BROOKHAVEN NATIONAL LABORATORY

BNL-90676-2009

Applying Human-performance Models to Designing and Evaluating Nuclear Power Plants: Review Guidance and Technical Basis

John M. O'Hara

November 30, 2009

Energy Sciences and Technology Department

Brookhaven National Laboratory
P.O. Box 5000
Upton, NY 11973-5000
www.bnl.gov

Human Factors of Advanced Reactors (NRC JCN Y-6529)
BNL Tech Report No. BNL-90676-2009

Applying Human Performance Models to Designing and Evaluating Nuclear Power Plants: Review Guidance and Technical Basis

Prepared for

Division of Risk Analysis
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555-0001

Prepared by

John M. O'Hara
Brookhaven National Laboratory
Energy Sciences and Technology Department
Upton, NY 11973-5000

November 30, 2009

ABSTRACT

Human performance models (HPMs) are simulations of human behavior with which we can predict human performance. Designers use them to support their human factors engineering (HFE) programs for a wide range of complex systems, including commercial nuclear power plants. Applicants to U.S. Nuclear Regulatory Commission (NRC) can use HPMs in their submittals for design certifications, operating licenses, and license amendments. In the context of nuclear-plant safety, it is important to assure that HPMs are verified and validated, and their usage is consistent with their intended purpose. Using HPMs improperly may generate misleading or incorrect information, entailing safety concerns. The objective of this research was to develop guidance to support the NRC staff's reviews of an applicant's use of HPMs in an HFE program. The guidance is divided into three topical areas: (1) HPM Verification, (2) HPM Validation, and (3) User Interface Verification. Following this guidance will help ensure the benefits of HPMs are achieved in a technically sound, defensible manner. During the course of developing this guidance, I identified several issues that could not be addressed; they also are discussed.

CONTENTS

TABLES

FIGURES

ABBREVIATIONS

AFRL	Air Force Research Laboratory
AMBR	agent-based modeling and behavior representation
ANS	American Nuclear Society
ANSI	American National Standards Institute
ARL	Army Research Laboratory
BNL	Brookhaven National Laboratory
BRIMS	Behavior Representation in Modeling and Simulation
CDF	core damage frequency
CFR	U.S. Code of Federal Regulations
CM	configuration management
DMSO	Defense Modeling and Simulation Office
HA	human actions
HARDMAN	Hardware vs. Manpower
HBR	human behavior representation
HED	human engineering discrepancy
HEP	human error probability
HF	human factors
HFE	human factors engineering
HPM	human performance models
HRA	human reliability analysis
HSI	human system interface
IEEE	Institute of Electrical and Electronics Engineers
IMPRINT	Improved Performance Research Integration Tool
LOCA	loss of coolant accident
M&Ts	methods and tools
NMSO	Navy Modeling and Simulation Office
NPP	nuclear power plants
NRC	Nuclear Regulatory Commission
PHRED	Plant-human Review and Effectiveness Decision Tool
PRA	probabilistic risk assessment
SGTR	steam generator tube rupture
SME	subject matter expert
V&V	verification and validation
VV&A	verification, validation, and accreditation
VACP	visual, auditory, cognitive, and psychomotor (component of workload)

Part 1:
Technical Basis and Guidance Development

1 INTRODUCTION

1.1 Background

The methods and tools (M&Ts) used to analyze, design, and evaluate the human factors engineering (HFE) aspects of nuclear power plants (NPPs) are changing rapidly. O'Hara et al., (2009) defined the current trends in HFE M&Ts, identified their applicability to designing and evaluating NPPs, and determined their significance to the safety reviews of the U.S. Nuclear Regulatory Commission (NRC). They identified seven M&T categories:

- Computer Applications for Performing Traditional Analyses
- Computer-aided Design
- Integration of HFE Methods and Tools
- Rapid Development Engineering
- Analysis of Cognitive Tasks
- Use of Virtual Environments and Visualizations
- Application of Human Performance Models

The authors considered the last one, application of human performance models (HPM), as important in NRC safety reviews. HPMs essentially are simulations of human behavior that predict human performance.[1] HFE professionals use them extensively in designing and evaluating many types of complex human-machine systems, including NPPs. HPMs support predictions of the effects of a wide range of factors on human performance, such as changes in tasks, cognitive workload, environmental degradation, personnel characteristics, and training (Baron, Kruser & Huey, 1990). The NRC's staff needs guidance to support their reviews of applicant submittals using HPMs as part of their HFE programs. The purpose of this research was to develop such guidance.

1.2 Using Human Performance Models in Human Factors Engineering

1.2.1 What are Human Performance Models

The term "human performance model" has a long history and was used to characterize theoretical approaches to explaining and predicting specific aspects of human behavior. For example, Wickens (1980, 1987, 1991) proposed a multiple-resource model of workload and attention allocation.[2] However, I am using the term HPM more specifically, that is, to refer to models that are (1) mathematical, programmable, and executable rather than purely explanatory; and (2) applied in the engineering design and evaluation of complex systems. Thus, HPMs are simulations of various aspects of human performance used in HFE applications. Consistent with this approach, Pew and Mavor (1998) define an HPM as:

> …a computer-based model that mimics either the behavior of a single human or the collective action of a team of humans. The term may be used in the context of a self-contained constructive computer simulation that is used to simulate a battle and is run once or many times to produce outputs that reflect the battle outcomes, either individually or statistically. Or it may be used in the context of a

[1] In the general literature, the term "human behavior representation" (HBR) is used synonymously with HPM.

[2] Sheridan (2005) gives a history of normative models of human behavior.

> distributed simulation of the behavior of selected battlefield elements that can be viewed by real crews performing in other battlefield element simulators, such as squads of individual soldiers, ground vehicles, or aircraft, so that the battle can be played out in the simulated world interactively. (p. 11)

Similarly, the DoD (2001a) defines an HPM (or HBR) as follows:

> All HBRs model the behavior of people at some level. The term HBR encompasses representations of parts of individuals (e.g., hands operating controls), individuals (e.g., a specific terrorist or equipment operator), aggregates of individuals (e.g., a crowd, a command staff), and aggregates of organizations (e.g., several organizations responding in concert to an emergency situation). An HBR may depict one or more classical cognitive functions (e.g., perception, inference, planning, control), human performance limitations (e.g., sensing bandwidth, decision latencies) and the effects of behavior moderators (e.g., stress, injury, fatigue, discomfort, motivation and emotion). HBR implementations vary from simple finite state machines to complex knowledge-based systems integrating multiple reasoning paradigms and augmented by simulations of the effects of various behavior moderators. (p. 2)

HPMs long have been employed in formulating and developing U.S. military systems (Allender et al., 2005). For example, the Improved Performance Research Integration Tool (IMPRINT) is one often used in defense applications, such as to analyze the needs for staffing complex systems like a military vessel.[3]

IMPRINT is based on discrete-event, task network simulation model. An analyst breaks a mission into discrete tasks, and then specifies the sequence of their performance. The analyst defines important attributes of the task, such as the persons responsible, the time required to carry out each one, the workload, and the likelihood of accurate performance. Models may cover the effects of outside stressors and training on performance. Having identified and sequenced the tasks, and described their essential attributes, the IMPRINT model is run. Its outputs include the time to accomplish the mission and the likelihood of success.

To illustrate how an HPM represents cognitive functions, I detail IMPRINT's modeling of a workload, for which it offers two options (Mitchell, 2000). One represents workload as a combination of visual-, auditory-, cognitive-, and psychomotor- (VACP) components. For each task, the analyst estimates on a seven-point scale the demand level on each VACP component. When IMPRINT is run, the workload values are summed across concurrent tasks, producing a workload profile for the mission. From this profile, the analyst can identify workload peaks that may compromise performance, and, thereafter, redesign the system or reassign tasks to other personnel to mitigate potential problems with workloads.

While the VACP approach provides workload profiles, it does not estimate their effects on task performance. IMPRINT has other features analysts can use to assess this. They support analysts in modeling workload management strategies by defining rules governing how tasks may be omitted, postponed, interrupted, or reallocated in response to high workloads.

IMPRINT and its predecessor, Hardware vs. Manpower (HARDMAN), were applied to a wide variety of design projects including analyzing the maintenance staffing needs in Army systems

[3] General information on IMPRINT is given in http://www.arl.army.mil/ARL-Directorates/HRED/imb/imprint/Imprint7.htm; retrieved 14 July 2008.

(Archer et al., 2005), and the staffing requirements for military vehicles (Mitchell et al., 2003).[4] I discuss the general scope of HPM applications in the next section.

1.2.1 Scope of HPM Applications

HFE professionals use HPMs for many aspects of design and evaluation, such as:

- assessing the effects of organizational structures on performance
- anticipating workload at the design stage to inform decisions on the requisite size of the crew
- estimating the effects of changes in interface design on the system's performance, and
- modeling differences among displays in supporting situation awareness

These uses of HPMs are summarized briefly below. The summary is intended to illustrate the range of applications, not as a review of HPM applications in HFE. Detailed reviews of individual models and their applications are published elsewhere (DoD, 2003a; Gluck & Pew, 2005; Pew, 2007; Pew, 2008; Pew & Mavor, 1998).

Hansberger and Barnett (2005) developed a task network model to represent organizational structures, individual personnel, tasks, and communication patterns. The output generated by the model include operator workload, situation awareness, completed tasks, delayed or omitted tasks, and likelihood of correct decision. They used the model to compare the effectiveness of two different configurations of personnel in vehicles in a tactical environment. Their findings revealed the influence of organizational structure on information flow (indicated by interrupted communications and dropped tasks). The authors extended the model to the operational level by modifying its architecture to represent complex organizational structures, e.g., individuals belonging to multiple groups. Their modeling of collaborative intelligence-gathering identified individuals with high workloads as bottlenecks in the information flow. The data aided in developing and testing flexible, adaptive workload-management strategies for the modeled activity.

Mitchell et al., (2003) used a task network model to compare and predict workload for two- versus three-soldier crews. They varied the distribution of functions among crew members, and the types of scenarios crews will encounter. The model calculated that a two-soldier crew will experience excessive workload multiple times during combat scenarios. Based on these findings, Mitchell et al., recommended using three-person crews.

Hautamaki et al., (2006) also employed a task network model of a submarine combat system to predict whether changes to the design of the operator interfaces would improve the system's performance. They compared mission performance under the existing configuration to that with a system designed to detect and alert operators to errors and hazards, wherein the intrusiveness of the alerts increased with the severity of the consequences. They found that an alerting system would be a useful enhancement.

Wickens et al., (2007) used a two-part model of situation awareness to compare the effectiveness of advanced flight-deck displays. The model's first part represented selective

[4] For more information on the range of IMPRINT and HARDMAN uses, see http://www.arl.army.mil/www/default.cfm?Action=445&Page=447, retrieved 19 October 2009

attention with four parameters: salience, effort, expectancy, and value. This part of the model described how people scan displayed information. The authors validated their model using scanning measures in simulations of operating an aircraft and a motor vehicle. The model's second part represented the operator's awareness of the information, incorporating a memory decay function that was refreshed by scanning; the model described how scanned information updates understanding. The two-part model revealed differences in situation awareness between alternative displays.

1.2.3 Use of HPMs in the Nuclear Industry

Researchers and plant designers have used HPMs in the nuclear industry, a primary application being to investigate staffing issues (Laughery & Persensky, 1994; Laughery, Plott, Engh, & Nash, 1996; Lawless, Laughery, & Persensky, 1995).

The NRC examined the feasibility of using task network models to model the operator's performance in several highly proceduralized scenarios, including a loss of coolant accident (LOCA), steam generator tube rupture (SGTR), and load maneuver (Laughery & Persensky, 1994; Lawless, Laughery, & Persensky, 1995). The authors in both studies used the procedures to develop models, and compared their predictions to actual human-performance data from simulator trials. Crews handled the events using either paper or computer-based procedures. The authors concluded the model's predictions "…were representative of actual performance," but the results "…were not good enough to declare a clear success of the modeling approach."

Sebok, Hallbert, Plott, and Nash (1997) evaluated how well the models incorporate cognitive- and decision-making activities. They developed HPMs to predict performance in the following scenarios: SGTR, LOCA, loss of off-site power, and interfacing systems LOCA. The operators' responses to the first three scenarios are significantly determined by procedures, i.e., "rule-based." Their reactions to the interfacing systems LOCA are less-well defined and more cognitively demanding, i.e., "knowledge based." The HPM's prediction of task times and the crews' diagnoses was compared with data from simulator trials with operators. The HPM's predictions were favorable in the rule-based scenarios, but not as good for the knowledge-based scenario.

Sebok and Plott (2008) used models to predict the effects of fewer staff in the control room, and the effects of varying crew size on workload and event timing. The authors focused on two scenarios used in a previous simulation with operators, viz., an SGTR and loss of offsite power, that compared normal versus minimal staffing. They adopted operators' descriptions of the tasks that a four-person crew must undertake in each scenario and the decision rules operators use in performing them. The operators also had estimated task times, task variability, and workload. Sebok and Plott's model predictions agreed well with the human-performance data for task-time measures. Running the model 100 times allowed them to specify points in each scenario at which significant delays (sufficient to impact safety) might occur when the crew comprised three members, rather than four. Agreement between workload predictions and the operator's workload measures was not as good. The authors postulated that this discrepancy might reflect differences in how the model predicted workload and how the operators' evaluated it.

In another study, Yow et al., (2005) compared HPM predictions of the crew's response to plant disturbances to that of actual crews responding to them in a plant simulator. The performance measures were task times and workload. There was a good correlation between the results.

The authors concluded that the model's performance was sufficiently good to support extrapolation of the findings to untested conditions.

In an earlier study with the same simulator data, Walters and Yow (2000) had developed a model to predict situation awareness. During pauses in the simulation, the operators were asked to recall whether selected process parameters were increasing, decreasing, or unchanging. The model's predictions did not differ significantly from the operators' recollections.

Overall, these results suggest that performance models are useful in predicting the performance of operators on procedural tasks, e.g., the models highlighted some tasks that might be delayed after a reduction in crew size. The models also could predict the general trend and magnitude of the operators' workload during the scenarios.

HPMs now are being employed by the nuclear industry in a variety of applications, and also to support plant design. Hugo (2006) used an HPM to evaluate event timing and error rate as part of the Pebble Bed Modular Reactor (PBMR) HFE program. PBMR designers also employed the HPM to evaluate the effects of performance-shaping factors and workload on task performance. As the sophistication of the HPMs improve, their application likely will be extended to more complex designs and evaluations.

1.2.4 Potential Benefits of Adopting HPMs for Engineering Design and Evaluation

After reviewing the application of human performance models in general as part of a National Academy of Sciences study, Baron et al., (1990, p. 86) concluded:

> Given the current state of the art in human performance modeling, is the methodology ready to be an integral part of the system design process? Although the methodology has a number of admitted weaknesses, it also has the ability to make a number of unique contributions to the process of system engineering.
>
> By beginning modeling efforts early in the design process, a formal means is provided for considering the impact of human performance capacities and limitations on the range of design issues that must be confronted while there is still time to resolve them. An early modeling effort can provide quantitative and qualitative analyses that allow design trade-off studies to include a variety of human performance factors along with other system variables. This process forces consideration of the assumptions and design decisions which underlie assertions that the system will work with available personnel.
>
> In all, there are compelling reasons to believe that systematic human performance modeling efforts should be regularly advocated and used along with expert judgment and manned part- and full-task simulation, as a regular part of the design process for large-scale human-machine systems.

Building an HPM requires the analyst to generate detailed evaluations of the tasks that people perform and to identify the links between activities inside and outside the control room, and the events in scenarios. Tasks are modeled at varying levels of detail, thus ensuring more precise modeling of the more important tasks.

HPMs can be run numerous times, while varying task characteristics and situational factors, to determine the effects on performance. Models also provide analysts with a tool for "sensitivity analyses;" i.e., to identify which factors most significantly affect human performance.

HPMs offer documentation and traceability, so that independent analysts can review the model and repeat analyses. Further, the HPM constitutes a lasting tool for testing future modifications to the plant.

The main benefits of modeling fall into the following areas:

- clear depiction of task flows, dependencies, and interrelationships
- imposition of detail and rigor on the analysis
- incorporation of the influence of many factors on performance
- ease of performing "sensitivity analyses" i.e., the sensitivity of human performance measures to changes in other factors, such as workload and staffing levels
- affords documentation and traceability

HFE professionals may use HPMs because of the difficulties in collecting human-performance data from actual operators:

- The availability of operators is very limited and often restricted to a very short time.
- Specialized facilities, such as simulators, are often needed for collecting data with operators; similarly, these facilities have very limited availability.
- The amount of data generated from testing operators in simulators often circumscribed because only a small set of scenarios are used.
- The cost of conducting such simulations is considerable, especially in view of the moderate amount of data produced.

Thus, HPMs afford a potentially satisfactory alternative to assess human performance under varying conditions.

1.2.5 Potential Limitations of Using HPMs for Engineering Design and Evaluation

Although HPMs offer potential benefits in research and design applications, there are limitations that both users and reviewers of models should consider carefully.

HPMs are an abstract representation of human behavior; an exact match is unlikely (DoD, 1999a). I gave examples of this limitation in the studies described earlier. Thus, while HPMs were quite good at predicting some particular aspects of performance, they were not as good at predicting other aspects.

As the models become more complex and incorporate elements designed to address various aspects of human performance, such as visual perception, decision-making, memory, and workload, I expect they may become more predictive. However, they may become more susceptible to the effects of modeling errors, partly because errors are additive (Topcu, 2003). Thus, even though each element of a model may be accurate, the model as a whole may not be. Integrating individual elements that form the complete model may not produce meaningful results because the effects of the errors in all of them may be summed.

Misuse is another potential problem, such as when an HPM that predicts performance in one context is applied to an entirely different one. Each HPM has an intended use. As I found in reviewing the studies described earlier, an HPM may predict behavior well in a narrow range of application, such as a particular scenario or staffing profile, but may be quite inappropriate to use for another.

When an analyst fails to consider these limitations, the model's human-performance predictions may be inaccurate or incorrect, and hence, using them as the basis for nuclear-plant design, e.g., for task allocation, task design, staffing, and HSI design, may jeopardize the safety of plant operation.

To be of value in engineering design and evaluation, models must be valid representations of human behavior, especially in safety-critical domains, such as nuclear power. As Campbell and Bolton (2005) noted, "…it is generally agreed that validation is tremendously important, and the risk of drawing erroneous conclusions from unvalidated models is unacceptable" (p. 365).

When HPMs are used to support NPP design, the NRC's staff should review them to ensure that they are appropriately validated and properly used.

1.3 Objectives

The objective of this research was to develop guidance to support the NRC's staff reviews of an applicant's use of HPMs as part of an HFE program.

1.4 Report Organization

The report is divided into two parts. Part one describes the research objectives, methodology, and technical basis to support the development of guidance. In addition to this introduction, there are five more sections. Section 2 describes the study's methodology and Section 3 provides the results of my assessment of the potential use by applicants of HPMs in their HFE programs. Section 4 describes the technical basis used for developing HPM review guidance. During guidance development, I identified several topics related to HPMs that can be addressed with additional research; they are discussed in Section 5.

Part 2 of the report contains the review guidance, and information intended to support applying it in regulatory reviews. Section 6 provides an overview to the guidance and an HPM characterization. Section 7, 8, and 9 provides review guidance for HPM verification, HPM validation, and user interface verification respectively.

References to cited works appear in Section 10. A glossary of related terms is included following the references. An Appendix details the recommended contents of an applicant's HPM submittal.

2 METHODOLOGY

Figure 2-1 is an overview of the methodology used in developing guidance to address HFE issues (O'Hara et al., 2008a). It was steered by the objective of establishing the validity of the guidelines. Validity is defined along two dimensions: internal and external validity. The former is the degree to which the individual guidelines are linked to a clear, well founded, and traceable technical basis. External validity is the degree to which independent peer review supports the guidelines. Peer review is a good method of screening guidelines for conformance to generally accepted HFE practices, and to industry-specific considerations, i.e., for ensuring that they are appropriate based on practical operational experience in actual systems.

Of the four steps shown, in this document, I primarily address the first two. The last two steps will be performed in future.

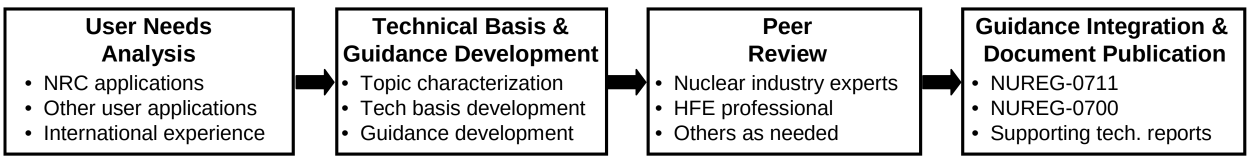

Figure 2-1 Major steps in developing the NRC's HFE guidance

<u>User Needs Analysis</u>

We partly analyzed the first step of the methodology, *User Needs Analysis*, earlier (O'Hara et al., 2008a, & 2008b). The HPM applications were summarized briefly in Section 1 of this report. To supplement our understanding of the needs of the NRC's reviewers, we analyzed the potential uses of HPMs in the HFE programs that the NRC's staff may have to review. Their review process is governed by Chapter 18 of the *Standard Review Plan* (SRP, NRC, 2007) that refers to NUREG-0711 for detailed review criteria for an HFE program. Thus, we used *Human Factors Engineering Program Review Model* (NUREG-0711; O'Hara et al., 2004) to determine where a reviewer might have a criterion for which the information supplied by an applicant is based on an HPM rather than on traditional techniques of analysis or data collection. In addition, we evaluated some unique aspects of the reviews of changes to important operator actions as addressed in the review criteria of *Guidance for the Review of Operator Actions* (NUREG-1764; Higgins et al., 2007), Section 3 describes the results of the analysis.

<u>Technical Basis and Guidance Development</u>

The second major step, *Technical Basis and Guidance Development*, involves: topic characterization, development of technical basis, development of guidance, and documentation.

Topic Characterization

A topic is an HFE issue or group of them for which design-review guidance is being formulated. The first step here is to develop a topic characterization identifying those areas where guidance is needed.

Technical Basis Development

The next step is to analyze information about the topic to generate the technical basis upon which guidance can be developed and justified. Figure 2-2 illustrates the use of several sources of information in order of preference for this task. Following the flow chart downwards, the sources of technical basis change in three ways. First, the information sources near the top are already in the HFE guidance format, or close to it. Towards the bottom, individual research studies must be synthesized and HFE guidelines abstracted. Second, the information at the top already possesses a degree of validity (as discussed earlier), while towards the bottom the validity of the data must completely be established. Third, using the information at the bottom of the flow chart for developing guidance generally is more costly; thus, the preference is to use sources higher in the figure.

Existing HFE standards and guidance documents are considered first, for example, the standards developed by the U.S. military and organizations such as the American National Standards Institute (ANSI). The authors of such publications have established HFE guidelines using research data, operational experience, and their own knowledge/expertise. In addition, most existing standards and guidance documents have been peer-reviewed. Thus, the documents have internal validity or external validity, or both. Since the information already is in guideline form, generally it is easier to use than information from other sources. Much of the technical-basis information used to develop HPM review guidance comes from this type of information. For example, the Department of Defense's (DoD's) Defense Modeling and Simulation Office (DMSO) offers an extensive, online collection of recommended-practices guides, as discussed in Section 4 of this report.

From this invaluable starting place, I utilized additional sources of information. I next sought documents providing good analyses and syntheses of existing literature, such as handbooks. An excellent example is Gluck and Pew's (2005) text, *Modeling human behavior with integrated cognitive architectures: Comparison, evaluation, and validation.* These documents are valuable because they constitute reviews of research and operational literature by knowledgeable experts. However, the information usually is not expressed in guidance form, so that guidance must be developed from it.

When those sources discussed above are insufficient to support the development of guidance, the basic literature is explored, such as papers from research journals and technical conferences. However, greater effort is required to translate such information into guidance than is involved in using the sources described earlier.

For the formulation of HPM review guidance, I employed information from these first three sources of information (see Section 4).

Industry experience and original research also can serve in developing a technical basis, although I used neither in this research.

When I noted topics for which the technical basis was inadequate to support guidance development, I identified unresolved issues, the topics reflecting these issues can be addressed in the future (see Section 5).

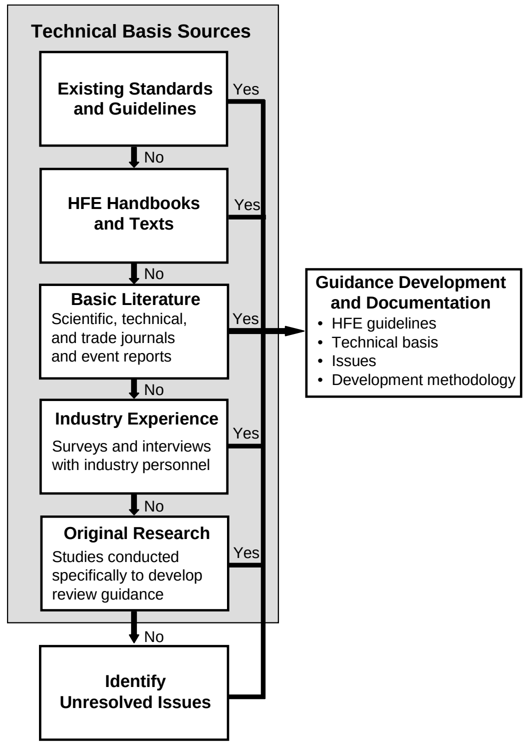

Figure 2-2 Technical-basis development, illustrating the sequential use of sources of data.

Guidance Development

Having completed the steps described above, review guidelines are developed based on the comprehensive technical basis. This guidance is documented in Part 2 of this report.

Peer Review

Subject-matter experts review the results of guidance development. These reviewers include personnel from the NRC with expertise in human factors engineering and engineering fields directly related to the topic.

After NRC comments are resolved, the document is ready for independent review; this document will undergo this process in the future. Independent HFE subject-matter experts undertake this outside review, along with experts in other technical domains based on the subject of the guidance. The review also covers evaluations of the topic characterizations. These reviewers evaluate the scope, comprehensiveness, technical content, technical basis (adequacy of its internal validity), and usability (guidance organization presentation) of the report. Their ensuing comments and recommendations are used in revising the guidance.

Guidance Integration and Document Publication

After the reviews are completed and the comments resolved, the document is ready for publication as a technical basis report. The guidance then is integrated into NUREG-0711, NUREG-0700, or other applicable NRC design review documents. These NUREGs provide only the information necessary for conducting HFE reviews. Reference is made to the technical report describing the technical basis.

3 POTENTIAL USE OF HUMAN PERFORMANCE MODELS IN APPLICANT SUBMITTALS

In this section, we examine the potential use of HPMs to analyze human performance by various applicants, viz., for license amendments, design certifications, and combined operating licenses. The NRC's staff will review their submittals relying on HPM-related information. Two areas of HFE reviews are discussed:

- Applicant's submittals for new plants (SRP Section II.A), and modifying control rooms (SRP Section II.B), reviewed using NUREG-0711
- Requests for licensing amendment requests involving credited and risk-significant human actions (SRP Section II.C), reviewed using NUREG-1764

3.1 Submittals for New Plants and Modifying Control Rooms

The SRP provides high-level guidance for conducting HFE reviews in Chapter 18, Human Factors Engineering (NRC, 2007a). New plant reviews are addressed in SRP Section II.A, and control room modifications (SRP Section II.B), which are reviewed using NUREG-0711

The *Human Factors Engineering Program Review Model* (NUREG-0711) details the criteria for review (O'Hara et al., 2004). The approach rests on the concept that the HFE aspects of NPPs should be developed, designed, and evaluated via a structured systems analysis, using accepted HFE principles, at the same time as other systems are undergoing design. The review encompasses the twelve elements shown in Table 3-1.

Based on our assessment, in seven of these twelve elements, HPMs potentially might be relevant in reviewing applicant submittals (see Table 3-1). The five review elements where HPMs are unlikely to be used include:

- HFE Program Management – this element addresses management aspects of the HFE program, including the team, its plans and procedures, and the tracking system for HFE issues
- Operating Experience Review – this element covers the analysis of operating experience related to the applicant's new plant design
- Training Program Development – this element is concerned with designing and implementing personnel-training programs
- Design Implementation – this element details verifying that the as-built design conforms to the validated one
- Human Performance Monitoring – this element discusses monitoring personnel performance in the early stages of new design operations

In the remainder of this section, I examine the specific applicability of HPMs to these elements.

Table 3-1 Potential HPM Use by Applicants in HFE Programs

Review Element	Implications
HFE Program Management	No
Operating Experience Review	No
Functional Requirements Analysis & Allocation	Yes
Task Analysis	Yes
Staffing and Qualifications	Yes
Human Reliability Analysis	Yes
Human-system Interface Design	Yes
Procedure Design	Yes
Training Program Development	No
Human Factors Verification and Validation	Yes
Design Implementation	No
Human Performance Monitoring	No

3.1.1 Functional Requirements Analysis and Function Allocation

One objective of this review element is to verify that functions are allocated between humans and automatic systems to assure that the role for plant personnel is acceptable; i.e., the allocations take advantage of human strengths and avoid functions that would be degraded human limitations; for example, allocating to operators a control action that must be completed more quickly than the time in which they can respond.

NUREG-0711, Section 4.4 has the review criteria for analyzing functional requirements and allocating functions. HPMs may be applicable to three specific review criteria, as follows:

<u>Criterion 6</u>

> *The technical basis for all function allocations should be documented; including the allocation criteria, rationale, and analyses method. The technical basis for functional allocation can be any one or combination of the evaluation factors (see Fig 4.1). For example, the performance demands to successfully achieve the function, such as degree of sensitivity needed, precision, time, or frequency of response, may be so stringent that it would be difficult or error prone for personnel to accomplish. This would establish a basis for automation (assuming acceptability of other factors, such as technical feasibility or cost).*

This criterion designates successful operating experience as an acceptable basis for allocating function, however, it may not be available, so alternative technical bases must be provided. HPMs might be a means to evaluate the technical basis for function allocations; they can be run to identify performance sensitivity and time. They also can assess the number of personnel best required to operate a system, one of their commonest applications, as discussed in Section 1.

Criterion 8

The allocation analysis should consider not only the primary allocations to personnel, but also their responsibilities to monitor automatic functions and to assume manual control in the event of an automatic system failure.

The demands from the proposed allocation of functions should be considered in terms of all other functions that might impose concurrent demands upon people. Modeling provides a means of exploring the effects of multiple responsibilities (concurrent demands), and their net impact of the workload on task performance. In fact, Leiden concluded from a review of human performance-modeling techniques, "…the most useful stand-alone application of task network modeling for error prediction is the context of multi-tasking activities and high operator mental workload" (quoted by Plott, Engh, & Barnes, 2004). Hence, applying human-performance modeling might satisfactorily enable managers to meet this criterion.

Criterion 10

The functional requirements analysis and function allocation should be verified:

- *all the high-level functions[5] necessary for the achievement of safe operation are identified.*
- *all requirements of each high-level function are identified.*
- *the allocations of functions result in a coherent role for plant personnel*

HPMs might offer a method to verify that the allocations of functions entail a coherent role[6] for plant personnel, in so far as modeling can determine that all allocated functions can be undertaken, and do not interact such that performing one task degrades the performance of another.

3.1.2 Task Analysis

The objective of this review element is to verify that the applicant's task analysis identifies the specific tasks needed to accomplish personnel functions and the informational-, control-, and task-support-requirements (such as personnel protection or tools) for each.

HPMs may be applicable to the following five review criteria:

Criterion 1

The scope of the task analysis should include:

- *selected representative and important tasks from the areas of operations, maintenance, test, inspection, and surveillance*

[5] In this context, a “function” is a process or activity required to achieve a desired goal. For example, safety functions, such as reactivity control, prevent or mitigate the consequences of postulated accidents that could damage the plant or unduely risk the public’s health and safety. .

[6] In NUREG-0711, a “coherent role” signifies a meaningful role for plant personnel that is accomplished by performing well-defined tasks. In contrast, allocating a function strategy based only on technological considerations (e.g., allocate everything possible, and let personnel do the rest) can result in unclear roles consisting of an *ad hoc* assortment of tasks to be performed.

- *full range of plant operating modes, including startup, normal operations, abnormal and emergency operations, transient conditions, and low-power and shutdown conditions*
- *human actions that have been found to affect plant risk by means of probabilistic risk analysis (PRA) importance and sensitivity analyses should also be considered risk-important. Internal and external initiating events and actions affecting the PRA Level I and II analyses should be considered when identifying risk-important actions*
- *where critical functions are automated, the analyses should consider all human tasks including monitoring of the automated system and execution of backup actions if the system fails.*

HPMs provide a basis for an analysis to meet this criterion. Because many HPMs can be linked with other models, such as plant models, task simulations can be run under a wide range of conditions. The HPM can evaluate situational factors (e.g., examining performance during overnight shifts), and performance-shaping factors (e.g., fatigue or high workload) that affect peoples' performance.

Criterion 2

> *Tasks should be linked using a technique such as operational sequence diagrams. Task analyses should begin on a gross level and involve the development of detailed narrative descriptions of what personnel have to do. The analyses should define the nature of the input, process, and output needed by and of personnel. Detailed task descriptions should address (as appropriate) the topics listed in Table 5.1 (of NUREG-0711).*

In a task-network model, human performance essentially is modeled as a series of individual tasks (Figure 3-1). The network defines the interrelationships amongst the tasks. Two potential advantages for the NRC reviewer of an applicant's use of modeling are documentation and traceability. The modeling process forces the analyst precisely to define tasks and their relationships. Thereafter, the model provides documentation of the tasks and their structure that the NRC staff can evaluate.

Criterion 3

> *Task analysis should be iterative and become progressively more detailed over the design cycle. It should be detailed enough to identify information and control requirements to enable specification of detailed requirements for alarms, displays, data processing, and controls for human task accomplishment.*

As noted above, modeling typically requires a detailed task analysis. Reasonably, the outputs of modeling are expected to be part of an iterative task-design process, since repeated modification and evaluation easily is accomplished with HPMs.

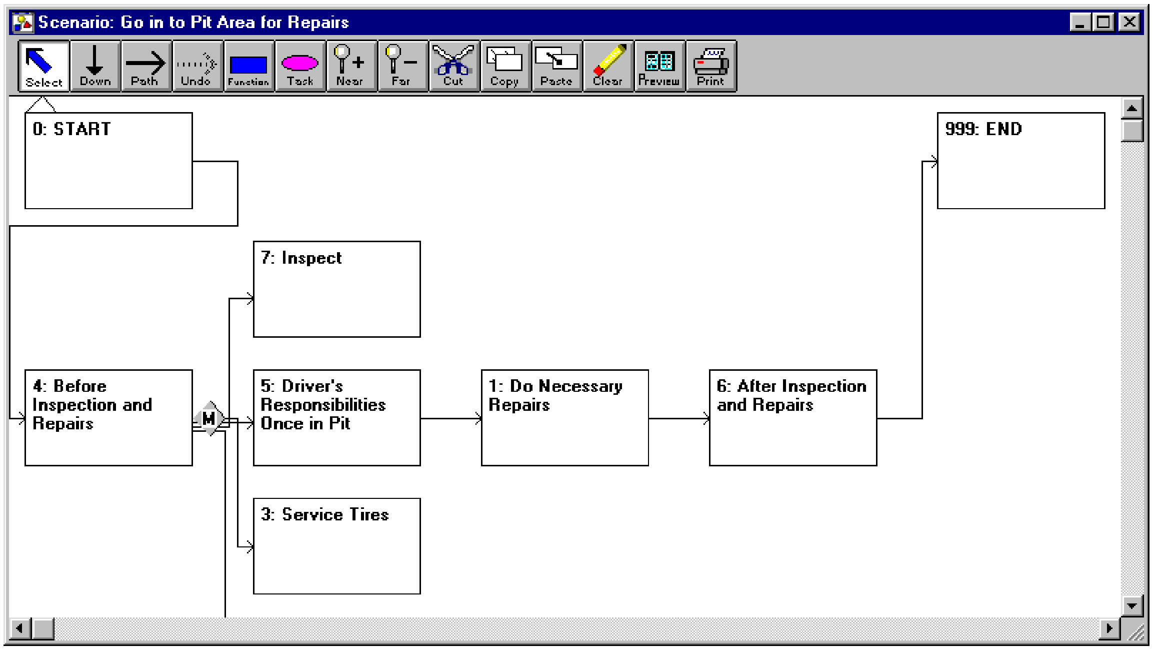

Figure 3-1 Task network model of a series of individual tasks
(Source is NRC, 1999)

<u>Criterion 4</u>

Task analysis should address issues such as:

- *the number of crew members*
- *crew member skills*
- *allocation of monitoring and control tasks to the (a) formation of a meaningful job and (b) management of crew member's physical and cognitive workload.*

Human-performance models support analyses that address the demands of multi-task situations. As noted in Section 3.1.1, modeling is well suited to questions about assigning tasks and workload.

<u>Criterion 6</u>

Task analysis should identify reasonable or credible, potential errors.

HPMs can aid in identifying human errors. A model can highlight the conditions under which personnel fail to perform a task (errors of omission). For example, they may link the failure to perform a task to conditions such as high fatigue, conflicting demands, or high workload. Some models (e.g., IMPRINT) also can identify situations wherein operators are likely to postpone or interrupt tasks as a workload-management strategy. The models also may specify an error of commission when the simulated human performance leads to an alternative action with an unwanted consequence.

3.1.3 Staffing and Qualifications

The objective of this review element is to verify that the applicant systematically has analyzed the need for the number and qualifications of personnel, and has demonstrated a thorough understanding of task- and regulatory-requirements.

HPMs might apply to two specific review criteria.

<u>Criterion 2</u>

> *The staffing analysis should determine the number and background of personnel for the full range of plant conditions and tasks including operational tasks (normal, abnormal, and emergency), plant maintenance, and plant surveillance and testing.*

Modeling is recognized as particularly useful in defining staffing requirements. The relative ease of evaluating a range of plant conditions is an important feature of many modeling approaches. (See the discussion of Criterion 3 in Section 3.1.6.3.)

One outcome of an applicant's analysis may be that fewer staff are needed than the NRC-specified requirements identified in *10 CFR 50.54(m)*. In such cases, applicants can apply for an exemption to the staffing requirements. Modeling was identified as a means of evaluating exemption requests from 10 CFR 50.54(m) (Plott, Engh, & Barnes, 2004; Persensky, Szabo, Plott, Engh & Barnes, 2005). NUREG-1791 offers guidance to the NRC staff for reviewing requests for exemptions from current control room staffing requirements in *Guidance for Assessing Exemption Requests from the Nuclear Power Plant Licensed Operator Staffing Requirements Specified in 10 CFR 50.54(m)*.

In validating the staffing plan, HPMs are identified as a means of providing data or demonstrations that the control room personnel identified therein satisfy performance requirements. HPMs are only one of several methods and data sources identified in NUREG-1791. Discussing using HPMs as part of the applicant's process, the authors state

> Data from human performance models can provide a robust representation of the performance of control personnel across the range of operational conditions. Models can easily incorporate the various conditions that may affect human performance, human performance variability, and measures of concepts, such as cognitive workload and situation awareness. Although human performance models historically have incorporated plant or system representations of limited fidelity, human performance models can now be linked to more sophisticated plant or system simulations. The human performance models also make it relatively easy to assess different staffing alternatives. As with the human-in-the-loop simulations, quantitative, objective measures and criteria can be captured to support the exemption request. (p. II-10-4)

NUREG-1791, Section 10.3.3, Data Sources and Demonstration Methods, gives the review criteria, stating:

> The reviewer should confirm that the following criteria have been met, as applicable:
>
> - The selected design of the staffing plan validation, the data sources, and the demonstration methods comprehensively address the dynamic aspects of the staffing plan and support the requested exemption.
> - The data sources and demonstration methods were used appropriately.
> - The appropriate quantitative, objective measures and criteria were defined and captured.

- The data collection and analysis were conducted appropriately.
- The scope and data quality were adequate.
- The outcomes were reasonable and valid. (p. II-10-8)

These criteria are general and applicable to any of the data sources and demonstration methods identified; that is, they are not specific to HPMs. Because the guidance is very general, an NRC reviewer has little basis to determine whether an HPM has met these criteria. Additional guidance is needed to support the NRC staff's evaluation of an applicant's use of HPMs for HFE analyses.

Criterion 4

> *The staffing analysis should be modified to address a broad range of HFE issues.*

NUREG-0711 lists a broad range of issues that applicants should address, including the following:

- personnel response time and workload
- availability of personnel considering other activities that may be ongoing and for which operators may take on responsibilities outside the control room (e.g., fire brigade)
- the effect of overall staffing levels and crew coordination for risk-important human actions

HPMs often are employed to examine task timing and workload, and the effects of multi-tasking on performance. As I noted above, HPMs are suitable for evaluating staffing levels and used as such for nuclear plants and other complex systems.

3.1.4 Human Reliability Analysis

The objectives of this review element are to verify that (1) the applicant addressed human-error mechanisms in designing HFE aspects of the plant to minimize the likelihood of personnel error, to support operators to detect errors when they occur, and to recover from them; and (2) the human reliability analysis (HRA) activity effectively integrates the HFE program, the PRA, and risk analysis. (I note that this area of HFE review does address conducting an HRA. Section 3.2 discusses some applications of HPM to the HRA.)

HPMs may be applicable to one specific review criterion.

Criterion 4

> *HRA assumptions such as decision making and diagnosis strategies for dominant sequences should be validated by walkthrough analyses with personnel with operational experience using a plant-specific control room mockup or simulator. Reviews should be conducted before the final quantification stage of the PRA.*

HPMs provide a means to test HRA assumptions, for example, that the performance of a risk-important action can occur within a time window defined by thermodynamic analysis. An HPM can be formulated to model the action and obtain the predicted time under a variety of credible scenarios.

3.1.5 HSI Design

The objectives of reviewing the HSI design are to verify that the processes by which the requirements for HSI is designed are developed properly, and the HSI is identified and refined. The review assures that the applicant appropriately translated functional- and task-requirements to the detailed design of alarms, displays, controls, and other aspects of the HSI by systematically applying HFE principles and criteria.

Here, HPM especially is applicable in the test and evaluation of HSI designs. NUREG-0711 states that such evaluations should be conducted iteratively throughout the HSI development process. Section 8.4.6.2, of that document, Performance-based Tests, gives the review criteria for evaluating the HSIs. Because HPMs apply to the entire section, the discussion is general, rather than at the level of individual criteria.

HPMs potentially can explore the effects of tradeoffs in HSI design on human performance. I discussed an example from the nuclear industry in Section 1.2.3, wherein researchers compared the crew's performance with paper vs. computer-based procedures (Laughery & Persensky, 1994; Lawless, Laughery, & Persensky, 1995). Similarly, task-completion time and workload can reveal the relative advantages of each method of presentation. .

3.1.6 Procedure Development

The objective of reviewing procedure development is to verify that the applicant has applied HFE principles and guidance, and all other design requirements, to generate procedures that are technically accurate, comprehensive, explicit, easy to use, and validated. HPMs are applicable here to two criteria.

Criterion 6

All procedures should be verified and validated, including:

- *A review should be conducted to verify they are correct and can be carried out.*

- *When procedures are modified, they should be verified to verify their adequate content, format, and integration. The procedures also should be assessed through validation if a modification substantially changes personnel tasks that are significant to plant safety. The validation should verify that the procedures correctly reflect the characteristics of the modified plant and can be carried out effectively to restore the plant.*

HPMs, such as task-network models, can determine if a procedure can be finished within the time requirements. They can examine the outcome of allocating tasks to crew members, and the effects of performance-shaping factors that might affect procedure performance. As discussed earlier, modeling methods are very acceptable for evaluating design modifications because of the ease of comparing alternative solutions.

Criterion 7

An analysis should be conducted to determine the impact of providing computer-based procedures (CBPs) and to specify where such an approach would improve procedure utilization and reduce operating crew errors related to procedure use. The justifiable use of

CBPs over paper procedures should be documented. An analysis of alternatives in the event of loss of CBPs should be performed and documented.

As discussed under HSI design, HPMs support comparisons of the crew's performance with CBPs and paper procedures.

3.1.7 Human Factors Verification and Validation

For V&V, HPMs might be useful in integrated system validation (ISV). The ISV's objective is to verify that the applicant validated the integrated system design (i.e., hardware, software, and personnel elements) via performance-based tests to determine whether it acceptably supports the plant's safe operation. NUREG-0711 Section 11.4.3.2, Integrated System Validation gives the review criteria. Because HPMs apply to the entire section, the discussion is a general one.

Designers might use HPMs to generate data demonstrating that an integrated system yields "human performance" that acceptably meets established criteria. However, because validation aims to confirm acceptable performance after unifying all constituent elements (trained personnel, HSIs, procedures, and plant model), modeling may have some general limitations and pose some concerns. It is not clear that HPMs are sufficiently accurate to meet the data requirements of validation. While those models simulate human performance, and a plant model represents the behavior of the plant (both in response to thermodynamic evolution and to human actions), they will yield only approximations of human performance. Although this may be acceptable for many of the HPM applications discussed earlier, it may not be for validation, where the intent is to generalize to actual plant operations. Thus, the demands for fidelity in the model are very high for this application.

Furthermore, should modeling be used to support the licensee's design and development work, the ensuing evaluation might lack "diversity and defense-in-depth". That is, if the models afford little opportunity of catching inherent errors or biases, there may be concerns that erroneous conclusions are reached about the design's acceptability.

With these potential limitations, hybrid validations may be of value, i.e., some data provided through modeling and other data obtained from testing crews. Using this approach, HPMs can be run repeatedly to examine the range of human performance while varying individual-, situational-, and performance-shaping factors.

3.2 License Amendment Requests Involving Credited and Risk-significant Human Actions

The SRP, Section II.C, provides high-level guidance for conducting HFE reviews of license-amendment requests involving credited and risk-significant human actions (HAs), especially those requiring changes in the plant's licensing basis, e.g., manually, rather than automatically operating the plant's safety systems. The SRP refers the NRC's staff to NUREG-1764, *Guidance for the Review of Changes to Human Action* (Higgins et al., 2007), for the detailed risk-informed review guidance and acceptance criteria to be used.

The evaluation method follows two-phase approach. First, a screening analysis of the licensee’s proposed modification and the affected HAs assesses their risk importance, and a graded, risk-informed approach determines the appropriate level of HFE review. This methodology is applicable to both risk-informed and non-risk-informed submittals. For the former, the first phase has four steps:

1. use of Regulatory Guide 1.174 (NRC, 1998) to determine the risk importance to the entire plant of a change or modification that involves the HA
2. quantify the risk importance of the HA itself
3. qualitatively evaluate the HA
4. undertake an integrated assessment to determine the appropriate level of the HFE review

Thereafter, the HAs are assigned a risk level, either high, medium, or low.

In the second phase, HFE review guidance is used to review the HAs, and verify that the proposed action can be reliably performed. The level of this human-factors review corresponds to the risk levels determined in the first phase. HAs in the high-risk level receive a detailed HFE review, and those of medium risk undergo a less detailed one. For human actions placed in the low-risk level, there is no HFE review or a minimal one. The review criteria in the second phase are tailored versions of the NUREG-0711 criteria discussed above, so the same HPM considerations apply.

For the evaluations undertaken during the first phase, HPMs may be germane to estimating human-error probabilities, and to qualitatively assessing the safety significance of human actions, as considered below.

In Phase 1, Step 3 of the screening process, HA safety significance is assessed qualitatively; the findings are used to adjust the level of the HFE review.

Three types of qualitative assessment are used. The first category, addressing "Personnel Functions and Tasks" includes:

- Operating experience
- New actions
- Change in automation
- Change in tasks
- Change in performance context

The second category, "Design Support for Task Performance" covers:

- Change in HSIs
- Change in procedures
- Change in training

The third category, "Performance Shaping Factors" encompasses:

- Change in teamwork
- Change in skill level of individuals performing the action

- Change in communication demands
- Change in environmental conditions

HPMs can contribute to evaluating all three, by providing a tool to examine changes to the plant and task design on human performance. Thus, they support a way to examine the impacts on performance of changes in automation, task design, HSIs procedures, communication demands, environmental conditions, and other factors.

3.3 Conclusions

In this section, I considered the potential use by applicants of HPMs to undertake human-performance analyses as part of their HFE programs. I found that HPMs have broad applicability within such programs and might well support applicants in allocating functions, analyzing tasks and staffing requirements, in designing HSIs and procedures, and in performing ISV.

The NRC staff will need review guidance for assuring the validity of the HPM applications models, and the appropriateness of their usage. I describe the development of such guidance in the next section.

4 DEVELOPMENT OF HPM REVIEW GUIDANCE

In this section, I consider the technical basis used to develop the HPM review guidance presented in Part 2 of this report.

HPMs are a class of modeling and simulation (M&S) applications. One reason for the surge in their use within HFE programs reflects the initiative taken by the DoD to incorporate M&S into their acquisition process for new military systems (DoD, 2003b). To ensure that the M&Ss are of sufficient quality to employ in engineering design and evaluation, the DoD requires that they be verified, validated, and accredited (VV&A). The DoD's process for accrediting M&S applications is applied to HPMs. Since NRC applicants use HMP in essentially the same manner as does the DoD, the latter's approach to HPM VV&A is directly applicable. An NRC reviewer needs to ensure that the properties of an applicant's HPM are suitable for using in engineering design and evaluation. Thus, the DoD VV&A process forms the basis for developing review guidance for employing HPMs in NPP applications.

DoD Instruction 5000.61 gives the high-level procedures and documentation requirements for the M&S VV&A, and clearly defines the following key concepts:

- *Accreditation* - The official certification that a model, simulation, or federation[7] of models and simulations and its associated data are acceptable for use for a specific purpose.
- *Model* - A physical, mathematical, or otherwise logical representation of a system, entity, phenomenon, or process.
- *Modeling and Simulation (M&S)* - The use of models and simulations to develop data as a basis for making managerial or technical decisions.
- *Simulation* - A method for implementing a model over time. Also, a technique for testing, analysis, or training in which real-world systems are used, or where real-world and conceptual systems are reproduced by a model.
- *Verification* - The process of determining that a model implementation and its associated data accurately represent the developer's conceptual description and specifications.
- *Validation* - The process of determining the degree to which a model and its associated data are an accurate representation of the real world from the perspective of the intended uses of the model.

Several military organizations originated detailed VV&A guidance for the implementation of DoD Instruction 5000.61, e.g.:

- Army Instruction 5-11 (DoD, 2005) and Army Pamphlet 5-11 (DoD, 1999a)
- Navy Instruction 5200.40 (DoD, 1999b)
- Air Force Instruction 16-1001 (DoD, 1996)

Their documents outline acceptable methods to set about verification, validation, and accreditation activities.

[7] DoD (2003b) defines a federation as "A system of interacting models and/or simulations, with supporting infrastructure, based on a common understanding of the objects portrayed in the system."

Accreditation is the official certification of the acceptability of an M&S. use. The accreditation review considers the M&S development process, configuration management, history of use, and evidence of verification and validation (DoD 1999a). The DoD established templates for use in assembling information for the accreditation process (DoD, 2003; DoD, 1999a; and Graffagnini & Youngblood, 2008, for examples). As part of the process, acceptance criteria are formulated by considering the context in which the M&S will be used. Table 4-1 gives examples of acceptance criteria. Subject matter experts (SMEs) review the documents to assure that the criteria are met.

Table 4-1 Examples of High-level Acceptability Criteria for M&S Accreditation

The levels of force structure and interaction have sufficient fidelity and resolution.
The M&S is suitable for the overall intended use (e.g., training, explanatory, predictive).
The M&S output/results may be used clearly, adequately and appropriately to address the problem.
The CM (configuration management) policy is in effect and responsive to the anticipated needs of the M&S users.
All required data values are well defined and data sources for obtaining accredited data have been identified.
There is availability of baseline scenarios, terrain data, threat data, and weapon performance data for the M&S.
The algorithms, terrain and environment representations are functionally adequate to address the issues.
The clarity, fidelity, complexity and level of detail of the simulated entities are acceptable for its intended usage.
The documentation, user training, and user help are adequate.
The M&S is suitable for the hardware and software platforms on which it will be used.
M&S demonstrate appropriate sensitivity to data perturbations and response at boundary (limiting value) cases.

Source is DoD 1999a, Figure 1-1, p. 44

More recently, the DoD's Defense Modeling and Simulation Office (DMSO) furnished an extensive, online collection of VV&A recommended practice guides.[8] Those most pertinent to evolving review guidance for HPMs in nuclear applications include

- Validation (DoD, 2004)
- Validation of Human Behavior Representation (DoD, 2001a)
- V&V Techniques (DoD, 2001b)
- V&V Tools (DoD, 2000b)

[8] The DMSO recommended practice guides are available at: http://vva.dmso.mil/default.htm; retrieved June 19, 2008. Specific guides used in formulating the guidance are identified in Section 5, References.

- T&E/V&V Checklist (DoD, 2000c)
- Subject Matter Experts and VV&A (DoD, 2000d)
- A Practitioner's Perspective on Simulation Validation (DoD, 2001c)
- V&V Agent's Role in the VV&A of New Simulations (DoD, 2001d)

The DMSO site's additional resources include an extensive glossary of M&S terminology.

The Australian Defense Simulation Office (2005) and the North Atlantic Treaty Organization (2008) recently published their M&S VV&A guides, based on the DoD process. In addition, the Institute of Electrical and Electronics Engineers published a standard on the VV&A of distributed interactive simulation systems[9] that was reaffirmed in 2002 (IEEE, 1977).

These documents apply to any type of M&S. However, it was noted that validating HPMs can be more difficult than M&Ss of physical phenomena. The HMSs are highly complex, with nonlinear relationships between variables, and are chaotic; i.e., small changes to inputs may entail large discrepant outputs (DoD, 2001a; DoD, 2003a; Campbell & Bolton, 2005).

Young (2003) summarized several attempts to validate HPMs. "Face validation" was the commonest approach, namely, the HPM's output was assessed by SMEs who judged its ability to simulate the human behavior of interest. Young concluded that models used for engineering need a fuller validation, a position echoed by others (Campbell & Bolton, 2005; Conwell et al., 2000). Accordingly, more formal approaches to VV&A were developed.

Allender et al., (1995) described a formal VV&A process applied to HARDMAN. HARDMAN is the predecessor of IMPRINT, the HPM described in Section 1.2 of this report. This HARDMAN VV&A effort followed the guidance in Army Pamphlet 5-11. Table 4-2 summarizes those aspects of the HPM that were appraised by a review board comprising representative users, policy-makers, technical experts, and soldiers. Allender concluded, "Based on the results of the verification and validation efforts, the accreditation board granted accreditation to the designated components of HARDMAN III with only limited caveats. Basically, HARDMAN III was found to be value-added over the current way of doing business and to be consistent with the philosophy of using the best available data."

[9] Distributed interactive simulation systems are ones that encompass multiple, interacting simulations models,

Table 4-2 Aspects of HARDMAN Evaluated in the VV&A

Aspect of HARDMAN	Description
Configuration management established	HPM configuration management includes specification of responsible organizations, software change management, distribution and update mechanism, software rights, and responsiveness to users.
Software verification performed correctly	Verification that the software is working properly.
Documentation and user help adequate	Evaluation of documentation to ensure the HPM is fully described and that it supports users to conduct the intended analyses.
Data input requirements defined	Evaluation that the data required to use the HPM are clearly defined.
Level of detail acceptable	The model supports analysis of the level of detail (e.g., mission, function, task) required by users.
Modeling techniques functionally adequate	Validation that the modeling techniques and algorithms can achieve their intended purposes. For HARDMAN this was examined for three aspects: task network modeling, workload estimation, and task performance degradation.
Output results useful to support decision making	Evaluation that the types of HPM results that are produced are presented in such a way to support design decisions.
Analysis feasible within project timelines	Evaluation of the HPM's capability to provide results in a timely manner to support design decisions.

Source is Allender et al., (1995).

I summarized some key points from this literature below:

1. Accreditation is the official certification that an M&S is acceptable for use. Considerations similar to those used in the DoD's process are applicable to determining that an HPM is acceptable for an HFE program. Thus, the accreditation process is a reasonable avenue to considering review guidance for nuclear applications of HPMs by applicants.
2. The process of "accrediting" M&Ss involves judgments based on evaluating the model description, development process, technical basis, verification, and validation.
3. M&Ss should be verified and validated before employing them in engineering design and evaluation. Campbell and Bolton (2005) noted, "…it is generally agreed that validation is tremendously important, and the risk of drawing erroneous conclusions from unvalidated models is unacceptable." (p. 365). Here, risk means that of designing a system that fails to meet human-performance requirements.
4. The validation of M&Ss must incorporate clear reference to a well-defined context of use. They are not validated in general; i.e., they should be used only in the validated context. .
5. Validation essentially is a decision-making process. That decision should rest on a set of process- and outcome-measures (e.g., goodness of fit of the model to measured human performance) that collectively offer converging evidence that the HPM's application is sound.
6. HPMs represent an abstract representation of human behavior; thus, an exact match to human data should not be expected. The DoD (1999a) notes that:

> It must be recognized that M&S are, by definition, abstractions and may not duplicate actual observed phenomena but rather provide an approximation of observed behavior. Therefore, accreditation procedures are the formal process by which the M&S application sponsor gains confidence in the model and simulation for its intended purpose. (p. 40)

7. Because the types of models and their applications differ, no one single approach to V&V suffices. V&V methodologies should to be tailored to the type of model, and its intended use.
8. Whilst validating the individual components of a model is important, it does not validate the fully integrated model. The interactions between the model's elements are an essential aspect of producing the final output.
9. When the opinions of subject matter experts constitute part of model's validation, a structured methodology should support the consistency and reliability of the information obtained.

In summary, the DoD's efforts in developing and applying a VV&A process offers a substantial technical foundation for elaborating review guidance for nuclear applications of HPMs. I employed the DoD's process to develop the review guidelines in the next section. Although many of the documents discussed above contributed, my principal sources of information were:

- Verification, Validation, and Accreditation of Army Models and Simulations (DoD, 1999a)
- The DMSO VV&A recommended practices guides listed above
- HBR Validation: Integrating Lessons Learned From Multiple Academic Disciplines, Applied Communities, and the AMBR[10] Project (Campbell & Bolton, 2005)

The following sources afforded additional information on the VV&A of HPMs applied to engineering design and evaluation applications:

- Defense Modeling and Simulation Office (DMSO): http://vva.dmso.mil; retrieved June 19, 2008
- Navy Modeling and Simulation Office (NMSO) https://nmso.navy.mil/NavyMSOffice/tabid/37/Default.aspx
- Human Factors and Ergonomics Society's Human Performance Modeling Technical Group (http://www.sys.virginia.edu/hfes/hpm/)
- Behavior Representation in Modeling and Simulation (BRIMS) Conference (http://www.sisostds.org/index.php?tg=articles&idx=More&article=20&topics=10)
- Simulation Interoperability Standards Organization's Behavior Representation in Modeling and Simulation (BRIMS) Conference (http://www.sisostds.org/index.php)

The technical basis information summarized in this section was used in creating the HPM review guidance presented in Part 2 of this report.

[10] Agent-Based Modeling and Behavior Representation (AMBR)

5 IDENTIFICATION OF ADDITIONAL TOPICS

During the development of guidance development, I noted several issues that were not addressed easily by the existing technical basis. I describe them below.

Grading and Risk-informing V&V Methods for HPMs

Like many aspects of the NRC's HFE review process, HPM evaluations may be graded according to the risk significance of the application. For example, an HPM intended to evaluate credited human actions should receive more scrutiny than one used to assess local operator actions that are not of high risk. The guidance discussed in this document does not distinguish the two. Additional research could tailor the review criteria to the application's risk importance.

Validating the Interaction between Multiple Models

Integrating multiple models is challenging (Topcu, 2003). In nuclear plants, HPMs are likely to be linked to plant models. The review criteria discussed in this document suggest that the flow of information between them should be validated (see Structural Validation); however, ensuring a complete, comprehensive validation necessitates having more detailed, explicit criteria.

Quantifying Goodness-of-fit

Goodness-of-fit measures are a mathematical measure of the degree to which two data sets agree. Additional guidance is required on quantifying the goodness-of-fit between HPM data and referent data, as well as criteria for judging its acceptability. As noted, an exact match between them is not expected, but guidance on "how good is good enough" will help reviewers.

Reviewing an Applicant's Request to Use a Validated Model Beyond its Intended Application

Applicants may want to extend the application of a verified, validated HPM beyond its context of use. A completely novel application might warrant a full V&V program; however, if it involves only a small extension beyond that validated, this may be unnecessary. Additional guidance is required for defining extensions of use and the associated V&V requirements.

V&V of HPMs Whose Use is Research Not Design

Validation requirements vary with the level of the simulation model's maturity (Harmon & Youngblood, 2005). In the present guidance, I assumed that the models are mature since they are used for engineering design and evaluation. Models also can support research. However, they may be undergoing development, so that applying the current guidance may be inappropriate, nevertheless, they warrant some assessment of appropriate use and validation.

Part 2:

Guidelines for Reviewing Human Performance Models

6 HPM GUIDANCE OVERVIEW

The guidance in this section addresses using HPMs for engineering design and evaluations. It does not consider them in a research context.

An HPM is characterized as a simulation of various aspects of human performance, used to support HFE programs. An HPM is (1) mathematical, programmable, and executable rather than purely explanatory; and (2) applied in the engineering design and evaluation of complex systems. HPMs comprise inputs, elements, and outputs. An element is an aspect of a model that simulates some aspect of human- and team-behavior. For example, an HPM's elements may reflect tasks, numbers of crew members, communication, situation awareness, and workload. Whilst they describe the model's framework, analysts can input data to characterize the elements for a specific application, such as task time or difficulty. The elements interact according to established rules in the HPM. When the HMP runs, the elements interact to generate output data, such as the total time to accomplish a mission, or the workload profile across a mission.

The following are the general objectives of an HPM review:

- Verify that the applicant's HPM works as intended, i.e., that the logic and code correctly perform their intended functions
- Validate that the applicant's HPM produces realistic results
- Verify that the design of the users' interfaces to the HPM support its proper use

The guidance for reviewing an HPM falls into three topical areas: Verification, Validation, and User Interface Verification.

Applicants should identify the use of an HPM in their HFE program and submit a document, hereafter called the "HPM submittal," to the NRC's staff addressing the topics outlined in the Appendix to this report. The staff will review the applicant's submittal using the guidance in this section.

Each guideline is numbered, and contains a criterion that the reviewer will use in evaluating the HPM's acceptability. For many guidelines, there is additional information to assist the reviewer in interpreting or applying the guideline. The information has clarifications, explanations of the technical basis, and gives examples.

I recommend that in addition to HFE experts, the NRC's review team includes I&C and operations experts. Additional expertise may be needed depending on the design of the HPM, other models it interacts with, and its specific application.

7 HPM VERIFICATION

Both logic- and code-verification of the HPMs should be conducted to confirm that they work as intended (DoD, 1999a). Logic verification assures that the M&S algorithms correctly represent the intended processes in relation to the M&S requirements and specifications, while code verification guarantees that the verified logic is implemented properly in code. The basis of the staff's review of HPM verification should be the applicant's methods and acceptance criteria for accomplishing verification.

(1) Logic verification - the applicant should provide evidence that HPM algorithms were successfully verified, ensuring that they correctly represent the intended processes in relation to the model's requirements and specifications.

Additional Information

The DoD (1999a, 2001b, DoD 2000b) identified an extensive list of logic- and code-verifications they developed. . The techniques fall into four categories: Informal, static, dynamic, and formal.

- Informal methods rest on subjective analyses, such as SME reviews (audits) and walkthroughs. Informal methods are the commonest techniques applied.
- Static V&V techniques evaluate the accuracy of the static model's design and source code; they do not require running the HPM. Many automated tools can support these techniques.
- Dynamic techniques assess the model's execution. As noted in DoD (2001b), models can be "instrumented;" i.e., code included in the individual model collects information about the functioning of its elements while it is running. This data on the performance of an individual model's elements reveal whether it is doing what it should be doing.
- Formal techniques use formal mathematical proofs; they mainly are recommended for models to which other methods are unsuitable.

Applicants should identify the particular method or methods used in their verifications.

(2) Code verification - the applicant should offer evidence that the logic was verified successfully to ensure that the code is implemented properly.

Additional Information

See Additional Information for Criterion 1.

(3) The applicant should provide evidence that the flow of information between the HPM elements was verified.

Additional Information

Verification of the flow of information between the individual elements of an HPM establishes that (1) each element receives the correct input from other ones, and. (2) each element outputs the correct information to others in the model.

(4) The applicant should provide evidence that the flow of information between the HPM and other M&Ss was verified.

Additional Information

HPMs can be linked with other simulations, such as plant models. Verifying the flow of information between HPMs establishes that (1) each M&S receives the correct input from other M&Ss, and (2) each M&S outputs the correct information to the others.

8 HPM VALIDATION

Validation is the process of determining the degree to which an HPMs output is realistic. It covers two aspects: structural, and integrated performance. The former addresses the internal composition of the HPM's elements. The latter deals with how well the HPM's results predict real human performance. In addition, review guidance is presented for the validation team.

8.1 Validation Team

(1) The team(s) performing validation should include the necessary subject-matter experts.

Additional Information

The subject-matter expertise needed to validate an HPM depends on the model's purpose and its mode of implementation. In addition to HFE- and I&C- (software) experts, the team should include people knowledgeable in the aspect of performance that the HPM addresses, e.g., operations experts for modeling operator performance. Additional expertise might be needed if the HPM is used in a federation including, for example, a plant model. Then, experts on the plant model should be included.

(2) The team(s) performing the validation should be technically independent from the HPM's developers.

Additional Information

The independence of the validation team is important to ensuring an unbiased process and to instilling confidence in their finding; HPM developers should be excluded. While the validation team should include other individuals, (Pew & Mavor, 1998), cooperation with the HPM developers will support the team's understanding of the HPM and its uses (DoD, 2001d).

8.2 Structural Validation

Structural validation addresses the internal composition of the HPM's elements. The staff's review should revolve around the applicant's methods and acceptance criteria for accomplishing the aspects of verification identified in the following criteria.

(1) The scope of the applicant's structural validation should include:

- the completeness of the model's elements for their intended purposes
- the technical basis for the model's elements
- the adequacy of the modeling of individual elements
- the reasonableness of the output at the element level

Additional Information

The list of bulleted items in this criterion was adapted from DoD (1999a).

(2) The applicant should offer evidence that individual model elements were validated successfully.

Additional Information

The applicant's conclusion about overall structural validation should be stated clearly, and the supporting rationale set out. Any remaining limitations that were not corrected must be identified and their implications on the model's overall performance discussed.

The DoD (2005) recommends functional decomposition; i.e., decomposing the M&S into functional units organizing these elements by the model's intended use. SMEs knowledgeable with each functional unit can assess its validity through evaluating the documentation, code, and output for that component.

The DoD (1999a) suggested that SMEs address the following questions in their review of the functional unit or element level:

- Does the HPM element produce expected results?
- Is the output/result reasonable compared to the inputs?
- Does a difference in input produce the expected proportional change in the output?
- How does the M&S element output compare to historical data, and data from tests, the laboratory, and exercise?

8.3 Integrated Performance Validation

Integrated performance validation addresses the realism of the HPM's output when all its elements are integrated. Validation is essential after integrating elements because errors are additive (Topcu, 2003); thus, while each element of a model may be valid, the model as a whole may not be. Although each element in the HPM may have an acceptable level of error, the integrated model may not generate meaningful results when the effects of all the errors are summed. Therefore, acceptable results must be demonstrated when all elements are integrated.

Performance of the HPM is validated by comparing its results with those from another source, called the "referent" such as actual human performance. The staff's review of performance validation covers the quality of the referent, the methodology of this comparison, and the criteria for determining the HPM is valid.

(1) The applicant should identify the referent information, and an acceptable rationale for using it as the basis for HPM validation.

Additional Information

The referent should represent the kinds of output the HPM is intended provide. That is, it should offer the same type of information that the HPM output produces, such as mission time, or workload profiles. Several different referents were identified for HPM (DoD, 2001a), including:

- SME's judgments
- empirical observations or experimental data from actual operations
- validated simulations of human behavior

The DoD (2001a) notes that referent(s) data should have the same level of detail as the HPM performance data; one with a higher or lower level of detail will complicate comparisons. For example, suppose an HPM, such as IMPRINT, produces a workload profile for visual-, auditory-, cognitive-, and psychomotor- workload during different phases of a crew's response to managing a LOCA. If the referent is a single workload measure obtained from real crews at the end of a LOCA scenario, it will be difficult to compare the two to determine whether the HPM is validated.

(2) The applicant should provide evidence that the range of conditions/scenarios used for HPM validation corresponds to the range of those for which the HPM will be used in HFE program.

Additional Information

The choice and design of the scenarios/conditions significantly affect the validation findings (DoD, 2004a). The scenarios included should reflect the HPM's intended purpose. Since every condition cannot be tested, sampling is necessary. To the extent practical, the sample should represent the range of scenarios/conditions that fall within the model's intended purpose. Selection should be such as to provide a comprehensive basis to permit generalization to other conditions or combinations thereof not addressed explicitly by the validation tests (see discussion in NUREG-0711, Section 11.4.1 of sampling operational conditions).

(3) The applicant should identify the methodology and acceptance criteria for comparing the HPM's performance output and the referent information, and give an acceptable rationale for its use as the basis for HPM validation.

Additional Information

At the heart of validation is the methodology used to define the relationship between the HPM's output and the referent information. Several qualitative- and quantitative- approaches were proposed (Campbell & Bolton, 2005; DoD, 1999a; 2000d; 2001a; 2001b; and 2001c). Some of them are discussed below. Since each method is limited in one way or another, a combination of them offers a strong basis for affirming validity when their findings converge.

The main approaches discussed are

- Face Validation
- Turing Test Validation
- Model Comparison Validation
- Sensitivity Analysis Validation
- Goodness-of-fit Validation

<u>Face Validation</u>

The DoD (1999a) defines face validation as "…the process of determining whether an M&S, on the surface, seems reasonable to personnel who are knowledgeable about the system." In this method, SMEs evaluate the model's output under various scenarios. The referent is the SME's judgment; thus, the precise bases for judgment and acceptance criteria are not clear cut. Hence, face validation is not recommended as the sole basis for establishing HPM validity.

SMEs' judgments should be made independently. SME evaluations improve when the methodology to obtain their assessments is standardized, objective, systematic, repeatable, and independent (Campbell & Bolton, 2005). Several other methods, below, rely on SME judgment, but in a more structured manner than face validation.

<u>Turning Test Validation</u>

In a Turning Test, SMEs are asked to distinguish data produced by a model from that produced by human subjects under highly similar scenarios (DoD, 1999a; DoD, 2001c). If the SME easily discriminates between them, the HPM is not validated. Should the SME have difficulty distinguishing the two data sets, the HPM is validated.

Model-comparison Validation

Model-comparison validation entails comparing the HPM output to that of a model validated for similar applications (DoD, 1999a). Similar to a Turning Test, if an SME can easily discriminate between two data sets, the HPM is not validated; but is validated when the SME has difficulty distinguishing between them

Sensitivity-analysis Validation

A sensitivity analysis involves running the model under various conditions that should produce variations of known effects in performance (DoD, 1999a; DoD, 2001c). SMEs evaluate whether the model's output is as expected.

Goodness-of-fit Validation

Campbell and Bolton (2005) recommend goodness-of-fit approaches to validating HPMs. This requires obtaining data from the HPM and from human participants under very similar conditions. The two data sets are analyzed and their goodness-of-fit quantified, e.g., using a measure, such as the coefficient of determination (r^2) that denotes the amount of variance in the actual performance data that the HPM predicts. Confidence intervals also should be provided so the data from the model and the human participants can be compared more meaningfully.[11]

When models are developed, human data are often used. These same developmental data are inappropriate for validating a model. However, if there are few data, a cross-validation approach (Campbell & Bolton, 2005) may suffice. The data are divided into two sets; the first is used for developing the model, while the second constitutes the referent in a validation test.

Campbell and Bolton (2005) suggested ways to improve the goodness-of-fit comparison:

- Pattern matching – the HPM is compared to a complex pattern of human data. One approach is to assess the goodness-of-fit of individual humans to the overall data. The same assessment then is done with the HPM data. An evaluation is made of whether the HPM fits the overall data in a similar way to that from individual humans.
- A priori prediction – the HPM is used to predict a pattern of results in scenarios for which the modeler has no data. Data then are collected in the new scenarios and compared to the model.

(4) The applicant should provide evidence that the HPM was validated.

Additional Information

The applicant should draw clear conclusions about the validation of overall performance, and supply the supporting rationale. Any known uncorrected limitations should be identified and their implications for the model's overall performance discussed.

Campbell and Bolton (2005) noted that qualitative and quantitative approaches should be employed for analyses due to the diversity of performance measures and methods used in HPM

[11] Note that traditional hypothesis testing statistics are not appropriate to model validation, since this would be an attempt to confirm the null hypothesis; i.e., to show that the model and the human data are the same.

validation. To draw conclusions on the acceptability of a model's performance, criteria must be established for comparison to the referent's performance.

There are four approaches to establishing criteria, based upon the type of comparisons undertaken: requirement-referenced, benchmark referenced, normative referenced, and expert-judgment referenced (O'Hara et al., 1997, 1999). When criteria are Requirement Referenced, the model's performance is compared with a quantified performance requirement; i.e., one defined through engineering analysis. When criteria are Benchmark Referenced, the performance of the integrated system is compared to a benchmark system, e.g., a current system predefined as acceptable. For Normative Referenced criteria, the performance criteria are established via many evaluations, rather than a single benchmark system. Finally, when the criteria are Expert-judgment Referenced, the performance of the HPM is compared to criteria established by expert judgment. Validation may use a combination of these approaches, since the types of performance to be measured are qualitatively different.

The degree of convergence of the multiple measures of performance should be evaluated (Campbell & Bolton, 2005; O'Hara, 1999). The applicant's conclusion regarding overall performance validation should be clearly stated along with the supporting rationale. Any limitations identified that were not corrected should be identified and discussed with respect to overall model performance for its intended application.

9 USER INTERFACE VERIFICATION

1. The user-interface documentation should describe the specific purpose of the HPM, including its intended usage, and the range of conditions in which it can be used.

   Additional Information

   HPMs are designed to address specific questions and validation typically focuses on them (Campbell & Bolton, 2005; DoD, 2003; DoD, 2005). It is essential to define the scope of the HPM well, so that users and regulatory personnel clearly understand it. For example, an HPM may be suitable for estimating the time for an operator to complete an individual manual task, but not for determining the time for a crew to perform a task requiring communication. User guidance should clearly identify the HPM version number so the user is sure the documentation is appropriate for the HPM's software.

2. The user interface documentation should describe clearly what the model cannot be used for.

   Additional Information

   Potential misuses of the HPM should be stated succinctly in the documentation to help ensure HPMs are used properly.

3. The user guidance should give precise step-by-step instructions on "how to" use the model, covering the following:

   - Initiate the model, and perform any necessary configurations or calibrations.
   - Provide model inputs, i.e., the states, parameters, and values that are set in the model before running it. The default conditions used if users do not set each model input should be identified.
   - Establish interfaces to other M&Ss, if necessary, and to verify that they are configured properly.
   - Select desired model outputs.
   - Obtain additional information and help.

4. The HPM user-interface should reflect principles of good human-system interface design.

   Additional Information

   In addition to having good documentation for users, the HPM user-interface should be designed to support user tasks, to be compatible with human cognitive- and physiological- characteristics, and to minimize user's errors. NUREG-0700 (O'Hara et al., 2002) contains HSI design review guidelines that will support this aspect of the review.

10 REFERENCES

Allender, L., Archer, S., Kelley, T. & Lockett, J. (2005). Human performance modeling in the Army: A long and winding road. In the *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting*. San Monica, CA: Human Factors and Ergonomics Society.

Allender, L., Kelley, T., Salvi, L., Lockett, J., Headley, D., Promisel, D., Mitchell, D., Richer, C. & Feng, T. (1995). Verification, validation, and accreditation of a soldier-system modeling tool. In *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting* (pp. 1219-1223). San Monica, CA: Human Factors and Ergonomics Society.

Archer, S., Gosakan, M., Shorter, P., & Lockett III, J. (2005). New capabilities of the Army's maintenance manpower modeling tool. *Journal of the International Test and Evaluation Association*. *26* (1), 19-26.

Australian Defense Simulation Office (2005). *Simulation verification, validation and accreditation guide*. Canberra, Australia: Commonwealth of Australia.

Baron, S., Kruser, D. & Huey, B. (Eds) (1990). *Quantitative modeling of human performance in complex, dynamic systems*. Washington, DC: The National Academies Press.

Campbell, G. & Bolton, A. (2005). HBR validation: Integrating lessons learned from multiple academic disciplines, applied communities, and the AMBR Project. In Gluck, K., & Pew, R. (Eds.). (2005). *Modeling human behavior with integrated cognitive architectures: Comparison, evaluation, and validation*. Mahwah, NJ: Lawrence Erlbaum Associates.

Conwell, C., Enright, R. & Stutzman, M. (2000). Capability maturity models support of modeling and simulation verification, validation, and accreditation. In *Proceedings of the 32nd SCS Winter Simulation Conference*. San Diego, CA: The Society for Modeling and Computer Simulation (pp. 819–828).

DoD (1996). *Verification, validation, and accreditation.* (Air Force Instruction 16-1001). Washington, DC: United States Air Force.

DoD (1999a). *Verification, validation, and accreditation of Army models and simulations* (Department of the Army Pamphlet 5-11). Washington, DC: United States Army.

DoD (1999b). *Verification, validation, and accreditation of models and simulations* (Navy Instruction 5200.40). Washington, DC: United States Navy.

DoD (2000b). *V&V tools*. Washington, DC: Defense Modeling and Simulation Office. (Retrieved June 19, 2008 from http://vva.dmso.mil/default.htm)

DoD (2000c). *T&E/V&V checklist*. Washington, DC: Defense Modeling and Simulation Office. (Retrieved June 19, 2008 from http://vva.dmso.mil/default.htm)

DoD (2000d). *Subject matter experts and VV&A*. Washington, DC: Defense Modeling and Simulation Office. (Retrieved June 19, 2008 from http://vva.dmso.mil/default.htm)

DoD (2001a). *Validation of human behavior representation*. Washington, DC: Defense Modeling and Simulation Office. (Retrieved June 19, 2008 from http://vva.dmso.mil/default.htm)

DoD (2001b). *V&V techniques*. Washington, DC: Defense Modeling and Simulation Office. (Retrieved June 19, 2008 from http://vva.dmso.mil/default.htm)

DoD (2001c). *A practitioner's perspective on simulation validation*. Washington, DC: Defense Modeling and Simulation Office. (Retrieved June 19, 2008 from http://vva.dmso.mil/default.htm)

DoD (2001d). *V&V agent's role in the VV&A of new simulations*. Washington, DC: Defense Modeling and Simulation Office. (Retrieved June 19, 2008 from http://vva.dmso.mil/default.htm)

DoD (2003a) *Human behavior representation (HBR) literature review.* Washington, DC: Defense Modeling and Simulation Office. (Retrieved June 19, 2008 from http://vva.dmso.mil/default.htm)

DoD (2003b). *Modeling and simulation (M&S) verification, validation, and accreditation* (DoD Instruction Number 5000.61).

DoD (2004). *Validation*. Washington, DC: Defense Modeling and Simulation Office. (Retrieved June 19, 2008 from http://vva.dmso.mil/default.htm)

DoD (2005). *Management of Army models and simulations* (Army Regulation 5–11). Washington, DC: Department of the Army.

Gluck, K. & Pew, R. (Eds.). (2005). *Modeling human behavior with integrated cognitive architectures: Comparison, evaluation, and validation*. Mahwah, NJ: Lawrence Erlbaum Associates.

Graffagnini, J. & Youngblood, S. *The verification, validation, and accreditation report templates*. (Retrieved June 19, 2008 from www.sisostds.org/index.php).

Hansberger, J. & Barnett, D. (2005). Human performance modeling for operational command, control, and communication. In *Proceedings of the HFES - 38th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Harmon S. & Youngblood, S. (2005). A proposed model for simulation validation process maturity. *Journal of Defense Modeling and Simulation*, 2 (4), 179–190.

Hautamaki, B., Bagnall, T. & Small, R. (2006). Human interface evaluation methods for submarine combat systems. (Retrieved June 19, 2008 from http://www.npt.nuwc.navy.mil/symposium/2006/06Tech/11%20HautamakiEtAl.pdf).

Higgins, J., O'Hara, J. Lewis, P., Persensky, J., Bongarra, J., Cooper, S. & Parry, G. (2007). *Guidance for the review of changes to human actions: Final report* (NUREG-1764). Washington, D.C.: U. S. Nuclear Regulatory Commission.

Hugo, J. (2006). A model-base approach to the integration of human factors and systems engineering. *ESSA 2006 Conference*, Pretoria, South Africa: CSIR.

IEEE (1977). *Recommended practice for distributed interactive simulation – verification, validation, and accreditation* (IEEE Std. 1278.4-1977 – reaffirmed in 2002). New York, NY: The Institute of Electrical and Electronics Engineers.

Laughery, R. & Persenksy, J. (1994). Network modeling of nuclear operator procedures. In *Proceedings of the HFES - 38th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Laughery, R., Plott, B., Engh, T. & S. Nash, S. (1996). Discrete event simulation as a tool to determine necessary nuclear power plant operating crew size. In *Proceedings of the 1996 Winter Simulation Conference*. San Diego, CA: The Society for Modeling and Simulation.

Lawless, K., Laughery, R. & Persensky, J. (1995). *Using Micro Saint to predict performance in a nuclear power plant control room: A test of validity and feasibility* (NUREG/CR-6159). Washington, D.C.: U.S. Nuclear Regulatory Commission.

Mitchell, D. (2000). *Mental workload and ARL workload modeling tools* (ARL-TN-161). Aberdeen Proving Ground: Army Research Laboratory.

Mitchell, D., Samms, C., Henthorn, T. & Wojciechowski, J. (2003). *Trade study: A two- versus three-soldier crew for the mounted combat system (MCS) and other future combat system platforms* (ARL-TR-3026). Aberdeen Proving Ground, MD: Army Research Laboratory.

NATO (2008). *Verification, validation, and accreditation (VV&A) of federations* (TR-MSG-019). NEUILLY-SUR-SEINE CEDEX, France: North Atlantic Treaty Organization.

NRC (1998). *An approach for using probabilistic risk assessment in risk-informed decisions on plant-specific changes to the licensing basis* (Regulatory Guide 1.174). Washington, DC: U.S. Nuclear Regulatory Commission.

NRC (1999). *Plant-human Review and Effectiveness Decision Tool (PHRED): User's guide*. Washington, D.C.: U.S. Nuclear Regulatory Commission.

NRC (2007). *Standard review plan* (NUREG-0800); Chapter 18, Human Factors Engineering. Washington, D.C.: U.S. Nuclear Regulatory Commission.

O'Hara, J. (1999). A quasi-experimental model of complex human-machine system validation. *International Journal of Cognition, Technology, and Work*, *1*, 37-46.

O'Hara, J., Brown, W., Lewis, P. & Persensky, J. (2002). *Human-system interface design review guidelines* (NUREG-0700, Rev 2). Washington, D.C.: U.S. Nuclear Regulatory Commission.

O'Hara, J., Higgins, J., Brown, W. & Fink, R., Persensky, J., Lewis, P. & Kramer, J. (2008a). *Human factors considerations with respect to emerging technology in nuclear power plants* (NUREG/CR-6947). Washington, D.C.: U. S. Nuclear Regulatory Commission.

O'Hara, J., Higgins, J., Brown, W. & Fink, R. (2008b). *Human factors considerations with respect to emerging technology in nuclear power plants: Detailed analyses* (BNL Technical Report No: 79947-2008). Upton, NY: Brookhaven National Laboratory.

O'Hara, J., Higgins, J., Persensky, J., Lewis, P. & Bongarra, J. (2004). *Human factors engineering program review model* (NUREG-0711, Rev. 2). Washington, D.C.: U.S. Nuclear Regulatory Commission.

O'Hara, J., Plott, C., Milanski, J., Ronan, A., Scheff S, Laux, L. & Bzostek J. (2009). *Trends in HFE methods and tools and their applicability to safety reviews* (BNL Tech Report No. BNL-90424-2009). Upton, NY: Brookhaven National Laboratory.

O'Hara, J., Stubler, W., Higgins, J. & Brown, W. (1997). *Integrated system validation: Methodology and review criteria* (NUREG/CR-6393). Washington, D.C.: U.S. Nuclear Regulatory Commission.

Persensky, J., Szabo, A., Plott, C., Engh, T. & Barnes, A. (2005). *Guidance for assessing exemption requests from the nuclear power plant licensed operator staffing requirements specified in 10 CFR 50.5t* (NUREG-1791). Washington, D.C.: U.S. Nuclear Regulatory Commission.

Pew, R. (2007). Some history of human performance models. In W. Gray (ed.). *Integrated models of cognitive systems.* New York, NY: Cambridge University Press.

Pew, R. (2008). More than 50 years of history and accomplishments in human performance model development. *Human Factors*, *50* (3), 489-496.

Pew, R. & Mavor, A. (1998). *Modeling human behavior and command decision making: Representations for military simulations*. Washington, D.C.: The National Academies Press.

Plott, C., Engh, T. & Barnes, V. (2004). *Technical basis for regulatory guidance for assessing exemption requests from nuclear power plant licensed operator staffing requirements specified in 10 CFR 50.54(m)* (NUREG/CR-6838). Washington, D.C.: U.S. Nuclear Regulatory Commission.

Sebok, A. & Plott, B.(2008). Staffing levels: Methods for assessing requirements. In A. B. Skjerve, A. Bye, F. Øwre, K. Haugset, G. Skraaning Jr., R. Boring, & J. Neeb (Eds.) *The HAMMLAB history - 25 years of experimental research and development.* Halden, Norway: Institute of Energy Technology.

Sebok, A., Hallbert, B., Plott, B. & Nash, S. (1997). Modeling crew behavior and diagnoses in the control room. In *Proceedings of the IEEE Sixth Conference on Human Factors and Power Plants*. Orlando, FL. New York: IEEE, pp. 7.13 - 7.17.

Sheridan, T. (2005). A brief history of normative models of human behavior. In *Proceedings of the HFES - 49th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Topcu, O. (2003). *Review of verification and validation methods in simulation* (DRDC Atlantic TM 2003-055). Canada: Defense Research and Development Canada.

Walters, B. & Yow, A. (2000). An evaluation of task network modeling for use in simulating situation awareness. In *Proceedings of the IEA 200/HFES200 Congress*. Santa Monica, CA: Human Factors and Ergonomics Society.

Wickens, C. (1991). Processing resources and attention. In D. Damos (Ed.), *Multiple task performance.* London: Taylor & Francis.

Wickens, C. (1987). Attention. In P. Hancock (Ed.), *Human factors psychology*. New York, NY: Elsevier Science Publishers.

Wickens, C. (1980). The structure of attentional resources. In R. Nickerson (Ed.), *Attention and performance VIII*. Hillsdale, NJ: Lawrence Erlbaum.

Wickens, C., Sebok, A., Bagnall, T. & Kamienski, J. (2007). Modeling of situation awareness supported by advanced flight deck displays. In *Proceedings of the HFES - 51th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Young, M. (2003). Human performance model validation: One size does not fit all. (Retrieved June 19, 2008 from http://www.scs.org/getdoc.cfm?id=2367).

Yow, A., Walters, B., Plott, B., Laughery, R. & Persensky, J. (2005). Predicting nuclear power-plant operator performance using discrete event simulation. *Journal Cognition, Technology & Work*, *7* (1), 29-35.

GLOSSARY

Accreditation - The official certification that a model, simulation, or federation of models and simulations and its associated data are acceptable for use for a specific purpose.

Code verification - a verification that assures that the human performance model's verified logic has been properly implemented in code.

Federation - A system of interacting models and/or simulations, with supporting infrastructure, based on a common understanding of the objects portrayed in the system.

Fidelity (model) - Fidelity is the exactness of a model's representation of human behavior.

Human performance model - Human performance models that are (1) mathematical, programmable, and executable rather than purely explanatory; and (2) applied in the engineering design and evaluation of complex systems. Human performance models are also called human behavior representations."

Integrated performance validation - The process of determining that the model as a whole, with all model elements integrated together, predicts human performance.

Logic verification - A verification that assures that the human performance model's algorithms correctly represent the intended processes in relation to the M&S requirements and specifications.

Model - A physical, mathematical, or otherwise logical representation of a system, entity, phenomenon, or process.

Modeling and simulation - The use of models and simulations to develop data as a basis for making managerial or technical decisions.

Model element - An aspect of a human performance model that simulates some aspect of human- or team-behavior.

Referent - In validation, the referent is the information/data against which human performance model's results are compared to determine the model's realism.

Resolution (model) - Resolution is the degree of precision of a human performance model's results.

Simulation - A method for implementing a model over time. Also, a technique for testing, analysis, or training in which real-world systems are used, or where real-world and conceptual systems are reproduced by a model.

Structural validation - The process of determining that the internal composition of the human performance model's elements produce realistic results.

Validation - Validation is the process of determining the degree to which a human performance model's output is realistic. (See structural- and integrated performance- validation).

Verification - The process of determining that a human performance model's implementation and its associated data accurately represent the developer's conceptual description and specifications. (See logic- and code- validation).

Appendix

Recommended Contents of an Applicant's HPM Submittal

Submittals made by applicants for HPM review should provide the information identified in this appendix.

1 Purpose and Scope

The submittal should describe the specific purpose of the HPM with respect its role in the HFE program.

The submittal should identify the scope of the HPM, including its intended usage and the range of conditions for which it can be used. This information should clearly identify the boundary conditions at which the model's validity is not assured.

2 HPM Technical Description

The submittal should provide identifying information, such as name of the model and its version number.

The submittal should describe the HPM, covering:

- *Model inputs* - the model's states, parameters, and values the user must set before using it. The submittal should identify the default conditions used if users do not set each model input, and the implications of using the default settings
- *Model elements* – a description of the model's constituent elements e.g., situation awareness, workload, communication, environmental factors, scenarios, and their algorithms, i.e., the means by which the elements are used in the model
- *Element interactions* – a description of how the elements interact
- *HPM interactions with other models* – a description of the HPM's linkages with other M&Ss, e.g., interactions between an HPM and a plant thermodynamic model
- *Model output* – a description of the results the model generates

Potential misuses of the model should be identified in the submittal, along with any constraints, limitations, and cautions in using it.

3 HPM Technical Basis

The submittal should describe the technical basis, psychological theories or data, for each of the model's elements and their interactions.

The submittal should identify any predecessor models that served as the basis for the HPM, their uses, and any information pertaining to their use, verification, and validation.

4 HPM Verification and Validation

The submittal should describe the expertise of the verification team, and their relationship to the HPM development process.

The submittal should describe the methods, acceptance criteria, and results of the verification performed to show that the HPM's logic and code correctly perform their intended functions.

The submittal should describe the applicant's validation methods, acceptance criteria, and results for determining that the HPM realistically predicts the behavior of interest.

5 HPM Operations and User Interfaces

The submittal should describe how users interact with the HPM to input data, run the model, and obtain output.

The submittal should describe the HPM's HSIs.

The submittal should describe the documentation available to users, such as user manuals and online help.