INL/CON-06-01256
PREPRINT

Developing Human Performance Measures

PSAM 8

Bruce P. Hallbert
Jeffrey C. Joe
Larry G. Blackwood
Donald D. Dudenhoeffer
Kent F. Hansen

May 2006

DEVELOPING HUMAN PERFORMANCE MEASURES

Bruce P. Hallbert/Idaho National Laboratory

Jeffrey C. Joe/Idaho National Laboratory

Larry G. Blackwood Idaho National Laboratory

Donald D. Dudenhoeffer Idaho National Laboratory

Kent F. Hansen Massachusetts Institute of Technology

SUMMARY/ABSTRACT

Through the reactor oversight process (ROP), the U.S. Nuclear Regulatory Commission (NRC) monitors the performance of utilities licensed to operate nuclear power plants. The process is designed to assure public health and safety by providing reasonable assurance that licensees are meeting the cornerstones of safety and designated crosscutting elements. The reactor inspection program, together with performance indicators (PIs), and enforcement activities form the basis for the NRC's risk-informed, performance based regulatory framework.

While human performance is a key component in the safe operation of nuclear power plants and is a designated cross-cutting element of the ROP, there is currently no direct inspection or performance indicator for assessing human performance. Rather, when human performance is identified as a substantive cross cutting element in any 1 of 3 categories (resources, organizational or personnel), it is then evaluated for common themes to determine if follow-up actions are warranted.

However, variability in human performance occurs from day to day, across activities that vary in complexity, and workgroups, contributing to the uncertainty in the outcomes of performance. While some variability in human performance may be random, much of the variability may be attributed to factors that are not currently assessed. There is a need to identify and assess aspects of human performance that relate to plant safety and to develop measures that can be used to successfully assure licensee performance and indicate when additional investigation may be required. This paper presents research that establishes a technical basis for developing human performance measures. In particular, we discuss: 1) how historical data already gives some indication of connection between human performance and overall plant performance, 2) how industry led efforts to measure and model human performance and organizational factors could serve as a data source and basis for a framework, 3) how our use of modeling and simulation techniques could be used to develop and validate measures of human performance, and 4) what the possible outcomes are from this research as the modeling and simulation efforts generate results.

INTRODUCTION

Human performance is a key component in the safe operation of nuclear power plants [1]. Through the reactor oversight process (ROP), the U.S. Nuclear Regulatory Commission (NRC) monitors plant performance in a number of cornerstones and crosscutting elements (e.g., human performance, the utilities' corrective action program, and the safety conscious work environment) to assure public health and safety. The reactor inspection program, together with performance indicators (PIs), assessment, and enforcement activities form the basis for the NRC's risk-informed, performance based regulatory framework. Many of the performance indicators currently used are described in NEI 99-02 "Regulatory Assessment Performance Indicator Guideline," which is endorsed by the NRC.

Performance indicators use objective data to monitor performance within each of the "cornerstone" areas. The data which make up the performance indicators is generated by the utilities and submitted to the NRC on a quarterly basis. Each performance indicator is measured against established thresholds which are related to their effect on safety [2].

While human performance is a key component in the safe operation of nuclear power plants and a designated cross-cutting element of the ROP, there is currently no direct inspection or performance indicator for assessing human performance. Rather, when human performance is identified as a substantive cross cutting element in any 1 of 3 categories (resources, organizational or personnel), it is then evaluated for common themes to determine if follow-up actions are warranted. Nevertheless, variability in human performance occurs from day to day, across activities that vary in complexity, workgroups, and across plants, contributing to the uncertainty in the outcomes of performance. Some of this variability may be random, though much of the variability may be attributed to factors that are not currently assessed. There is a need to identify and assess aspects of human performance that related to plant safety and to develop measures of these performance aspects that could be used to assess licensee performance and indicate when additional actions may be required.

The objective of this paper is to demonstrate an approach for establishing a technical basis from which human performance measures could be developed. Specifically, historical data that may indicate some connection between human performance and overall plant performance, potential measures derived from industry models of human performance and processes in nuclear power plants, the means for modeling and simulating their relationship to human and plant performance, and potential relevance to reactor oversight processes are explored.

PAST STUDIES CONNECTING HUMAN AND PLANT PERFORMANCE

As one of our initial steps to establishing a technical basis for human performance measures, we explored whether there was any prior evidence that established an empirical relationship between human performance and nuclear plant performance. The data for this analysis came from two sources. The human performance data was obtained from the NRC's Human Factors Information System (HFIS). HFIS is a database that provides a general overview of the types and approximate numbers of human performance issues documented by either the NRC or licensees. Performance indicator data for the cornerstones of the NRC's ROP served as the measures of plant performance. The HFIS data analyzed were the number of human performance related "hits" or causal factors extracted from NRC Inspection Reports (IRs) and Licensee Event Reports (LERs). These data were available for the years 2000 through 2004. Three ROP measures were analyzed: unplanned scrams per 7,000 critical hours, unplanned power changes per 7,000 critical hours and safety system unavailability. These are referred to as initiating events #1and #3, and mitigating systems #1 respectively, which corresponds to how they are listed in [2].

These data were binned into contingency tables. There were two reasons for pursuing this line of analysis. First, the sparseness of the data at higher levels of HFIS hits and the difference in the trends at those high levels suggest that considering HFIS hits on a more coarse measure, such as high versus low might be worthwhile. Secondly, the ROP initiating events measures (unplanned scrams and unplanned power changes) measures, although calculated as rates, really represent very few events per facility per year. This clustering would make it unlikely that normality or other distributional assumptions required for regression or other similar modeling methods would be valid. Hence, a categorization of these variables seemed to make sense.

The HFIS and ROP measures were both dichotomized to give high vs. low comparisons. The cutpoints used for the dichotomization measure in each case were based on exploratory analysis to determine which value would most likely give the best overall results in terms of statistical significance while maintaining sufficient cell sizes in the tables for analysis. The values chosen were 175 for HFIS hits, 0.4 for unplanned scrams, 0.8 for unplanned power changes, and 0.4 for safety system unavailability.

The first set of analyses looks at simple 2x2 tables and the simple relationship between the ROP measures and HFIS hits the previous year. In the second set of analyses, the data are expanded to 2x2x2 tables in which the previous year's ROP measures are also included as controlling variables. The idea is to see if the previous year's HFIS hits still have an effect after controlling for the year-to-year self-correlation of the ROP measures. For these analyses the data were combined over the years. This boosts the sample sizes, increasing the likelihood of detecting significant effects, provided there are no major differences between the relationships across years.

Results are reported in terms of odds ratios. The odds ratio is a fundamental parameter of contingency table models and it measures the degree of association between the variables in the table. The chi-square test indicates whether the deviation from independence is statistically significant. An odds ratio of 1.0 corresponds to independence. The further away the odds ratio is from 1.0 in either direction, the stronger the relationship.

The chi-square tests for independence showed statistically significant departures from independence for the ROP measures unplanned scrams and safety system unavailability and a marginally significant value for unplanned power changes. Interestingly, in the 2x2 contingency table analysis, safety system unavailability showed the strongest relationship with previous year's HFIS hits while it showed the least evidence of a trend in our initial plot analysis. The odds of a high value for safety system unavailability were 8.0 times greater for those cases where HFIS hits were high vs. low. This compares to odds ratio values of 4.6 and 2.6 for unplanned scrams and unplanned power changes respectively. These initial results show that human performance is significantly related to a plant's safety performance.

The results of the 2x2x2 contingency table analysis were analyzed using ANOVA type log-linear models. The odds ratio was based on the expected frequencies for the model with the previous year's HFIS and ROP effects included. Results showed that controlling for the previous year's ROP measurement value raises the p-value and lowers the odds ratio for the effect of HFIS hits in the previous year. However, the effects remain marginally significant and some of the lack of significance may be due simply to the small counts in some of the cells. If more data were available (e.g., by adding results from additional years) the cell counts would be larger and the effects possibly more significant. The odds ratios, although smaller than for the 2x2 table analysis, still range from 2.3 to 5.3, indicating at least a doubling of the odds of being at the high end of the ROP measures if the previous year's HFIS hits are high.

The general conclusion from this analysis of existing data is that there is, in fact, empirical support showing human performance is related to and affects overall plant performance. The analysis shows that as human performance degrades at nuclear power plants, the likelihood increases of plant performance also degrading. These results are corroborated by past research, namely, NUREG/CR-6753 [1], which studied the direct contributions of human performance to risk in significant operating events at commercial nuclear power plants. The NRC Accident Sequence Precursor (ASP) Program and the Human Performance Events Database (HPED) were used to identify safety significant events in which human performance contributed to changes in risk. The sensitivity analyses performed using these data showed that human performance contributed significantly to analyzed events. Two hundred and seventy human errors were identified in the events reviewed and multiple human errors were involved in every event. Latent failures (i.e., errors committed prior to the event whose effects are not discovered until an event occurs) were present four times more often than were active errors (i.e., those occurring at or following event initiation). The latent errors included failures to correct known problems and errors committed during design, maintenance, and operations activities. Based on both of these findings, we believe there is evidence that gives some indication of connection between human performance and plant outcomes.

INDUSTRY FRAMEWORKS AND HUMAN PERFORMANCE PROGRAMS

Various nuclear industry groups and associations, such as the Institute of Nuclear Power Operations (INPO), the Electric Power Research Institute (EPRI), and the Nuclear Energy Institute (NEI) have written guidance documents on human performance and work processes in nuclear power plants. These reports have been written in part to support ongoing research with the goal of providing practical tools for performance improvement (e.g., measures of human performance). One document jointly written by NEI, EPRI, and INPO [4] could serve as the initial starting point for a framework that could be used to develop an approach to measuring human performance. For example, one framework in this document (i.e., the human performance process map) has the potential to be an excellent starting point because it provides an integrated framework of organizational factors and performance improvement, and because it explicitly shows how organizational factors can contribute to a variety of human errors in the workplace. Specifically the human performance process map shows the actions and resource inputs needed to achieve the desired outputs. This process map can be used as a framework in a modeling and simulation approach because it has the basic input and output elements for nuclear power plants already mapped out. With a valid and representative framework developed by the nuclear industry in place, one of the major hurdles in using a modeling and simulation approach for the development of human performance measures is addressed.

In addition, a number of individual utilities, under their own initiative, have created a variety of in-house programs and efforts to track and trend human performance. These programs are based on guidance documents created by industry organizations and research by Reason [5], and could serve one of many data sources for the modeling and simulation approach we plan to use to develop human performance measures. In particular, we have met with a number of individuals from a variety of nuclear power plants in the US, have seen their programs to track and trend human performance, and in some instance been given data from their program to use in our research efforts. As we continue to develop our modeling and simulation approach, we anticipate that these data will provide valuable input and possibly validate our approach.

HUMAN PERFORMANCE MODELING AND SIMULATION

The impact of human performance on nuclear power plant safety and operations is acknowledged to be very important, but it is also difficult to ascertain. The difficulty arises from the complexity of nuclear plant operations and the many pathways by which human performance can affect operations. Regulators want to be assured that a plant is operating safely. This means that the physical system must be adequately designed, operated, and maintained. What the regulators would also like is some means of knowing that the managerial and organizational system is also adequately designed, operated and maintained. This latter issue is where modeling can be very useful.

The difficulty is that operators of a nuclear power plants cannot address human performance issues experimentally. For example, senior management cannot create a new reward system to encourage changed behavior in the operations or maintenance staff and quickly discover whether or not the change is working. It would take years to decide if a new policy is producing desired results. Indeed, even knowing what to measure to ascertain if the performance is improving is difficult. Complicating matters is the complexity of operations systems and the fact that changes may propagate in very diffuse ways.

This difficulty is what gives modeling and simulation approaches their great power and usefulness. If one can adequately represent the system in a model then system performance can be assessed by multiple simulations. The impacts of various proposed new policies, be they managerial or regulatory, can be analyzed, studied, and assessed. It is the rapid, quantitative, nondestructive testing of policy options that makes models so useful.

The key issue is to find a means of constructing a model that adequately represents the real system being modeled. We believe that the issue is partially addressed with the use of the human performance process map described by the NEI [4], but we consider other possible issues and approaches below. Another issue in this modeling approach is to show how its representation of the real system, (i.e., the system's behaviors, how changes in the system produce different outcomes, etc.) provide useful information to decision-makers (e.g., regulators). Modeling can help inform regulatory assessments by showing how interrelated plant outcomes such as safety, productivity, and quality can be associated with human performance. While human performance only partially explains plant outcomes, the association is consistent and predictable. In what follows, we outline a possible modeling and simulation approach that makes explicit the relations between the plant organizational structure, the nature of the work done in different sectors of the organization, and the performance of the workforce that accomplishes the work. This approach is based on modeling and simulation work developed by HGK Associates [6], which compliments our efforts to use the frameworks and process maps [4]. The purpose of this is to present in more detail an initial representation, or model, of our effort to incorporate human performance into overall performance at nuclear power plants. Two basic models (for plant management and maintenance department) are described below with a particular emphasis placed on how human performance issues can be incorporated.

We begin the model development by identifying the set of activities within the plant and how these activities are conducted. It is convenient to distinguish activities according to the different components of the organization. A typical organization chart for a nuclear power plant is shown in Figure 1.

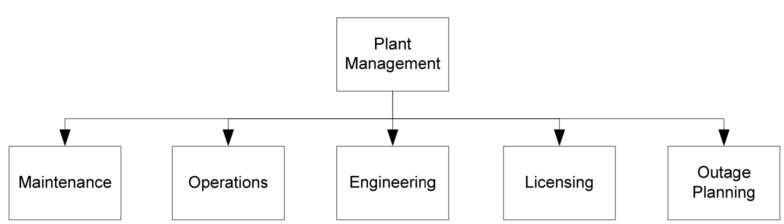

Figure 1. Typical Nuclear Power Plant Organization

We include those sectors of the organization that are important to safety but omit other sectors and functions such as personnel, payroll, taxes and accounting, etc.

The key to creating a model is to identify what types of work occurs in a sector, and what types of personnel carry out the work. Also, we must identify the policies that guide allocation of resources to various tasks. The plant management sector is discussed first.

Plant Management Sector

Plant management is composed of a small workforce that includes the plant manager, any assistant mangers, and, to some extent, the other sector managers. In addition there is a support staff to assist management. The plant management staff must carry out several different types of work including, but not limited to:

- Administrative work
- Resource allocation
- Review and assessment work
- Meetings
- Communication with corporate management
- Communication with regulators

Work to be done in this sector is represented in Figure 2.

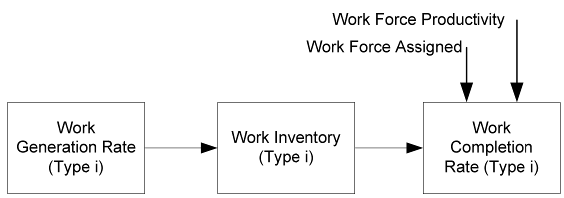

Figure 2. Typical Flow of Management Work

Note that the breakdown of operations contains flows of materials in the form of rates. These rates enter and leave quantities that obey a conservation principle. That is,

Content (t+dt) = Content (t) + (Flow rate in – Flow rate out) * dt Eq. 1

In the simple model above, the workforce is a conserved quantity. Flows into and out of the manpower levels are not shown explicitly. Moreover, there is a separate flow for each type of work done within a sector.

The work generation rate is different for each type of work. Usually there is a steady component and a time varying component that depends upon plant conditions. For example, there are usually regularly scheduled meetings for the sector management. The number and duration of meetings could change dramatically if unusual conditions arise such as an NRC or INPO visit.

The accomplishment of work depends upon the number of people doing the work and their productivity. The productivity of managers is itself a function of many factors including:

- Available support staff
- Workload
- Pressure from corporate management
- Pressure from regulators
- Schedule pressure
- Motivation

Motivation is a composite of several factors including the corporate reward system, the work environment, and growth prospects. Obtaining information on all of these factors is a key issue in model development. Working with plant managers it is possible to acquire data to help quantify such matters.

Maintenance Sector

The maintenance sector is more complex than the management sector, both in terms of the quantity and nature of the work to be done, and also in the composition of the workforce. The workforce is usually composed of:

- Sector management
- Management support staff
- Craft supervisors
- Craft labor
- Craft support personnel

The different types of workers carry out different types of work. Each of these types of work has a workflow similar to that in Fig. 2, but frequently more complex. Craft labor work is illustrated. The work types for the crafts include:

- Inspection work
- Corrective work
- Preventative maintenance
- Quick fix work

The work accomplishment rate will depend upon the craft productivity. However, we expect that not all of the work will be done correctly the first time. We introduce a quality factor that is a dimensionless quantity that represents the fraction of the work done correctly the first time. The overall flow of work is shown in Figure 3.

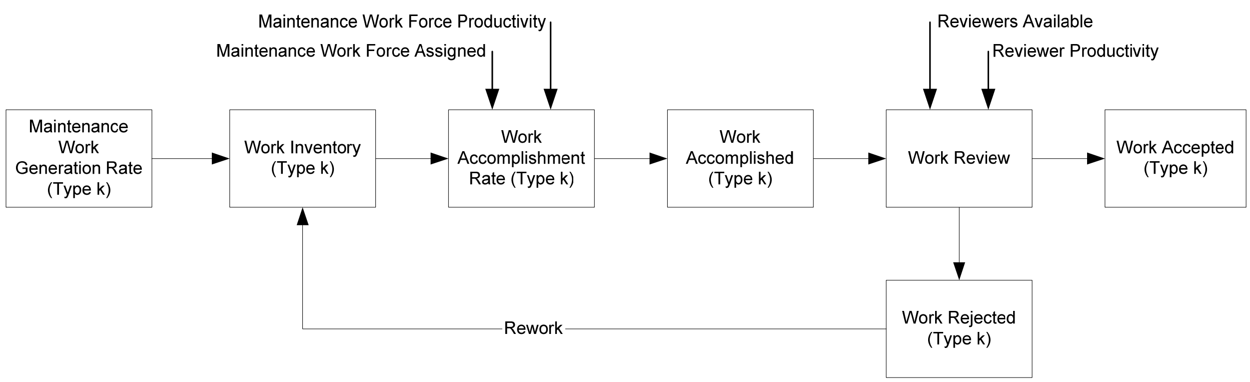

Figure 3. Maintenance Work Flow

The generation rate of work is different for each type of work. Thus, preventative maintenance work is usually on a long-term schedule that changes with plant conditions and/or sector management decisions. On the other hand, inspections are usually mandated and cannot be slipped easily. Part of the task of the sector management is to allocate manpower according to a priority schedule.

The rate of completion of work is determined by the level of manpower and the craft productivity. Maintenance staff productivity can be represented as a function of many factors such as:

- Skill level
- Experience
- Training
- Supervisor availability
- Tools available
- Spare parts inventory
- Procedure quality
- Planning quality
- Motivation

In general we represent the productivity of craft personnel as a multiplicative factor such as:

$$P(t) = P0*f1(t) * f2(t) * * * fn(t) \qquad \text{Eq. 2}$$

where P0 is the nominal productivity and the fk (k = 1, 2,…, n) are dimensionless multipliers on the base number. Thus, a workforce composed of all journeymen would have a maximum skill multiplier whereas a workforce composed of all apprentices would have a minimum skill multiplier.

All of the multipliers are time dependent and can change with many conditions. The overall staff skill level is a weighted composite of the skill of the total craft labor staff. As time goes on the experience level improves and apprentices migrate into the journeyman pool. One of the most rapidly varying factors is the supervisor availability. If conditions in the plant become abnormal the supervisors may spend too much time in meetings and insufficient time with the workers reviewing work products, planning and scheduling work, and maintaining worker motivation.

The quality of craft work is represented in a similar fashion. That is,

$$Q(t) = Q0*q1(t) * q2(t) * * * qn(t) \qquad \text{Eq. 3}$$

The modulating factors could be the same as for productivity, or different depending upon the beliefs of the plant operators.

All of the other sectors can be represented in a similar manner. That is, we breakdown the sector work into its various parts and analyze the creation rates and completion rate. The work is done by a multi-component workforce whose individual performance characteristics are represented by productivities and work qualities.

The sectors do not operate in isolation. The model must incorporate interactions in the form of collaborative work. Thus, the preparation and scheduling of a corrective work order will involve joint work between maintenance, operations, engineering, and possibly licensing. The degree to which coordination is done properly will impact both the productivity and quality of the repair. Staff is assigned to coordination work from all the involved sectors. Any sector that fails to assign adequate staff puts all the other sectors at risk of reduced performance.

Developing Model Parameters

Many of the variables in a model are easy to represent. Thus, manpower levels can be obtained from historic records. Each sector has records that can be used to create a picture of the distribution of manpower throughout the system. Similarly, the flow of most types of work can be obtained from existing sector records. Given the time dependent values of work levels and schedules, it is usually easy to find the flow rates into and out of inventories. However, some generation rates will have to be obtained indirectly. For example, the amount of coordination work for a work order may have to be obtained by averaging over many examples.

Obtaining data on the factors that influence productivity and quality is much more complex than work flow data. Much of the data will have to be inferred from real experience. Thus, if plants have data on work output as a function of time it may be possible to relate changes in productivity to other conditions in the plant. The most likely pathway to obtaining the factors is by interviews with plant staff and management via a Delphic process. Experience with models suggests that certain factors will emerge as the most significant factors in overall performance and one can concentrate extra effort on representing these factors.

Use of Models

The fundamental purpose for creating a model such as this is to help decision makers, whether they are regulators or the plant’s management, understand how human performance can improve, or threaten, safe operation of a nuclear power plant. For example, a key concern regarding plant safety is the availability of safety systems. A decrease in availability could be a reflection of wear and tear in equipment, which is easily recognized. More subtle is a decrease in availability due to poor human performance in service and maintenance activities, or in surveillance and test activities. These deficiencies in turn may relate to a host of sources such as poor scheduling and coordination, inadequate work processes, inadequate work supervision, etc. A model of the type we propose would be useful in identifying and anticipating conditions that can lead to reduced safety.

Once a model has been created and validated it is possible to undertake numerous “what if” studies to better understand the impact of human performance on the system safety. Because of the explicit nature of the model one can trace the impact of policy changes. For example, retirement of senior supervisors can lead to reduced maintenance productivity, which can lead to an increasing work order backlog, which can lead to management concern and increased meetings, which can then propagate into other sectors of the plant. Thus, the model can be useful to decision makers by focusing on the human performance factors that impact work productivity and work quality in all sectors of the plant.

It is also important to remember that a plant model is organic. As new ideas emerge about what types of factors may impact performance it is easy to add these ideas into the model. Thus, there may be many factors that affect motivation of people. These ideas can be introduced into the model with out reworking the entire model. Likewise, it is easy to remove factors that are not significant.

INTEGRATING HUMAN PERFORMANCE MEASURES WITH CURRENT INDUSTRY PRACTICES

The research questions this modeling and simulation approach address are what do measurable plant outcomes such as productivity, quality, and safety tell us about human performance at nuclear power plants? Additionally, can these outcomes be characterized by specific measures and then related to other measures that provide decision makers with advanced notification of declining human performance or degrading conditions? If so, how? More precisely let M = Set of plant outcome measures. The challenge is then to develop a function $F(X): X \subseteq M$ through modeling and simulation that correlates human performance with measurable plant outcomes. In broad terms the desired mapping could be:

$$F(X) = \begin{cases} \text{Green} \rightarrow \text{Human performance within the expected range} \\ \text{White} \rightarrow \text{Human performance outside of expected range, but is still acceptable} \\ \text{Yellow} \rightarrow \text{Human performance is acceptable, but safety margin has been reduced} \\ \text{Red} \rightarrow \text{A significant reduction in the safety margin due to human performance} \end{cases} \quad \text{Eq. 4}$$

The simulation of a model such as the ones described in this paper will establish the mapping of this function. The goal of such a function would be to provide an indication for additional actions such that potential problem areas are addressed prior to their manifestation into a reportable safety related event. When a boundary condition is exceeded (i.e., a Yellow or Red condition), further investigation and follow-up actions may be warranted. Under normal conditions (i.e., Green and White), follow-up investigations may not be necessary.

CONCLUSION

We recognize the development and acceptance of performance measures is nontrivial. They must be defendable and obtainable within the bounds of industry. In addressing human performance assessment, Wreathall [7] has suggested that human performance measures should have the following characteristics:

- *Objective*: it should not be easily manipulable by plants or involve judgments that can be arbitrary.
- *Quantitative*: this allows it to be trended and compared with other measures.
- *Available*: if possible, additional measurements by plants should be avoided as an issue of efficiency.
- *Simple to understand/represent worthy goals / possess face validity*: since plants will tend to 'manage the measure', having the measure represent a worthy goal will tend to improve performance in itself.
- *Related to / compatible with other programs*: if possible, measures should be integrated into existing programs to affect efficiency as minimally as possible.

These are an example of the standards to which individual or collective measures must be assessed, especially if such measures will be utilized as a basis for further diagnosis.

However, the advantages of the modeling and simulation approach we are pursuing are numerous. The framework for the approach was developed by the industry and the measures to be developed and tested are to a large degree based on data already collected and even used in current business practices. Thus, the information should be obtainable without significant burden to the utilities, the measures developed should be valid and reliable, and the approach should meld with the current measures of plant performance and thus is not a significant change in philosophy. Additionally, this overall approach would provide industry with self-assessment measures and trending tools standardized across the industry that are specifically designed to assess human performance.

There is a need to identify and assess aspects of human performance that relate to plant safety and to develop measures that can be used to successfully assure licensee performance. From the evidence presented in this paper, we believe that a technical basis for the development of human performance measures exists through modeling and simulation approaches that use industry developed frameworks of organizational and human performance and existing human performance data from nuclear power plants. The next step is to develop and test candidate measures for use as measures of human performance via modeling and simulation and to continue considering the issues related to their potential use within the existing regulatory environment.

ACKNOWLEDGMENTS

The research was sponsored by the U.S. Nuclear Regulatory Agency's Office of Research.

REFERENCES

[1] NUREG/CR-6753, 2002, "Review of Findings for Human Performance Contribution to Risk in Operation Events".

[2] NEI 99-02 (Rev. 3), 2005, "Regulatory Assessment Performance Indicator Guideline".

[3] OIG-05-A-06, 2004, "Office of the Inspector General: Audit of NRC's Baseline Inspection Program".

[4] NEI LP002, 2001, "Human Performance Process Benchmarking Report".

[5] Reason, J., 1997, *Organizational Accidents: The Management of Human and Organizational Factors in Hazardous Technologies*, Cambridge University Press, Cambridge, England.

[6] Hansen, K.F., Golay M.W., and Kwak, S., 2000, "OPSIM: The Nuclear Power Plant Management and Operations Simulator," HGK Associates, LLC

[7] Wreathall, J., 2000, "Leading Indicators of Human Performance," Paper presented at the ASQ Energy and Environmental Division - Annual Conference, Tucson, AZ.