experience levels. Contrary to the first BWR experiment, in the current PWR experiment the reported mental workloads of the SROs who were supported by STAs were significantly lower than the workloads of SROs not supported. At a more detailed level, additional differences surfaced. Common to both experiments were the findings that the STAs reported significantly lower mental workloads compared to SROs and ROs, and there were significant differences in subjective workload for the four operating sequences. In the first experiment, crews led by SROs having less experience tended to have shorter task performance times. In the current experiment, some limited correspondence was found between the task performance measures and the subjective ratings of performance by experts.

## GENERAL RESEARCH ISSUES

Through the conduct of training simulator experiments and analysis of operator performance data, several research issues have surfaced promising improvements to the research methodology in general, and to the asssment and measurement of NPP operator performance in particular. The discussions of the following issues are clustered according to their relationships with the research methodology and performance measurement.

Four research issues seem generally associated with the research methodology. First, the use of the training simulator as a field setting for the design and conduct of quasi-experiments brings with it a number of constraints. These include limitations with operators having qualifications meeting some experimentally-desired mix, and the reliability of simulator hardware and software to minimize potential loss of data. It is important to note that in the use of this research methodology the question should be raised as to whether the simulator experiment is the best approach for assessing the particular research hypothesis or objective. Variations of the methodology may provide more applicable approaches, such as increased use of expert judgment in assessing operator actions and more detailed post-hoc analysis of the data.

Second, the experiments have shown that training simulators can be successfully used for human factors research. These experiments have dealt primarily with off-normal emergency operating sequences, reflecting the NRC's concern with operator safety and the factors shaping operator performance. The experiments would seem to offer some increased understanding of operator performance during normal operations as well. By studying the performances of superior and less-than-superior crews and operators, the best techniques for performing tasks may be identified and then shared with all operators through training. The intent with this effort is to increase operational efficiency by bringing all operators up to the same high level of performance during both normal and off-normal conditions using operators demonstrating superior performance as the target goal.

Third, a concern exists with the extrapolation of simulator data to the actual NPP control room. This issue of generalizability seems especially problematic for rare events that place high stress and workload on the operators. Considering that as part of the experiments the operators are undergoing simulator training, it seems reasonable for them to expect to see emergency events, and so be more sensitive to promptly detecting symptoms. More field data on similar actual plant events need to be collected to support more accurate calibration of the simulator data.

Fourth, increased attention should be given to the development of an operator computer simulation model for analytic studies paralleling the simulator experiments. The OPPS model was an initial step in assessing safety-related operator actions. The operator model would provide increased understanding of operator performance by systematically varying input and process variables. The model could be used to develop data for a broader set of operating sequences and operator characteristics than would be feasible within the bounds of the simulator experiments.

There are several research issues that are associated with human performance measurement. Some of these issues are tied not only to the simulator experiments but have significant linkages with operator training and the operator simulator licensing examination.

At a high level is the development and validation of adequate human performance measures. This issue encompasses four other considerations as discussed below. From a system perspective, operator performance measures must be both comprehensive, in terms of content validity, and sensitive, in terms of differentiating operators according to their competencies.

The first consideration is that the measures are a conceptual characterization of skilled performance. This characterization includes a specification of what dimensions of operator performance are important in an absolute sense, as well as what relative range of performance on those dimensions is permissible for a licensed operator. In the operator simulator licensing examination, the NRC is most concerned with distinguishing between safe and unsafe performance leading to a pass/fail decision. In that sense, the difference between "superior" and "good" performance is of lesser importance. From an operational standpoint, however, a question may be raised as to the cost consequences to the plant and its power generation from any operator performance that is less than superior.

The second consideration is the continuing need to measure individual performance proficiency apart from the performance of the crew. Task performance and functional measures have been discussed in this context. The NRC is, of course, concerned with individual operator performance in that operators are tested and licensed as individuals and not as crews. A larger question remains, though, as to what increment in crew performance may result from some formal sorting of operators based on individual proficiency with regards to qualifications, skills, and knowledges.

The third consideration is the development and assessment of measurement techniques for cognitive tasks, especially those performed under emergency conditions. Taxonomies of operator cognitive behavior are useful for describing behavior but commonly lack measurement techniques that are adequately rigorous. More research is needed to develop techniques that can in some manner capture the decision making and troubleshooting that operators perform. In the absence of modeling the operator's cognitive processes, it is still possible to study their effects in terms of operator actions and the relationships of those actions with system dynamics.

The fourth consideration is assessing the relationship of operator performance to overall system performance. The contribution of the human element to system performance needs to be better understood and measured. One aspect of this relationship involves the rate at which system performance will increase in response to an improvement in operator performance, which is an indication of the cost-effectiveness of human factors improvements.

While the above considerations have addressed the definition of human performance measures, a separate issue is the manner in which data from the training simulator are collected and analyzed. Some attention has been given to the use of automated performance mesurement, which combines automatic data collection with real- or near-real-time scoring on specified performance measures. Clearly, the PMS already provides a method for automatic data collection. Research is still needed to identify system and operator performance measures, as previously described. The automation of scoring would seem to provide a valuable tool supporting operator training and licensing.

In conclusion, there are several research issues that continue to challenge researchers in finding improvements to the research methodology and the measurement of operator performance. The use of the training simulator as a field setting for the collection of operator data involves a refined research methodology applicable for addressing many current human factors issues, although it certainly is not a panacea for all research questions.

## REFERENCES

1. American Nuclear Society, "Criteria for Safety--Related Operator Actions," ANSI-N660/ANS-58.8, August 1982 (DRAFT).

2. P. M. Haas and T. F. Bott, "Criteria for Safety-Related Nuclear Plant Operator Actions: A Preliminary Assessment of Available Data," USNRC Report NUREG/CR-0901, ORNL/NUREG/TM-330, July 1979.

3. T. F. Bott, E. Kozinsky, C  Crowe, and P. M. Haas, "Criteria for Safety-Related Nuclear Power Plant Operator Actions: Initial Pressurized Water Reactor (PWR) Simulator Exercises," USNRC Report NUREG/CR-1908, ORNL/NUREG/TM-434, September 1981.

4. A. N. Beare, D. S. Crowe, E. J. Kozinsky, D. B. Barks, and P. M. Haas, "Criteria for Safety-Related Nuclear Power Plant Operator Actions: Initial Boiling Water Reactor (BWR) Simulator Exercises," USNRC Report NUREG/CR-2534, ORNL/TM-8195, November 1982.

5. A. N. Beare, R. E. Dorris, E. J. Kozinsky, J. J. Manning, and P. M. Haas, "Criteria for Safety-Related Nuclear Power Plant Operator Actions: Initial Simulator to Field Data Calibration," USNRC Report NUREG/CR-3092, ORNL/TM-8599, February 1983.

6. D. S. Crowe, A. N. Beare, E. J. Kozinsky, and P. M. Haas, "Criteria for Safety-Related Nuclear Power Plant Operator Actions: 1982 Pressurized Water Reactor (PWR) Simulator Exercises," USNRC Report NUREG/CR-3123, ORNL/TM-8626, June 1983.

7. D. W. Jones, D. K. Baer, and C. C. Francis, "Nuclear Power Plant Training Simulators: Their Use in Operator Training and Requalification," USNRC Report NUREG/CR-1482, ORNL/TM-395, July 1980.

8. C. A. Belsterling, S. Pandey, J. C. Praskievicz, E. J. Kozinsky, L. Felkins, and P. M. Haas, "Specification and Verification of Nuclear Power Plant Training Simulator Response Characteristics, Part I: Summary of Current Practices for Nuclear and Non-Nuclear Simulators," USNRC Report NUREG/CR-2353, Vol. 1, ORNL/TM-7985/P1, December 1981.

9. P. M. Haas, D. L. Selby, T. W. Kerlin, and L. Felkins, "Specification and Verification of Nuclear Power Plant Training Simulator Response Characteristics, Part II: Conclusions and Recommendations," USNRC Report NUREG/CR-2353, Vol. 2, ORNL/TM-7985/P2, May 1982.

10. D. B. Barks, E. J. Kozinsky, and S. Eckel, "Nuclear Power Plant Control Room Task Analysis: Pilot Study for Pressurized Water Reactors," USNRC Report NUREG/CR-2598, ORNL/Sub/79-40432/1, May 1982.

11. D. B. Barks, F. E. Gomer, E. J. Kozinsky, and G. F. Moody, "Nuclear Power Plant Control Room Task Analysis: Pilot Study for Boiling Water Reactors," USNRC Report NUREG/CR-3415, ORNL/Sub/79-40432/1, September 1983.

12. E. J. Kozinsky, L. H. Gray, A. N. Beare, D. B. Barks, and F. E. Gomer, "Safety-Related Operator Actions: Methodology for Developing Criteria," USNRC Report NUREG/CR-3515, ORNL/TM-8942, March 1984.

13. A. N. Beare, R. E. Dorris, and L. H. Gray, "Simulator Experiments: Effects of Experience of Senior Reactor Operators and of Presence of a Shift Technical Advisor on Performance in a Boiling Water Reactor Control Room," ORNL/TM-9327, December 1984.

14. A. N. Beare, M. D. Donovan, and D. L. Lassiter, "The Effects of Supervisor Experience and the Assistance of an STA on the Performance of Requalification Training Crews in a PWR Control Room Simulator," draft NUREG/CR, April 1985.

15. D. T. Campbell and J. C. Stanley, Experimental and Quasi-Experimental Designs for Research, Rand McNally, Chicago, IL, 1963.

16. G. B. Ried, C. A. Shingledecker, and F. T. Eggemeier, "Application of Conjoint Measurement to Workload Scale Development," pp. 552-526 in Proceedings of the Human Factors Society - 25th Annual Meeting, The Human Factors Society, Santa Monica, California, 1982.

17. D. Burgy, C. Lempges, A. Miller, L. Schroeder, H. VanCott, and B. Paramore, "Task Analysis of Nuclear Power Plant Control Room Crews: Project Approach Methodology," USNRC Report NUREG/CR-3371, Volume 1, September 1983.

18. E. J. Kozinsky and R. W. Pack, "Performance Measurement System for Training Simulators," Electric Power Research Institute Report EPRI-NP-2719, November 1982.