SAND2012-1318C

HRA Method Analysis Criteria

Stacey M. L. Hendrickson[a*], John A. Forester[a], Vinh N. Dang[b], Ali Mosleh[c], Erasmia Lois[d], Jing Xing[d]

[a]Sandia National Laboratories, Albuquerque, NM, USA
[b]Paul Scherrer Institute, Villigen PSI, Switzerland
[c]University of Maryland, College Park, MD, USA
[d]US Nuclear Regulatory Commission, Washington, DC, USA

Abstract: The Nuclear Regulatory Commission's (NRC) Office of Nuclear Regulatory Research (RES) is working under a memorandum of understanding with the Electric Power Research Institute (EPRI) in addressing the Commission's Staff Requirements Memorandum (SRM) [1] to the Advisory Committee on Reactor Safeguards (ACRS) to develop improved guidance and methods for human reliability analysis (HRA). A hybrid approach has been proposed to address this SRM combining (at least conceptually) elements from A Technique for Human Error Analysis (ATHEANA) [5, 6], Cause-Based Decision Tree (CBDT) [7], and the information, decision, and action in crew context (IDAC) model [19]. An initial step in forming this model was the creation of a list of criteria determined to be important in an HRA method. A group of experts was convened to develop this list of criteria. Elements were determined to be important in the following areas: 1) content validity, 2) empirical validity, 3) reliability, 4) traceability, 5) construct validity, 6) level of analysis, and 7) usability. Following the development of these criteria, a selection of existing HRA methods were judged on their ability to meet each of the criteria. This paper will present the list of criteria and provide a discussion of how the following methods fared when judged on the criteria: 1) ATHEANA, 2) CBDT, 3) Standardized Plant Analysis Risk HRA (SPAR-H) [8], 4) Technique for Human Error Rate Prediction (THERP) [9], and 5) Accident Sequence Evaluation Program (ASEP) [10]. Although not discussed in this paper, based on the evaluations of current methods and a study of the criteria, a plan was developed for moving forward in creating a hybrid or integrated HRA method that would fulfill (to the extent possible) as many elements of the criteria as possible. A companion paper at this conference discusses the plans for the new method [2].

Keywords: Human Reliability Analysis (HRA), IDHEAS, Method Comparison

1. INTRODUCTION

The U.S. Nuclear Regulatory Commission (NRC) Office of Research (RES) is working with the Electric Power Research Institute (EPRI) under a Memorandum of Understanding to address a Staff Requirements Memorandum (SRM) [1] given to the Advisory Committee on Reactor Safeguards (ACRS) by the Commission. The SRM directs the ACRS to "work with the staff and external stakeholders to evaluate the different human reliability models in an effort to propose a single model for the agency to use or guidance on which model(s) should be used in specific circumstances." An overview of the project is provided in a companion paper at this conference [2].

To appreciate what is being done in responding to this SRM, an understanding is needed of what is currently done in HRA. Furthermore, an appreciation of what elements are necessary to be addressed within an HRA and what technical gaps currently exist is needed. Therefore, two tasks were undertaken to gain this understanding. First, a group of experts in the area of HRA was convened to develop a list of criteria that should be addressed within HRA. Second, several NRC and US industry developed methods that are in current use were compared against this list of criteria so best practices could be determined and technical gaps could be identified. This information was combined with what has been learned from the development of the good practices within HRA [3, 4] to make a determination of how to address the challenges posed by the SRM.

2. CRITERIA TO BE ADDRESSED WITH HRA

In March 2009 a two-day workshop, comprised of NRC, national laboratory, industry and private sector experts and practitioners in the areas of HRA, PRA, and human factors, was conducted to determine the path forward for addressing the SRM on HRA modeling. During the workshop the SRM guidance was discussed and debated. It became clear that the SRM indicated that the "single model" approach was the most desirable solution and that if many models were to be retained, a justification for such decision should be provided and guidance should be developed for which models should be used for which applications, as well as guidance for the use and application of the particular model in the particular application. However, it also became clear that stakeholders preferred a "tool box" approach so that analysts could perform a high-level screening type of analysis or a more detailed analysis, depending on the application needs. In other words, even though a single detailed method was desirable, different methods or different versions of a single method might be needed for different levels of analysis.

In order to be able to address these issues, HRA desirable attributes were identified to provide a guide for determining a path forward. The attributes reflect the accumulated experience of the participants in the workshop based on their understanding of the HRA needs and comparison with existing methods and implementation practices. They are intended to cover both qualitative and quantitative aspects.

1) Content validity – this criterion is a form of internal validity and addresses how well the method represents all elements of the construct it claims to measure.
    a. Context characteristics including:
        i. Crew: For HRA purposes, when used in the context of a PRA, the method should address the average crew performance. However, when used retrospectively (e.g., for event assessment), the method should be able to address important, systematic individual crew differences to the extent possible.
        ii. Plant: Characteristics of the plant as the scenario evolves (cues, expected conditions, scenario timing, available time for action, etc.) should be accounted for by the method.
        iii. Performance Shaping Factors (PSFs) or other conceptualizations of the factors/elements affecting performance should be included.
    b. Cognition and diagnosis as well as action and execution should be handled by the method.
    c. PSF (or other factor) interactions should be accounted for. That is, differing effects of some PSFs in the presence of other PSFs should be considered to the extent possible.
    d. Errors of Commission (EOCs) and Errors of Omission (EOOs): The method should be able to identify, characterize, and quantify EOCs and EOOs.
    e. Dependency between HFEs should be covered.
    f. Recovery within an HFE (not cutset recovery). A key question is whether the method credits recovery (e.g., the presence of a second checker, self or crew recovery based on additional information obtained as the scenario develops) for both diagnosis and execution.
    g. Failure mechanisms should be identified in the qualitative analysis. Failure mechanisms describe how the failure would come to pass. This idea is contrasted against failure mode, which describes how the failures can be recognized as having occurred. One issue addressed in this requirement is the notion that the qualitative and quantitative analyses should match, such that the work done in the qualitative analysis should be addressed in the quantitative analysis. For example, the likelihood of particular failure mechanisms occurring as a function of the conditions should be addressed during quantification.
    h. Uncertainty analysis – The method should be able to address uncertainty and provide uncertainty bounds.
    i. The method should define the basic unit of quantification and level of granularity. In other words, is the method covering a single operator or the crew? Does the method help the analyst determine the level of decomposition for the HFE?
2) Reliability – this criterion addresses how consistently the method measures its intended construct. It asks a series of questions of the method:
    a. Reproducibility – If analysts make the same assumptions, will they get the same answer?
    b. Consistency –Will analysts choose the same PIFs and assume the same influence level (e.g., weight the PIFs consistently)?
    c. Inter- and Intra-rater (analyst) reliability (testing reliability)

i. Inter-rater reliability – Do multiple raters produce the same human error probability (HEP)?
ii. Intra-rater reliability – Does a single rater produce a HEP consistently?

3) Traceability (transparency) – This criterion refers to the ability to reverse engineer the analyst's work mathematically in quantification and conceptually in qualitative steps.
4) Validity – In general this criterion asks if the method is consistent with its underlying model.
   a. Construct Validity – This element refers to the method measuring what it states to be measuring.
   b. Empirical validity – This criterion asks if the method has an empirical basis. This criterion could include an empirical basis for the underlying model and one for the HEPs included in the method.
5) Adaptability/Scalability (graded approach) – This element addresses whether a method can be adapted to different application domains and levels of analysis in a consistent manner.
6) Usability – The method should provide a practical tool without compromising the technical integrity of the analysis.

These attributes cover the HRA process; however, a firm technical basis, as well as implementation tools are necessary for users to correctly and efficiently perform an analysis. The aim of the response to the SRM is to set the foundation and develop an HRA method that meets the desirable attributes (to the extent possible) so that, among other things, unwanted variability in results is addressed at a sufficient level.

3. METHOD COMPARISON TO CRITERIA

Following the introduction of the analysis criteria, several currently used HRA methods were examined to see how well they address each of the elements. The methods examined with regard to these criteria included:

- A Technique for Human Error Analysis (ATHEANA) [5-6]
- Cause-Based Decision Tree (CBDT)[1] [7]
- Standardized Plant Analysis Risk HRA (SPAR-H) [8]
- Technique for Human Error Rate Prediction (THERP) [9]
- Accident Sequence Evaluation Program (ASEP) [10]

The methods found to have elements of each of the criteria are represented in Table 1. The table presents a simplified answer to each of the criteria; however, a "yes" within the table should not be interpreted as no improvement could be done to that element. In places within the table, only a partial or minimal coverage is given by the method. The concept of reliability is not addressed in the table below. However, this concept was addressed in the International and Domestic Empirical Studies [11, 12]. The general findings of the empirical study demonstrated that there is no single method that addresses all levels of reliability. Two other criteria missing from the list below are empirical and construct validity. These are difficult elements to measure or give definitive answers to. However, in the long run it should be possible to validate the concepts that a method uses. In principle a method's constructs should be able to be tested and validated through simulator studies.

[1] CBDT was evaluated with the understanding that it addresses only the cognitive failure probability.

Table 1. Comparison of methods to the analysis criteria

Criteria	ATHEANA	CBDT	SPAR-H	THERP	ASEP
Content Validity					
Context - Crew	Yes	No	Minimal	No	No
Context - Plant	Yes	Yes	Minimal	No	Minimal
Context - PSF	Yes	Yes	Yes	Yes	Yes
Cognition/Diagnosis	Yes	Yes	Yes	Yes	Yes
Action/Execution	Yes	Yes	Yes	Yes	Yes
PSF interaction	Yes	Partial	No	No	No
Dependency between HFEs	Yes[2]	Yes	Yes	Yes	Yes
EOC	Yes	No	No[3]	No	No
EOO	Yes	Yes	Yes	Yes	Yes
Recovery	Yes	Yes	Indirectly	Yes	Yes
Failure mechanisms	Yes	Yes	No	No	No
Uncertainty	Yes	No	Yes	Yes	Yes
Unit of quantification	Yes	Yes[4]	No	Yes	Minimal
Traceability	Qualitative	Mathematical	Mathematical	Mathematical	Mathematical
Level of Analysis					
Screening	No	No	No	Yes	Yes
Scoping	No	No	Semi-detailed	No	Yes
Detailed	Yes	Yes	Semi-scoping	Yes	No
Usability	Resource intensive	Straight-forward	Straight-forward	Somewhat resource intensive	Straightforward

NOTE: CBDT, SPAR-H and ASEP all rely on the guidance given in THERP for addressing the dependency between HFEs. THERP does not explicitly account for dependency between HFEs, but the THERP dependency model has in practice been easily and reasonably adapted to address dependency between HFEs

Table 1 provides a simplified appraisal of methods against the analysis criteria. Examining the criteria in more depth points to ATHEANA [5-6] having the strongest qualitative analysis and providing a fairly comprehensive coverage of all the elements included within the content validity requirement. For instance, ATHEANA provides guidance for developing a full description of the context (including crew effects, plant conditions, and other influencing factors), accounts for cognitive and diagnosis failures, and strives to identify error-forcing conditions and failure paths that could contribute to HFEs. Furthermore, ATHEANA can be used to identify errors of commission as well as errors of omission. However, ATHEANA does not provide a readily traceable mathematical account of the quantification of the HEPs, and it lacks standardization in the manner in which it is applied, and relies heavily on expert judgment even though a structured approach is provided for performing expert elicitation to support quantification. Therefore, the method is subject to producing inconsistent results. Furthermore, application of ATHEANA can be time and resource intensive.

CBDT [7], on the other hand, is a causal-based, structured approach providing a standard format that allows for traceability of the calculation of the HEP, even though the basis for the decisions still need to be documented by the analysts. The reliability of the method may also be fairly high, but the reliability is largely dependent on the level of the qualitative analysis done. However, the qualitative analysis conducted through CBDT appears to lack full coverage of all the elements specified under the content validity criterion (particularly consideration of the appropriate range of PSFs and plant conditions) [11]. In addition, CBDT does not cover errors of commission and does not offer any guidance on how to perform task decomposition (but does reference SHARP1 [15]).

---

[2] In principle ATHEANA evaluates dependency between HFEs, but it relies on analyst judgment for a complicated assessment.

[3] The base failure rate within SPAR-H is a composite for omissions and commissions, but it does not address identification and quantification of diagnosis-based mistakes.

[4] CBDT follows guidance given in SHARP1 [15] for HFE identification.

The other methods evaluated each had their strengths and limitations. For instance, SPAR-H [8] offers a nice simplicity in the limited number of PSFs evaluated making the method fairly easy to use and offering traceability to some extent. However, partly due to its simplicity, the method may be missing some key contextual elements necessary for fully representing the situation being evaluated and needs additional guidance for scaling the PSFs. ASEP [10] suffers a similar problem in that it is relatively easy to use, which translates to it being less resource intensive and easier to understand, interpret, and review. However, this ease comes at a cost as the method potentially misses or incorrectly treats many PSFs that have been seen to affect the final HEP evaluation [11].

Finally, THERP [9] offered some strong elements as well (e.g. performance of a task analysis). THERP, one of the earliest HRA methods developed, aimed to ground its HEPs in empirical data. This goal was somewhat satisfied with the underlying bases for all the values being a mix of actual data and expert judgment. The actual data used is sparse and comes from reports and databases involving experiments using artificial tasks, experiments and field studies of actual tasks associated with industrial and process industries, available military data on human failures, simulations in nuclear power plants (NPPs), and actual events in NPPs. One caveat to note in the use of the method today is the age of most of these data, which date back to the 1960's through the early 1980's. If an analyst used the tables provided in THERP to ultimately calculate the HEP, many important contextual factors may be missed. However, if the analyst goes beyond the tables and uses expert judgment (an option in the method) in determining additional PSFs that must be evaluated, the reproducibility of the analysis is called into question.

4. GOOD PRACTICES WITHIN HRA

The evaluation of methods against the analysis criteria complements an earlier evaluation and determination of "good practices". NUREG-1792 [3] developed the Good Practices to HRA application as applied to Level 1 or Level 2 internal events PRAs with the reactor at full power. Following the introduction of the Good Practices in NUREG-1792, an evaluation of various HRA methods that are commonly used in regulatory applications was performed, with a particular focus on their capabilities to satisfy the good practices, as well as their respective strengths and limitations regarding their underlying knowledge and data bases. These evaluations were developed on the basis of the good practices and by eliciting input and feedback from HRA experts representing the NRC, national laboratories, the private sector, and international organizations including the ACRS, and documented in NUREG-1842 [4]. The methods examined included the same as compared in Section 3 of this paper as well as:

- Human Cognitive Reliability (HCR)/Operator Reliability Experiments (ORE) Method [7]
- Success Likelihood Index Methodology (SLIM) Multi-Attribute Utility Decomposition (MAUD) [13]
- Failure Likelihood Index Methodology (FLIM) [14]
- A Revised Systematic Human Action Reliability Procedure (SHARP1) [15]

The following subsections provide some of the general observations made through the evaluations done in NUREG-1842 [4].

4.1. Limitations in HRA Process

The discussion of the methods compared against the analysis criteria combined with the findings from NUREG-1842 [4] and the empirical studies [11, 12] identifies gaps or limitations that exist in current HRA methods. Each of these limitations may be covered in a few methods, they are largely left unaddressed in the remaining methods. However, no method was found that addressed all the limitations.

All the methods promote, albeit at varying degrees, the preference to use a multi-disciplinary team for performing HRA, so that no potentially important performance influencing factor is missed and a clear understanding of the performance conditions can be obtained. Further, HRA and human factors knowledge and expertise is found to be strongly desirable in the implementation of many methods. This is a desirable characteristic that is lacking in several methods. HRA methods should emphasize this preference much more strongly in their current guidance (especially those that can be very easily implemented without such expertise or corresponding training).

While most methods address the subject of using walkdowns, talk-throughs, and simulations as part of the HRA process, surprisingly this is emphasized more in some earlier methods than in many of the later methods. It may be that the latter take the use of these practices as given, given their emphasis in the good practices and coverage in the PRA standards [16] endorsed by RG 1.200 [17]. In any case, without such techniques to ensure the proper inputs and necessary understanding to properly judge the influencing factors and crew behavior, too much speculation or unsubstantiated judgments may be required by the HRA analyst, leading to undesirable variability in HRA results.

Virtually all methods agree on the framework of treating an HFE as having both a diagnostic (more cognitive) component and a response execution (implementation) component. This is a convenient logical distinction used by the various methods and is consistent with current models in the human behavior sciences. However, there is variability as to what human performance influencing factors are explicitly treated by the methods to address errors in both the diagnostic and execution phases of human actions, and some methods allow the diagnosis phase to be ignored when crews are following a procedure after an initiating event has been diagnosed, which can lead to inadequate assessment of crew cognitive activity and influencing factors [11].

4.2. Strengths of Features in Current Methods

Many methods have made great strides in addressing the gaps that have been uncovered in the HRA process. For instance, the need for a fuller, richer qualitative analysis is reflected in the use of ATHEANA [5-6] and MERMOS [18]. The use of task analysis techniques (e.g., as suggested by THERP [9]) can greatly assist in identifying and modeling HFEs and, in particular, can help to understand potential dependencies among human actions.

The automation and consistent nature of software tools like the EPRI HRA Calculator [19] is a positive enhancement in HRA. It takes away some of the burden of executing the analyses and may reduce inconsistency (computer screens remind the analyst what to consider each time). Additionally, such computerization can significantly assist HRA documentation, making it easier to review and reproduce.

The more current HRA methods (e.g., CBDT [7], ATHEANA [5, 6], MERMOS [18]) examine causes that could affect not only the implementation portion of an HFE, but also the diagnostic portion. This allows for a better understanding and more thoughtfully-based qualitative insights as to potential diagnostic vulnerabilities and their effects on the HEP than if a simple TRC (time reliability correlation) model (such as those used in THERP [9] or ASEP [10], for instance), by itself, is used for diagnosis errors.

4.3. Quantification Approach

Generally, and at a high level, the HRA methods that address quantification use one of three quantification approaches. One approach adjusts basic HEPs or otherwise determines the HEPs according to a list of influencing factors specifically addressed by the method. Another uses a more flexible context-defined set of factors and more expert judgment to estimate the final HEP. A third approach uses (to the extent practicable) empirical information based on simulations of accident scenarios in power plant simulators. All of these approaches have associated strengths and limitations that should be understood, so that thoughtful application of a method can be performed. The following gives a brief summary of the pros and cons of each of the different methods for quantification:

1. <u>Based on Influencing Factors</u>. Methods using specific influencing factors, associated guidance, and set multipliers as measures of the effects of influencing factors, may (at least in principle) better support the ability to reproduce results, compare results for different human actions, and lessen unwanted variability when implementing the method. However, because of the generally fixed approach of such methods, the ability to evaluate or even identify, for instance, other potentially relevant influencing factors not covered by the method, or account for interactions among the influencing factors, can be difficult and require modified use of the method or other compensations

with little guidance. In addition, most of these methods appear to need additional guidance on how to address scaling issues associated with the PSFs as a function of the context and PSF interactions.

2. Expert Judgment. Methods that more generally tend to employ a process whereby the analyst is more free to investigate the overall context associated with a human action and decide, through a systematic process, what influencing factors to address and how to weigh their effects, provide a level of flexibility desirable to ensure the most relevant factors and even interactions among the factors are indeed addressed. However, without prescribed or otherwise calibrated quantification guidance to fit the myriad combinations of factors that may come up, such flexibility may lead to greater analyst-to-analyst variability in results.
3. Empirically Based. Empirically based quantification can provide a level of credibility in the results that may be considered superior to analytical techniques. However, as a limitation, it is not practicable to obtain empirical evidence about each human action in the performance conditions that may be of interest for all types of sequences. This necessitates using limited empirical evidence for situations/sequences that were not simulated, potentially questioning the suitability of applying the information to these other situations; hence, the need for thoughtful use of the limited data and appropriate justification of its applicability wherever used.

No one approach is necessarily always better than another if good HRA practices are followed and appropriate information can be effectively incorporated into the application. Which approach is best to use will depend on the application and the potential tradeoffs involved – for example, how close the empirical evidence fits the situation being assessed or whether a method's list of treated influences captures those most relevant to the action being addressed.

5. CONCLUSION

The results of this investigation and development of good practices provided a basis for determining a path forward for addressing the SRM on HRA model differences by identifying the features needed in a HRA method and suggesting the development of a hybrid approach capturing the positive elements of many of the existing methods. The approach being developed is conceptually a hybrid of the ATHEANA [5-6] and CBDT [7] HRA methods and the IDA information processing model [20], along with the development of improved processes to address some of the limitations identified in the International Empirical Study [11].

The aim of the response to the SRM is to set the foundation and develop an HRA method that meets the desirable attributes to the extent possible (based on lessons learned over the years of applying and studying HRA methods) so that the reliability and validity of HRA can be improved to a more sufficient level. One of the objectives of the project is to demonstrate that the conceptual model adopted has these attributes. One caveat to note, however, is that some attributes (e.g., empirical validity) will need to be demonstrated through specific activities that may be outside the scope and resources of this effort. Furthermore, as with any product, accumulated experience should result in improvements of the method and associated application tools.

ACKNOWLEDGEMENTS

This work was funded by the U.S. Nuclear Regulatory Commission (USNRC) at Sandia National Laboratories. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. The University of Maryland participated under subcontract from Sandia.

This paper was prepared in part by employees of the USNRC. It presents information that does not currently represent an agreed-upon staff position. The USNRC has neither approved nor disapproved its technical content.

REFERENCES

[1] US Nuclear Regulatory Commission. *Staff Requirements – Meeting with Advisory Committee on Reactor Safeguards, SRM Mo61020.* US Nuclear Regulatory Commission, Washington, DC, 2006.

[2] Hendrickson, S. M. L., Parry, G. W., Forester, J. A., Dang, V. N., Whaley, A. M., Lewis, S., Lois, E., Xing, J. “Towards an Improved HRA Method.” Proceedings of the 11th International Conference on Probabilistic Safety Assessment and Management (PSAM11) (this conference).

[3] Kolaczkowski, A., Forester, J., Lois, E., & Cooper, S. “Good Practices for Implementing Human Reliability Analysis.” NUREG-1792, US Nuclear Regulatory Commission, Washington, DC, April 2005.

[4] Forester, J., Kolaczkowski, A., Lois, E., & Kelly, D. “Evaluation of Human Reliability Analysis Methods against Good Practices.” NUREG-1842, US Nuclear Regulatory Commission, Washington, DC, September 2006.

[5] NUREG-1624, Rev. 1. “Technical Basis and Implementation Guidelines for A Technique for Human Event Analysis (ATHEANA).” NUREG-1624, Rev. 1, US Nuclear Regulatory Commission, Washington, DC, May 2000.

[6] Forester, J., Kolaczkowski, A., Cooper, S., Bley, D., & Lois, E. “ATHEANA User’s Guide.” NUREG-1880, US Nuclear Regulatory Commission, Washington, DC, June 2007.

[7] Parry, G., Lydell, B. O Y., Spurgin, A. J., Moieni, P., & Beare, A. “An Approach to the Analysis of Operator Actions in PRA.” EPRI TR-100259, Electric Power Research Institute, Palo Alto, CA, 1992.

[8] Gertman, D. I, Blackman, H. S., Byers, J., Haney, L., Smith, C., & Marble, J. “The SPAR-H Method.” NUREG/CR-6883, US Nuclear Regulatory Commission, Washington, DC, August 2005.

[9] Swain, A. D. & Guttmann, H. E. “Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications.” NUREG/CR-1278/SAND80-0200, Sandia National Laboratories for the US Nuclear Regulatory Commission, Washington, DC, August 1983.

[10] Swain, A. D. “Accident Sequence Evaluation Program Human Reliability Analysis Procedure.” NUREG/CR-4772/SAND86-1996, Sandia National Laboratories for the US Nuclear Regulatory Commission, Washington, DC, February 1987.

[11] Lois, E., Dang, V. N., Forester, J., Broberg, H., Massaiu, S., Hildebrandt, M., Braarud, P. Ø., Parry, G., Julius, J., Boring, R., Männistö, I., & Bye, A. “International HRA Empirical Study—Phase 1 Report: Description of Overall Approach and Pilot Phase Results from Comparing HRA Methods to Simulator Data.” NUREG/IA-0216, Vol. 1, US Nuclear Regulatory Commission, Washington, DC, 2009.

[12] Forester, J., Hildebrandt, M., Broberg, H., Nowell, R., Liao, H., Dang, V. N., Presley, M., Bye, A., Marble, J., Lois, E., Hallbert, B., & Morgan, T. “US HRA Empirical Study – Comparison of HRA Method Predictions to Operating Crew Performance in a US Nuclear Power Plant Simulator and an Assessment of the Consistency of HRA Method Predictions.” Draft Report.

[13] Embrey, D. E., Humphreys, P., Rosa, E. A., Kirwan, B., & Rea, K. “SLIM-MAUD: An Approach to Assessing Human Error Probabilities Using Structured Expert Judgment.” NUREG/CR-3518, Vols. I & II, Brookhaven National Laboratory for the US Nuclear Regulatory Commission, Washington, DC, 1984.

[14] Chien, S. H., Dykes, A. A., Stetkar, J. W., & Bley, D. C. “Quantification of Human Error Rates Using a SLIM-Based Approach.” IEEE Fourth Conference on Human Factors and Power Plants, Monterey, CA, June 1988.

[15] Wakefield, D., Parry, G., Hannaman, G., & Spurgin, A. “SHARP1: A Revised Systematic Human Action Reliability Procedure.” EPRI TR-101711, Tier 2, Electric Power Research Institute, Palo Alto, CA, December 1992.

[16] ASME/ANS RA-Sa–2009, Addenda to ASME/ANS RA-S–2008, *Standard for Level 1/Large Early Release Frequency Probabilistic Risk Assessment for Nuclear Power Plant Applications*. The American Society of Mechanical Engineers, New York, NY, February 2009.

[17] US Nuclear Regulatory Commission. An Approach for Determining the Technical Adequacy of Probabilistic Risk Assessment Results for Risk-Informed Activities, Draft Regulatory Guide 1.200, US Nuclear Regulatory Commission, Washington, DC, February 2004.

[18] Bieder, C., Le-Bot, P., Desmares, E., Bonnet, J-L., & Cara, F. “MERMOS: EDF’s new Advanced HRA Method.” Proceedings of the 4th International Conference on Probabilistic Safety Assessment and Management (PSAM 4), New York, NY, September 1998.

[19] Julius, J., Grobbelaar, J., Spiegel, D., & Rahn, F. “The EPRI HRA Calculator® User’s Manual, Version 3.0.” Electric Power Research Institute, Palo Alto, CA, May 2005.

[20] Smidts, C., Shen, S. H., & Mosleh, A. “The IDA cognitive model for the analysis of nuclear power plant operator response under accident conditions. I. Problem solving and decision making model.” *Reliability Engineering & System Safety*, *55*(1), pp. 51-71, 1997.