NUREG/IA-0216, Vol. 1

International Agreement Report

International HRA Empirical Study – Phase 1 Report

Description of Overall Approach and Pilot Phase Results from Comparing HRA Methods to Simulator Performance Data

Prepared by:
E. Lois [1], V.N.Dang[2], J. Forester[3], H. Broberg[5],
S. Massaiu[5], M. Hildebrandt[5], P. Ø. Braarud[5],
G. Parry[1], J. Julius[4], R. Boring[3], I. Männistö[5],
A. Bye[5]

[1] U.S. Nuclear Regulatory Commission, USA
[2] Paul Scherrer Institute, Switzerland
[3] Sandia National Laboratories, USA
[4] Scientech, USA
[5] OECD Halden Reactor Project, Norway

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

November 2009

Prepared for
U.S. Nuclear Regulatory Commission (United States)
OECD Halden Reactor Project (Norway)

Published by
U.S. Nuclear Regulatory Commission

AVAILABILITY OF REFERENCE MATERIALS IN NRC PUBLICATIONS

NRC Reference Material

As of November 1999, you may electronically access NUREG-series publications and other NRC records at NRC's Public Electronic Reading Room at http://www.nrc.gov/reading-rm.html. Publicly released records include, to name a few, NUREG-series publications; *Federal Register* notices; applicant, licensee, and vendor documents and correspondence; NRC correspondence and internal memoranda; bulletins and information notices; inspection and investigative reports; licensee event reports; and Commission papers and their attachments.

NRC publications in the NUREG series, NRC regulations, and *Title 10, Energy*, in the Code of *Federal Regulations* may also be purchased from one of these two sources.

1. The Superintendent of Documents
   U.S. Government Printing Office
   Mail Stop SSOP
   Washington, DC 20402–0001
   Internet: bookstore.gpo.gov
   Telephone: 202-512-1800
   Fax: 202-512-2250
2. The National Technical Information Service
   Springfield, VA 22161–0002
   www.ntis.gov
   1–800–553–6847 or, locally, 703–605–6000

A single copy of each NRC draft report for comment is available free, to the extent of supply, upon written request as follows:

Address: U.S. Nuclear Regulatory Commission
Office of Administration
Mail, Distribution and Messenger Team
Washington, DC 20555-0001
E-mail: DISTRIBUTION@nrc.gov
Facsimile: 301–415–2289

Some publications in the NUREG series that are posted at NRC's Web site address http://www.nrc.gov/reading-rm/doc-collections/nuregs are updated periodically and may differ from the last printed version. Although references to material found on a Web site bear the date the material was accessed, the material available on the date cited may subsequently be removed from the site.

Non-NRC Reference Material

Documents available from public and special technical libraries include all open literature items, such as books, journal articles, and transactions, *Federal Register* notices, Federal and State legislation, and congressional reports. Such documents as theses, dissertations, foreign reports and translations, and non-NRC conference proceedings may be purchased from their sponsoring organization.

Copies of industry codes and standards used in a substantive manner in the NRC regulatory process are maintained at—

The NRC Technical Library
Two White Flint North
11545 Rockville Pike
Rockville, MD 20852–2738

These standards are available in the library for reference use by the public. Codes and standards are usually copyrighted and may be purchased from the originating organization or, if they are American National Standards, from—

American National Standards Institute
11 West 42nd Street
New York, NY 10036–8002
www.ansi.org
212–642–4900

Legally binding regulatory requirements are stated only in laws; NRC regulations; licenses, including technical specifications; or orders, not in NUREG-series publications. The views expressed in contractor-prepared publications in this series are not necessarily those of the NRC.

The NUREG series comprises (1) technical and administrative reports and books prepared by the staff (NUREG–XXXX) or agency contractors (NUREG/CR–XXXX), (2) proceedings of conferences (NUREG/CP–XXXX), (3) reports resulting from international agreements (NUREG/IA–XXXX), (4) brochures (NUREG/BR–XXXX), and (5) compilations of legal decisions and orders of the Commission and Atomic and Safety Licensing Boards and of Directors' decisions under Section 2.206 of NRC's regulations (NUREG–0750).

NUREG/IA-0216, Vol. 1

International Agreement Report

International HRA Empirical Study – Phase 1 Report

Description of Overall Approach and Pilot Phase Results from Comparing HRA Methods to Simulator Performance Data

Prepared by:
E. Lois [1], V.N.Dang[2], J. Forester[3], H. Broberg[5], S. Massaiu[5], M. Hildebrandt[5], P. Ø. Braarud[5], G. Parry[1], J. Julius[4], R. Boring[3], I. Männistö[5], A. Bye[5]

[1] U.S. Nuclear Regulatory Commission, USA
[2] Paul Scherrer Institute, Switzerland
[3] Sandia National Laboratories, USA
[4] Scientech, USA
[5] OECD Halden Reactor Project, Norway

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

November 2009

Prepared for
U.S. Nuclear Regulatory Commission (United States)
OECD Halden Reactor Project (Norway)

Published by
U.S. Nuclear Regulatory Commission

ABSTRACT

Volume 1 of NUREG/IA-0216 documents the Pilot Phase of the International Human Reliability Analysis (HRA) Empirical Study. This three-phase study is a multinational, multiteam effort supported by the Organization for Economic Cooperation and Development (OECD) Halden Reactor Project, the Swiss Federal Nuclear Safety Inspectorate, the U.S. Electric Power Research Institute, and the U.S. Nuclear Regulatory Commission (NRC). The Pilot has also been documented as a Halden publication: HWR-844, October 2009.

The objective of this study is to develop an empirically based understanding of the performance, strengths, and weaknesses of different HRA methods used to model human response to accident sequences in probabilistic risk assessments (PRAs). The empirical basis was developed through experiments performed at the Halden Reactor Project HAMMLAB (HAlden huMan-Machine LABoratory) research simulator, with real crews responding to accident situations similar to those modeled in PRAs. The scope of the study is limited to HRA methods thought appropriate for use in PRAs evaluating internal events during full power operations of current light water reactors. The study consists of performing HRAs for predefined human actions, with different HRA teams using different methods. Nuclear power plant crews perform these human actions at the Halden simulator, Halden experimentalists collect and interpret the data to fit HRA data needs, and an independent group of experts compare the results of each HRA method/team to the Halden crew performance data.

The Pilot Phase consisted of developing, testing, and revising the study's methodology and design. Phase 2, which will be documented in Volume 2, consists of the comparison of HRA predictions for all nine steam generator tube rupture human actions. Phase 3, which will be documented in Volume 3, consists of the comparison of four loss-of-feedwater human actions, as well as documentation of the overall study results. The results of the Empirical Study will provide a technical basis for improving individual methods, improving existing guidance documents for performing and reviewing HRAs (e.g., NUREG-1792, HRA Good Practices), and developing additional guidance and training materials for implementing individual methods.

FOREWORD

Volume 1 of NUREG/IA-0216 documents the Pilot Phase of the International Human Reliability Analysis (HRA) Empirical Study. This three-phase study is a multinational, multiteam effort supported by the Organization for Economic Cooperation and Development (OECD) Halden Reactor Project. The Project provided facilities, crews, and expertise to collect and analyze simulator crew performance data, and, with HRA teams from multiple organizations, used various methods to analyze and predict the performance of these crews. Halden's signatory organizations provided analyst teams to perform HRA of the simulated human actions. The Swiss Federal Nuclear Safety Inspectorate, the U.S. Electric Power Research Institute, and the U.S. Nuclear Regulatory Commission (NRC) and its contractor, Sandia National Laboratory, compared the HRA method predictions with the empirical data generated at Halden.

The objective of this study is to develop an empirically based understanding of the performance, strengths, and weaknesses of different HRA methods used to model human response to accident sequences in probabilistic risk assessments (PRAs), particularly as applied in PRAs for internal events analysis at full power for the current light water reactors. The widespread use of different HRA methods within PRA, combined with the potential impact of the different methods on PRA results, led to the focus on this specific application.

The Pilot Phase consisted of developing, testing, and revising the study's methodology and experimental design, which involved designing simulator experiments for assessing HRA methods (i.e., developing PRA-type scenarios and selecting and defining human failure events for analysis); developing a methodology for collecting human performance data/observations for comparison with the human performance perspectives considered by analysts; developing an information package with plant and crew information needed for HRA; HRAs of the simulated human actions, performed by different HRA teams using different methods; and comparing the crew performance with the HRA predictions. These facets of the study were applied to human actions from two different steam generator tube rupture (SGTR) case scenarios, one simple and one complex. This NUREG/IA report documents these results.

Lessons learned in the Pilot Phase (Phase 1) were used to improve the data analysis to be incorporated in the evaluation of the remaining SGTR human actions (Phase 2) and to improve the design of the study incorporated in Phase 3, which will involve loss-of-feedwater (LOFW) scenarios. The Pilot Phase also included an initial evaluation of the HRA methods. Based on these initial findings, insights were developed for improving HRA methods and practices, as well as HRA-focused simulator experiments.

Volume 2 will document the findings from the comparison of HRA predictions to crew performance for all nine SGTR human actions, while Volume 3 will document the findings for the four LOFW human actions, as well as results for the overall study.

The results of the Empirical Study will provide a technical basis for improving individual methods, improving existing guidance documents for performing and reviewing HRA (e.g., NUREG-1792, HRA Good Practices), and developing additional guidance and training materials for implementing individual methods. Moreover, the results of this study will provide a technical basis to support the work that addresses the Commission's Staff Requirements Memorandum (SRM)-M061020, which directed the staff to address the issue of HRA model differences—including examining whether the NRC could adopt a single model for all HRA applications, or whether it should adopt more than one—and to provide explicit guidance on the applicability and implementation of each model.

Christiana Lui, Director
Division of Risk Analysis
Office of Nuclear Regulatory Research

TABLE OF CONTENTS

ABSTRACT iii
FOREWORD v
EXECUTIVE SUMMARY xiii
ACKNOWLEDGMENTS xxiii
ABBREVIATIONS xxvii
1. OVERALL STUDY DESIGN AND METHODOLOGY 1-1
1.1 Background 1-1
1.2 Overview of the Study Design 1-2
1.2.1 Four High-Level Study Tasks 1-2
1.2.2 Study Organization, Participants, and Roles 1-3
1.2.3 Phases of the Empirical Study 1-3
1.3 Study Design: The Overall Methodology 1-4
1.3.1 Scenario Design 1-4
1.3.2 Predictive HRAs 1-5
1.3.2.1 Inputs to HRA Teams 1-5
1.3.2.2 Interaction of HRA and Accident Sequence Modeling 1-6
1.3.2.3 Documentation of HRAs 1-6
1.3.3 Empirical Data Collection and Analysis 1-7
1.3.4 Comparison Methodology 1-8
1.4 Specifics of the Pilot Study 1-9
1.4.1 HRA Methods Represented in the Pilot Study 1-9
1.4.2 Implementation of the Methodology in the Pilot Study 1-10
1.4.3 Review of Methodology and Results in a Workshop 1-11
1.4.4 Inter-Analyst Team and Inter-Method Consistency 1-11
1.4.5 Details of the Information Package Given to the HRA Teams 1-11
1.4.6 Experimental Method and Measures 1-12
1.5 Overview of the Remainder of this Report 1-13
2. EMPIRICAL DATA: HALDEN DATA COLLECTION AND ANALYSIS 2-1
2.1 Introduction and Overview 2-1
2.2 Methodology 2-1
2.3 Simulator Study Design and Experimental Procedure 2-2
2.3.1 Scenario Presentation Order 2-2
2.3.2 Participants' Daily Schedule 2-3
2.3.3 Operator Training in HAMMLAB 2-3
2.4 Characterization of Crews, Work Practices, and Training 2-3
2.4.1 Crew Organization 2-3
2.4.2 Crew Experience 2-4

2.4.3 Leadership Styles, Team Orientation, Crew Dynamics, Communication Style....... 2-5
2.4.4 Use of and Adherence to Procedures........ 2-5
2.4.5 Criteria for Having Crew Meetings During Accident Scenarios........ 2-6
2.4.6 Scenario-Relevant Training........ 2-6
2.5 Scenario Description........ 2-6
2.5.1 Operating Procedures........ 2-6
2.5.2 SGTR Base Case Scenario........ 2-7
2.5.3 SGTR Complex Case Scenario........ 2-8
2.5.4 SGTR HFE Definitions and Event Tree........ 2-8
2.6 Simulator Data Analysis Methodology........ 2-11
2.6.1 Initial Data Screening........ 2-12
2.6.2 Selection of Crews for In-Depth Analysis........ 2-12
2.6.3 Operational Stories and Influencing Factor Identification........ 2-12
2.6.4 Identification of Influencing Factors, Modified HERA Form........ 2-14
2.7 Overall Performance of Individual Crews and Aggregated Operational Stories ... 2-15
2.7.1 Overall Performance of Individual Crews on HFE 1A (SGTR Base Case)........ 2-15
2.7.2 Overall Performance of Individual Crews on HFE 1B (SGTR Complex Case)..... 2-15
2.7.3 Aggregated Operational Story on HFE 1A (SGTR Base Case, All Crews)........ 2-16
2.7.4 Aggregated Operational Story on HFE 1B (SGTR Complex Case, All Crews)..... 2-18
2.8 Identification of Driving Factors–Aggregated Across All Crews........ 2-22
2.8.1 Performance Shaping Factors (PSFs) Used in the Empirical Study........ 2-22
2.8.1.1 Background to PSF Selection........ 2-22
2.8.1.2 PSFs Used for Comparison and Data Analysis........ 2-26
2.8.1.3 PSFs Used in First Pilot Phase–Working Definitions........ 2-26
2.8.2 Rating of Driving Factors........ 2-29
2.8.3 Methodology for the Identification of Driving Factors........ 2-30
2.8.4 Influencing Factor Matrices for HFE 1A and HFE 1B........ 2-33
2.8.5 Driving Factors for Performance on HFE1A (Base Case)........ 2-36
2.8.6 Driving Factors for Performance on HFE 1B (Complex Case)........ 2-37
2.9 Discussion........ 2-41
2.9.1 Issues on Empirical Information Aggregation and Formatting........ 2-41
2.9.1.1 Failure vs. Performance........ 2-41
2.9.1.2 The Difficult Treatment of "Variable" PSFs........ 2-42
2.9.1.3 The Interaction Between PSFs: Factor Models vs. Models of Observed Influences2-43
2.9.1.4 Interpretation of Driving Factors........ 2-44

3. COMPARISON OF METHODS TO DATA........ 3-1
3.1 Comparison Methodology........ 3-1
3.1.1 Summarizing the HRAs—the HRA Submittals and Predicted Outcomes........ 3-1
3.1.2 Criteria for the Comparison of HRA Submittals vs. Empirical Data........ 3-2
3.2 Comparison Results........ 3-4
3.2.1 ASEP/THERP (NRC)........ 3-4

3.2.1.1 Short Overviews of the ASEP and THERP Methods 3-4
3.2.1.2 Main References to the ASEP and THERP Methods 3-6
3.2.1.3 NRC, ASEP/THERP Analysis of HFE 1A 3-6
3.2.1.4 NRC, ASEP/THERP Analysis of HFE 1B 3-9
3.2.1.5 Additional Comments on NRC ASEP/THERP Analysis 3-12
3.2.1.6 NRC ASEP/THERP Team Comments on the Original Comparison 3-13
3.2.2 CBDTM + THERP (EPRI) 3-13
3.2.2.1 Short Overview of the CBDT Method 3-13
3.2.2.2 Main References to the CBDTM + THERP Method 3-15
3.2.2.3 EPRI, CBDTM + THERP Analysis of HFE 1A 3-15
3.2.2.4 EPRI, CBDTM + THERP Analysis and Comparison, HFE 1B 3-18
3.2.2.5 Additional Comments on CBDTM + THERP Analysis 3-20
3.2.2.6 CBDTM + THERP Team Comments on the Original Comparison 3-21
3.2.3 SPAR-H (INL) 3-21
3.2.3.1 Short Overview of the SPAR-H Method 3-21
3.2.3.2 Main References to the SPAR-H Method 3-22
3.2.3.3 INL, SPAR-H Analysis of HFE 1A 3-22
3.2.3.4 INL SPAR-H Analysis of HFE 1B 3-25
3.2.3.5 Additional Comments on INL SPAR-H Analysis 3-28
3.2.3.6 INL SPAR-H Team Comments on the Original Comparison 3-30
3.2.4 SPAR-H (NRC) 3-30
3.2.4.1 NRC, SPAR-H Analysis of HFE 1A 3-30
3.2.4.2 NRC, SPAR-H Analysis of HFE 1B 3-32
3.2.4.3 Additional Comments on NRC SPAR-H Analysis 3-35
3.2.4.4 NRC SPAR-H Team Comments on the Original Comparison 3-36
3.2.5 CESA-Q (PSI) 3-36
3.2.5.1 Short Overview of the CESA-Q Quantification Method (by Luca Podofillini and Bernhard Reer) 3-36
3.2.5.2 Main References to the CESA-Q Method 3-37
3.2.5.3 PSI, CESA-Q Analysis of HFE 1A 3-37
3.2.5.4 PSI, CESA-Q Analysis of HFE 1B 3-40
3.2.5.5 Additional Comments on PSI, CESA-Q Analysis 3-43
3.2.5.6 PSI CESA-Q Team Comments on the Original Comparison 3-45
3.2.6 Decision Trees + ASEP (NRI) 3-45
3.2.6.1 Short Overview of the NRI "Decision Trees + ASEP" Method (by Jaroslav Holy). 3-45
3.2.6.2 Main References to the "Decision Trees + ASEP" Method 3-46
3.2.6.3 NRI, DT + ASEP Analysis of HFE 1A 3-47
3.2.6.4 NRI, DT + ASEP Analysis of HFE 1B 3-50
3.2.6.5 Additional Comments on NRI DT+ASEP Analysis 3-53
3.2.6.6 NRI DT+ASEP Team Comments on the Original Comparison 3-54
3.2.7 MERMOS (EDF) 3-54
3.2.7.1 Short Overview of the MERMOS Method (by Pierre Le Bot) 3-54
3.2.7.2 Main References to the MERMOS Method 3-55
3.2.7.3 EDF, MERMOS Analysis of HFE 1A 3-56
3.2.7.4 EDF, MERMOS Analysis of HFE 1B 3-59
3.2.7.5 Additional Comments on the EDF MERMOS Analysis 3-63
3.2.7.6 EDF MERMOS Team Comments on the Original Comparison 3-65
3.2.8 HEART (Ringhals) 3-66

3.2.8.1 Short Overview of the HEART Method (by Steve Collier) 3-66
3.2.8.2 Main References to the HEART Method 3-68
3.2.8.3 Ringhals, HEART Analysis of HFE 1A 3-68
3.2.8.4 Ringhals, HEART Analysis of HFE 1B 3-70
3.2.8.5 Additional Comments on the Ringhals, HEART Analysis 3-72
3.2.8.6 Ringhals, HEART Team Comments on the Original Comparison 3-73
3.2.9 PANAME (IRSN) 3-73
3.2.9.1 Short Overview of the PANAME Method (by Véronique Fauchille) 3-73
3.2.9.2 Main References to the PANAME Method 3-74
3.2.9.3 IRSN, PANAME Analysis of HFE 1A 3-74
3.2.9.4 IRSN PANAME Analysis of HFE 1B 3-76
3.2.9.5 Additional Comments on the IRSN PANAME Analysis 3-80
3.2.9.6 IRSN PANAME Team Comments on the Original Comparison 3-80
3.2.10 Enhanced Bayesian THERP (VTT) 3-81
3.2.10.1 Short Overview of the VTT-Enhanced Bayesian THERP Method (by Jan-Erik Holmberg) 3-81
3.2.10.2 Main References to the Enhanced Bayesian THERP Method 3-82
3.2.10.3 VTT, Bayesian THERP Analysis of HFE 1A 3-82
3.2.10.4 VTT, Bayesian THERP Analysis of HFE 1B 3-84
3.2.10.5 Additional Comments on the VTT THERP Analysis 3-87
3.2.10.6 VTT THERP Team Comments on the Original Comparison 3-88
3.2.11 ATHEANA (NRC) 3-88
3.2.11.1 Short Overview of the ATHEANA Method 3-88
3.2.11.2 Main References to the ATHEANA Method 3-89
3.2.11.3 NRC, ATHEANA Analysis of HFE 1A 3-89
3.2.11.4 NRC, ATHEANA Analysis and Comparison, HFE 1B 3-92
3.2.11.5 Additional Comments on NRC ATHEANA Analysis 3-97
3.2.11.6 NRC ATHEANA Team Comments on the Original Comparison 3-97
3.2.12 K-HRA (KAERI) 3-98
3.2.12.1 Short Overview of the K-HRA Method (by Wondea Jung) 3-98
3.2.12.2 Main References to the K-HRA Method 3-100
3.2.12.3 KAERI, K-HRA of HFE 1A 3-100
3.2.12.4 KAERI, K-HRA of HFE 1B 3-103
3.2.12.5 Additional Comments on KAERI K-HRA 3-106
3.2.12.6 KAERI K-HRA Team Comments on the Original Comparison 3-107
3.2.13 CREAM (NRI) 3-107
3.2.13.1 Short Overview of the CREAM Method (by Jaroslav Holy, NRI) 3-107
3.2.13.2 Main References to the CREAM Method 3-108
3.2.13.3 NRI, CREAM Analysis of HFE 1A 3-108
3.2.13.4 NRI, CREAM Analysis of HFE 1B 3-111
3.2.13.5 Additional Comments on NRI CREAM Analysis 3-113
3.2.13.6 NRI CREAM Team Comments on the Original Comparison 3-114
3.3 Comparison Summary 3-114
3.3.1 The Nature of PSFs-Insights from the Empirical Data 3-114
3.3.2 Implications for the Comparison with HRA Predictions 3-115
3.3.3 Implications for the Conclusions to be Drawn from this Comparison 3-116
3.3.4 Implications for Model Improvement and the Next Phase of this Research 3-116

4. CONCLUSION AND OUTLOOK 4-1
4.1 Overview of Results of Comparison of Methods to Data 4-1
4.2 Observations on the Methodology Used in the Pilot Study 4-3
4.3 Upcoming Work: Pilot Study Second Phase and LOFW Study 4-4
5. REFERENCES 5-1
APPENDIX A 1
CREW SUMMARIES 1
APPENDIX B 1
FORM A 1
FORM B 0
APPENDIX C 1
FORM B 1
APPENDIX D 0
FORM B INSTRUCTIONS 0

EXECUTIVE SUMMARY

Background

The Office of Nuclear Regulatory Research (RES) of the U.S. Nuclear Regulatory Commission (NRC) supported the initiation and execution of a research project that would develop an empirical basis for evaluating human reliability analysis (HRA) methods. This project is an international collaborative effort, and involves the use of the Organization for Economic Co-Operation and Development (OECD) Halden Reactor Project HAMMLAB (HAlden huMan-Machine LABoratory) research simulator, a full-scope nuclear power plant simulator located in Halden, Norway. The study aims to develop an empirically based understanding of the performance, strengths, and weaknesses of different HRA methods. The empirical basis is developed through experiments performed at the HAMMLAB simulator, with real crews responding to accident situations similar to those modeled in probabilistic risk assessments (PRAs). The scope of the study is limited to HRAs for internal events during full-power operations of current light water reactors. The results will provide a technical basis for the development of improved HRA guidance, and, if necessary, improved HRA methods.

Overview of the Study Design

Four High-Level Study Tasks

The International HRA Empirical Study (hereafter called "Empirical Study") focused on the HRA of control room personnel actions required in response to PRA-initiating events. The study consisted of four high-level tasks:

Task 1. Definition of the scenarios and the human failure events (HFEs) to be analyzed and compilation of information packages for the HRA teams

Task 2. Predictive analysis of the HFEs by different HRA teams applying different methods.

Task 3. Production of the empirical data through the collection of raw crew performance data from the simulator runs and subsequent HRA-oriented data analysis and aggregation.

Task 4. Review of the HRA submittals, comparison of HRA predictions to the empirical data, and development of insights for improving HRA methods and HRA practices.

Study Organization, Participants, and Roles

The description of the Empirical Study high-level tasks shows that there was a simulator study embedded within the overall study (Task 3). There were therefore four sets of study participants:

- **Halden experimental staff** (Tasks 1, 3), who were involved in the design and execution of the simulator experiments at the HAMMLAB research facility. The HAMMLAB staff, in collaboration with PRA experts, was involved in the design of the experiments and was largely responsible for analyzing the experimental data.

- **HRA teams** (Task 2), which applied different HRA methods to obtain predictions for the human failure events (HFEs) in the scenarios defined for the study. Organizations representing industry, regulators, and the research community participated.
- **Operator crews** (Task 3), who performed the simulator runs at the HAMMLAB simulator. Each crew responded to four scenarios consisting of one simple/base and one "complex" variant for two different scenario types.
- **Assessment and evaluation group** (Overall responsibility and Task 4), which was responsible for the organization and implementation of the study, including experiment design, information package preparation (analysis input) for the HRA teams, and reviewing and evaluating the HRA submittals.

To avoid bias in the comparison, a "blind" study protocol was used. That is, the operator crews had no prior knowledge of the scenarios. Likewise, the assessment and evaluation group did not receive any information about the actual crew performances in HAMMLAB until after the HRA submittal results were summarized by group members and reviewed by the pertinent HRA teams. In addition, the Halden staff analyzed and documented the crew performance data without knowledge of the HRA predictions.

Phases of the Empirical Study

The Empirical Study is performed in three phases. Phase 1, the Pilot, consists of the development and testing of the methodology. Due to scheduling issues, the HRA teams received information and analyzed all nine HFEs involved in the two variants of steam generator tube rupture (SGTR) scenarios; however, for the purposes of the Pilot, empirical data were developed and compared to HRA predictions for only two of the HFEs (HFE1A and 1B in Table ES-1). Furthermore, the pilot was limited to the comparison and evaluation of the qualitative portion of the HRAs. In phase 2 and 3, the quantitative results are also used; the predicted HEPs are used as a measure of the difficulty of the HFEs and are compared to a ranking of the HFEs developed from the empirical data. These results will be presented Volumes 2 and 3 documenting the work of Phases 2 and 3.

Table ES-1. Human Failure Events (HFEs) in the SGTR Scenarios

Simple/Base Case		Complex Case	
HFE1A	Identify/isolate SGTR	HFE1B	Identify/isolate SGTR w/masked indicator
HFE2A	Cool down RCS	HFE2B	Cool down RCS (Same as 2A)
HFE3A	Depressurize RCS (PRZ spray or PORV)	HFE3B	Depressurize RCS (No PRZ spray, PORV only)
HFE4A	Terminate SI (Stop spray, close PORV)	—	(PORV stuck open)
—		HFE5	Try to close PORV / \ 5B1: PORV indicator shows closed 5B2: PORV indicator shows open

The pilot phase allowed the study participants (Halden staff, assessment/evaluation group, and HRA teams) to review and revise the methodology. A particular emphasis was given to address the HRA teams' feedback given during the October 2007 workshop in Washington, D.C., and by reviewing draft versions of this document. Furthermore, the pilot was externally reviewed; recommendations from this review will also be incorporated into Phases 2 and 3.

Phase 2, the SGTR Study, consists of the development of the empirical data, review of HRA submittals and comparisons of HRA predictions to the empirical data for the remaining seven SGTR human actions; it also consists of the , and documenting the findings and insights for the whole SGTR scenario. Phase 3, consists of a similar analysis of the four loss-of-feedwater (LOFW) human actions, and the documentation of the overall perspectives gained from this study.

The Overall Methodology

Scenario Design

Full-power PRA scenarios have been used in the Empirical Study. In PRAs, the post-initiator operator actions, assessed with HRA methods, are frequently associated with postulated scenarios that are beyond the design basis of nuclear power plants. For instance, PRA scenarios often include multiple equipment failures. The use of simulator data is therefore necessary, due in part to the infrequency of these multiple failure scenarios; however, designing and implementing these scenarios in simulator studies is challenging because they must postulate component and system failures in combinations that lead to the required actions of interest while remaining plausible to the operators. Such scenarios were designed and utilized in this study, thereby providing information on how well HRA methods can identify factors that may lead to crew failure, and, to a lesser degree, on how well they can estimate failure probabilities.

To provide a varied set of situations for the crews, several scenarios were necessary; thus, as illustrated in Table ES-1, both simple/base and complex variants for each SGTR and LOFW scenario were designed. This was particularly useful in the context of the Empirical Study, because the two variants included similar or related tasks that differed only in terms of their performance contexts. The use of a simple/base variant provided a baseline performance for each scenario. The subsequent use of a complex variant therefore allowed for a comparison and an analysis of the differences in performance in order to determine the effects of the change in context difficulty. To control for order effects, such as learning and other potential biases, the scenarios were presented to the crews in a semi-randomized order. This approach provides a more complete understanding of the crew actions than would be allowed by individually examining unrelated scenarios. It also allows for an evaluation of whether the HRA methods are sensitive to such scenario differences, and whether their predictions are adjusted accordingly.

Approach for Predictive HRA

In principle, the HRAs required for the Empirical Study are no different from those performed for a PRA. In practice, however, the study methodology needed to address three issues that arose due to limitations in the study.

Inputs to HRA Teams

A prerequisite for HRA within a PRA is analyst familiarity with the background, training, and experience of the performers (the crews), as well as the performance conditions, including human-system interface (HSI) and the availability of job aids (e.g., guidance, procedures, etc.). Although much of this information could be given directly to the HRA teams, there were limited opportunities for HRA teams to perform familiarization tasks, such as plant visits, observations of the crews, task walk-throughs, and interviews with crews or training personnel. To compensate for this, the assessment group compiled an information package that included as much of this information as possible, and the HRA teams had the opportunity to request and receive additional information in a question-and-answer process.

Reporting of HRAs and Predicted Outcomes

HRA methods differ in many aspects, including but not limited to the type of analysis for which they are intended (e.g., full power internal event PRA); the underlying human behavior models and terminology they use; and the number and types of performance shaping factors (PSFs) they may potentially utilize to estimate human error probabilities (HEPs). To address issues related to these differences in HRA methods, analysts were asked to provide more detailed documentation of their qualitative analysis than is typically provided in a PRA. Specifically, the HRA teams were asked to document their analysis using the following three methods:

- Form A (see Appendix B), an "open-form" questionnaire where, for each HFE, the teams were asked to report (1) the estimated HEP, (2) the identified driving factors (i.e., factors that can contribute either to success or failure), and (3) associated "operational expressions" (i.e., a description of how crews are expected to deal with the identified factors for the specific scenarios).
- "Normal" documentation of their HRAs and quantification, as in a PRA.
- Form B (see Appendices C and D), a "closed-form" questionnaire based on the taxonomy from the Human Event Repository and Analysis (HERA) system (NUREG/CR-6903, Vols. 1-2), with adaptations for the study.

Form B was utilized in an attempt to "standardize" the predictions in a common terminology, with a predefined taxonomy. However, because the evaluation team identified many terminology mapping issues, they did not extensively use the information provided in Form B.

Comparison Methodology

The outcomes predicted in the HRAs performed by the teams were compared with the outcomes obtained from the HAMMLAB experiments. Analytical predictions were compared with experimental outcomes for each of the following elements of response Form A:

- The level of difficulty associated with the operator actions of interest (i.e., those associated with the HFEs); for the HRA predictions, the level of difficulty is represented by the HEP.
- The factors that most influence the performance of the crews in these scenarios (i.e., "driving factors").
- The reason for the level of ease or difficulty with which the crews perform the tasks associated with each HFE, and how these difficulties are expressed in operational and scenario-specific terms (i.e., "operational expressions").

Several other criteria were also evaluated:

- The insights given by the HRA method for error reduction.
- Sensitivity issues, such as the impact of qualitative choices made by the analysts on the HEP.
- Traceability of the analysis, an important aspect of HRA.

The design of the study anticipated that the HAMMLAB experiments would not support the derivation of HFE failure probabilities from the experimental data. Although large for a simulator study, the number of sessions and crews (sample size) remains small in relation to the expected levels of performance of the crews. The study featured 14 actual crews, but most HRA methods would estimate for the simple HFEs average failure probabilities of about 1E-3, or 1 per 1000 occurrences. As a result, the comparative analyses in the pilot phase intentionally focus on the qualitative insights rather than on the quantitative results obtained with the methods or on the ability of the methods to predict the tendencies of behavior and performance in the scenarios. The comparison of the quantitative HRA predictions will be addressed in the second phase. It is worth noting here that the pilot phase of the Empirical Study was aimed primarily at establishing the methodology, not at comparative analysis.

HRA Methods in the Empirical Study

Table ES-2 summarizes the HRA methods that were evaluated in this study, along with the organizations that supported the HRA teams that agreed to participate. Summary descriptions of each method are in Chapter 4, Comparison of Methods to Data. Note that SPAR-H was applied by two teams, and that one team (NRI) used two methods. Note also that the "NRC staff and consultants" teams are three separate teams.

Table ES-2. HRA Methods in the Pilot Study[1]

Method: HRA Team					
ASEP/THERP	NRC staff and consultants, USA	CBDT	EPRI (Scientech), USA	HEART	Vattenfall & Ringhals NPP, Sweden
THERP with Bayesian Enhancement	VTT, Finland	Decision Trees + ASEP	NRI, Czech Rep.	KHRA	KAERI, Korea
ATHEANA	NRC staff and consultants, USA	MERMOS	EDF, France	CREAM	NRI, Czech Rep.
SPAR-H	NRC staff and consultants, USA	PANAME	IRSN, France	CESA	PSI, Switzerland
	Idaho National Laboratory, USA				

Initial Findings and Conclusions

This pilot study provided the opportunity to develop a methodology for comparing crew performance data to HRA predictions and identify issues related to the experimental design and data analysis for future phases of the study. However, the pilot also points to areas of improvements in HRA methods and practices.

Preliminary Results from the Comparison of Methods to Data

This section summarizes preliminary findings regarding strengths and weaknesses of HRA in general as well as of HRA methods and method implementation.

- All HRA methods identified some of the important factors driving performance in the SGTR scenarios. Thus, from an overall PRA perspective, the pilot demonstrated that existing HRA methods, if appropriately applied, are capable of identifying important underlying drivers of human success or failure and can therefore identify potential areas for safety improvement.
- Different methods address different performance driving factors in terms of PSFs or other causal factors. The importance of these differences became more evident in the complex scenarios where it appeared that analysts were limited by their methods to characterize crew performance. For example, some methods include a limited set of PSFs, which appeared to constrain analysts to incorporate all factors identified in the qualitative analysis into the HEP estimation.
- Differences in the level or nature of the analysis performed by the HRA analysts to understand the scenario and the factors likely to affect the crews' performance also appeared to contribute to differences in the results. Therefore, an early lesson from the

[1]In addition to the listed teams performing analyses using HRA methods, three teams have used the data as input for their simulation models to test the applicability of these methods in such a setting: QUEST-HP (Risø, Denmark), Microsaint (Alion, USA), and IDAC (University of Maryland, USA). Additionally, another team has used the data to test a selection algorithm (Politecnico di Milano, Italy).

pilot is that HRA methods should include guidance on how to analyze scenario characteristics to assist the analysts in understanding the cognitive and execution demands on operating crews. Furthermore, it appeared that many of the methods could benefit from additional guidance on how to accommodate these qualitative insights into the quantification of the HEPs.

- The evaluations of the degree to which a driver influences performance appear to be another important issue. The judgments involved can be difficult in many cases; also, there is evidence that drivers do interact. The pilot indicates that HRA results of can be sensitive to these sometimes subtle judgments. Frequently, the guidance provided by methods on the evaluation of the strength of an influencing factor is limited. Furthermore, although most methods note that driver interactions can be important, they do not provide guidance on how they should be handled. It is noted however that, although some methods may provide better guidance for evaluating the strength of a factor and handling factor-interactions, this is an issue that goes beyond individual methods.
- The pilot indicates that crew factors, such as team dynamics, work processes, communication strategies, sense of urgency, and willingness to take knowledge-based actions, can have significant effects on crew performance. Since the effects from these factors can be moderated or reinforced by other crew characteristics and/or situational features, they can be positive for some crews and negative for others. While such factors are certainly worth investigating in the context of an HRA, the effects may often have to be evaluated using sensitivity analyses on the HRA results, since the variability of these factors is not normally evaluated by most HRA methods.

Observations on the Study Methodology

The study developed a methodology for collecting crew performance observations suitable for comparisons to HRA results and demonstrated that it is possible to benchmark HRA methods using simulator experiments. Observations on the methodology and related issues included:

- Many issues related to the use of simulator data can be addressed through thoughtful experiment design, relative to the needs of HRA methods. For example, the perceived differences between human actions analyzed in a PRA versus that in a simulated scenario can be addressed. Constraints related to the use of the simulator can be identified a priori and documented in the "information package" provided to the HRA analysts, thereby ensuring that they understand the intricacies of the simulated scenario and focus their analysis on the simulated actions. In this way, the pilot study examined the application of HRA methods within a well-defined context that enabled comparison of method results to data.
- The study allowed the development of a detailed understanding of how HRA methods are applied. In particular, since all analyst teams were given the same information package, this experiment demonstrated how and to what extent analysts use the information provided in applying their method. This is an important aspect of the study, since, based on the comparison, insights can be developed with respect to the ability of

HRA methods to consider both the relevance of certain information and its potential impact on the HRA results.

- Although the challenging complex scenarios simulated in the study may be of low probability, the pilot shows that comparison of data obtained from such scenarios with corresponding HRAs can provide important insights into the methods.
- Since, with only one exception, one HRA team applied each method, the pilot as well as the Empirical Study in its entirety, will not be able to separate analysts' effects from methods' effects. To separate these effects, future studies should include multiple teams per HRA method and control for analysts' experience level. Such comparisons will help addressing the issue of analyst-to-analyst variability in HRA.
- Although the issue of resources required to perform a thorough analysis are of interest in HRA method selection, trade-off evaluations relevant to this issue were not addressed in this study. The goal of this study was simply to examine the validity and reliability of the HRA methods, regardless of their resource demands. However, as noted above the level of analysts' understanding of the scenario and the likely performance factors appeared to make a difference in the results
- The pilot phase did not examine all of the capabilities of some HRA methods, such as those related to the identification of HFEs and the treatment of errors of commission. This was partly due to the pilot phase's focus on the identification of qualitative drivers as the common capability of all HRA methods. In addition, the pilot study design had to consider practical limitations related to providing all HRA teams with the opportunity to observe and interact with the crews participating in the experiment. The impact of these limitations on the present results needs to be examined further and adjustments made in future studies as needed.

Upcoming Work: Phase 2-SGTR Study and Phase 3-LOFW Study

In this first phase of the Empirical Study, a comparison of method predictions with the outcomes observed in two variants of SGTR scenarios has been performed. This comparison has focused on the qualitative predictions in the form of (1) the factors driving performance and (2) the operational expressions of these factors. The comparison has been limited to two of the nine HFEs defined for the two SGTR scenarios; these HFEs addressed the identification and isolation of the faulted steam generators in each of the two scenario variants. The following work is planned for Phase 2:

- Simulator data analysis for the remaining SGTR human actions in terms of factors driving performance as well as in terms of operational expressions.
- For each HRA method, comparison of the qualitative predictions to the observed outcomes.
- Development of a "qualitative ranking" of all HFEs in the SGTR scenarios by ranking the HFEs in terms of the level of difficulty observed.
- For each HRA method, establishment of a "predicted ranking" using the HEPs for HFEs in the SGTR scenarios.

- For each HRA method, comparison of the predicted HFE ranking to the observed HFE ranking.
- Documentation of the resulted comparisons development of an understanding of the methods in terms of strengths and weaknesses.
- Identification of needed improvements that go beyond individual methods.

The SGTR study will allow the development of a more in-depth understanding of the methods and needs. It is noted however, that although it will be based on a more comprehensive comparison of HRA results to empirical data, its results will still be considered as preliminary.

Phase 3, the LOFW Study will include similar tasks. It will also include observations regarding the impact of the changes made on the methodology for performing HRA using simulator runs, analyzing the simulator data and comparing the methods-to-data as a result of this Pilot. Thus, Phase 3 will include an overall assessment of the HRA methods as well as the Empirical Study. More importantly, Phase 3 will include overall lessons learned for improving individual HRA methods and the HRA technology as a whole.

The results of the Empirical Study will provide a technical basis for improving individual methods, improving existing guidance documents for performing and reviewing HRA (e.g., NUREG-1792, HRA Good Practices), and developing additional guidance and training materials for implementing individual methods. Moreover, the results of this study will provide a technical basis to support the work that addresses the Commission's Staff Requirements Memorandum (SRM)-M061020, which directed the staff to address the issue of HRA model differences—including examining whether the NRC could adopt a single model for all HRA applications, or whether it should adopt more than one—and to provide explicit guidance on the applicability and implementation of each model.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the HRA teams who participated in the study and whose contributions are invaluable. In addition to their analyses, their interest in the study, combined with the views and feedback they provided, were essential for the design and execution of the study. The authors also acknowledge Alan M. Kolaczkowski and Bruce P. Hallbert for their assistance in the review and study design work.

This study is a collaborative effort between the Joint Programme of the Organization for Economic Cooperation and Development (OECD) Halden Reactor Project (in particular, Halden's signatory organizations, which provided the analyst teams), the U.S. Nuclear Regulatory Commission (NRC), the Swiss Federal Nuclear Safety Inspectorate (DIS-Vertrag Nr. 82610), and the U.S. Electric Power Research Institute.

The HRA teams who participated in the analyses of the SGTR scenarios are:

ASEP/THERP (NRC)

- Y. James Chang, U.S. NRC, Office of Nuclear Regulatory Research, USA
- Mary Drouin, U.S. NRC, Office of Nuclear Regulatory Research, USA
- Stacey Hendrickson, Sandia National Laboratories, USA
- Rick Grantom, C.R. Grantom P.E. & Assoc. Inc., USA

ATHEANA (NRC)

- Susan E. Cooper, U.S. NRC, Office of Nuclear Regulatory Research, USA
- Dennis C. Bley, Buttonwood Consulting, USA
- Mark King, U.S. NRC, Office of Nuclear Reactor Regulation, USA
- Michael A. Junge, U.S. NRC, Office of New Reactors, USA
- John E. Thorp, U.S. NRC, Office of Nuclear Reactor Regulation, USA

CESA_Q (PSI)

- Luca Podofillini, Paul Scherrer Institute (PSI), Switzerland
- Bernhard Reer, Swiss Federal Nuclear Safety Inspectorate-HSK, Switzerland (before July 2007: PSI)

Decision Trees + ASEP (NRI)

- Jaroslav Holy, Nuclear Research Institute (NRI), Czech Republic
- Jan Kubicek, Nuclear Research Institute (NRI), Czech Republic

CREAM (NRI)

- Jaroslav Holy, Nuclear Research Institute (NRI), Czech Republic
- Jan Kubicek, Nuclear Research Institute (NRI), Czech Republic

CBDTM + THERP (EPRI)

- Jan Grobbelaar, Scientech, USA

Enhanced Bayesian THERP (VTT)

- Jan-Erik Holmberg, VTT (Technical Research Centre of Finland), Finland
- Kent Bladh, Vattenfall Power Consultant, Sweden
- Johanna Oxstrand, Ringhals AB, Sweden
- Pekka Pyy, Teollisuuden Voima Oy, Finland

K-HRA (KAERI)

- Wondea Jung, KAERI, Korea
- Jinkyun Park, KAERI, Korea

MERMOS (EDF)

- Pierre Le-Bot, Electricité de France (EDF), France
- Hélène Pesme, Electricité de France (EDF), France
- Patrick Meyer, Electricité de France (EDF), France

PANAME (IRSN)

- Véronique Fauchille, IRSN, France
- Vincent Ridard, IRSN, France
- Manuel Lambert, IRSN, France

SPAR-H (INL)

- April Whaley, Idaho National Laboratory, USA
- Harold Blackman, Idaho National Laboratory, USA

SPAR-H (NRC)

- Gary M. DeMoss, U.S. NRC, Office of Nuclear Regulatory Research, USA
- Bruce B. Mrowca, Nuclear Systems Analysis Division, ISL, Inc., USA

HEART (Ringhals)

- Johanna Oxstrand, Ringhals AB, Sweden
- Kent Bladh, Vattenfall Power Consultant, Sweden
- Steve Collier, OECD Halden Reactor Project, Norway

Three teams have used the scenario descriptions or data as input for their simulation models in order to test the applicability of these methods. These teams have participated by giving general input to the study, but the results have not been analyzed or compared to the HAMMLAB data.

ECAT, Discrete event simulation, MicroSaint (NRC/Sandia/Alion)

- Beth M Plott, Alion Science, USA

QUEST-HP (Risø)

- Igor Kozine, Risø National Laboratory, Denmark

IDAC (University of Maryland)

- Kevin Coyne, University of Maryland, USA
- Ali Mosleh, University of Maryland, USA

Another team has used the data to test a selection algorithm for fuzzy classification of results.

Fuzzy data classification (Politecnico di Milano)

- Piero Baraldi, Politecnico di Milano, Italy
- Massimo Librizzi, Politecnico di Milano, Italy
- Enrico Zio, Politecnico di Milano, Italy

ABBREVIATIONS

AFW	auxiliary feedwater
AFWS	auxiliary feedwater system
ARO	assisting reactor operator
ASEP	Accident Sequence Evaluation Program
ATHEANA	A Technique for Human Event Analysis
CBDTM	Cause-Based Decision Tree Method
CCF	common cause failure
CD	core damage
CESA	Commission Errors Search and Assessment Method
CESA-Q	Commission Errors Search and Assessment Method (Quantification Module)
CICA	important configurations of accident operation
COCOM	Contextual Control Model
CPC	common performance conditions
CREAM	Cognitive Reliability and Error Analysis Method
CRIEPI	Central Research Institute of Electric Power Industry
DDD	detect, diagnose, or make a decision
DT	decision tree
EDF	Electricité de France
EFC	error-forcing context
EOC	error of commission
EOO	error of omission
EOP	emergency operating procedure
EPC	error-producing condition
EPRI	Electric Power Research Institute
ERG	emergency response guidelines
FFD	fitness for duty
FO	field operator
FR	functional restoration
GRS	Gesellschaft für Anlagen- und Reaktorsicherheit
GTT	generic task-type
HAMMLAB	Halden Human-Machine Laboratory
HAT	HEART assessment team
HCR/ORE	Human Cognitive Reliability/Operator Reliability Experiments
HEART	Human Error Assessment and Reduction Technique
HEP	human error probability
HERA	Human Event Repository and Analysis
HFE	human failure event
HFPP	human factors and power plants
HMI	human-machine interface
HPLV	human performance limiting value
HRA	human reliability analysis
HWR	Halden work report
INL	Idaho National Laboratory
IRSN	French Institut de Radioprotection et de Sûreté Nucléaire
KAERI	Korea Atomic Energy Research Institute
K-HRA	Korean Human Reliability Analysis Method
LTA	less than adequate
LOFW	loss of feedwater

MERMOS	Methode d'Evaluation de la Realisacion des Missions Operateur la Sureté
MMI	man-machine interface
MSIV	main steamline isolation valve
MSLB	main steamline break
MWe	megawatts
NEA	Nuclear Energy Agency
NPP	nuclear power plant
NRC	Nuclear Regulatory Commission
NRI	Nuclear Research Institute
OPA	operator performance assessment system
PANAME	French acronym: new action plan for the improvement of the human reliability analysis model
PIR	problem identification and resolution
PLG	Pickard, Lowe, and Garrick
PORV	power operated relief valve
PRA	probabilistic risk assessment
PSAM	probabilistic safety assessment and management
PSF	performance shaping factor
PSI	Paul Scherrer Institute
PWR	pressurized water reactor
PRZ	pressurizer
RCS	reactor coolant system
RO	reactor operator
RWST	refueling water storage tank
SG	steam generator
SGTR	steam generator tube rupture
SHARP1	systematic human action reliability procedure
SI	safety injection
SLB	steamline break
SLIM	Success Likelihood Index Methodolgy
SPAR-H	Standardized Plant Analysis Risk-Human
SS	shift supervisor
THERP	Technique for Human Error Rate Prediction
TRC	time/reliability correlation
USNRC	U.S. Nuclear Regulatory Commission
VTT	Technical Research Centre of Finland
WEC-LLC	Westinghouse Electric Power Company

1. OVERALL STUDY DESIGN AND METHODOLOGY

1.1 Background

A number of diverse human reliability analysis (HRA) methods are currently available to treat human failure in probabilistic risk assessments (PRAs). This range of methods reflects traditional concerns, such as human-machine interfaces and basic feasibility of actions in PRA scenarios. Many of the methods have also been developed to address errors of commission and decision-making performance. Given the differences in the scope of the methods and their underlying models, there is substantial interest in assessing HRA methods, and, ultimately, in validating the approaches and models underlying these methods. In addition, such a validation is warranted to assess the credibility of HRA results when decision makers have to use those results to make risk-informed decisions.

A literature review of benchmarking and HRA has been performed as both a start and a basis for this study [1]. Initial efforts in designing and implementing validation studies have identified a number of issues associated with structuring the studies in order to allow an adequate and appropriate test of the different methods. These issues can impact (a) the ability to test the consistency of HRA results across the different methods (inter-method) and across the same methods using different analysis teams (inter-analyst team), and (b) the ability to test the validity or accuracy of HRA results by comparing the predictions of the different methods to the observed crew performance. Some of the issues to be considered include:

- Which methods to include in the evaluation and how many teams to include.
- How to equalize the teams in terms of their experience using the methods.
- How to construct the scenarios for evaluation and what factors to focus on within the scenarios.
- Whether to focus the comparison of HRA results on qualitative or quantitative analyses.
- How to assess inter-analyst team and inter-method consistency.

With these issues in mind, an international evaluation study of HRA methods was begun with the support of the Organization for Economic Co-operation and Development (OECD) Halden Reactor Project's HAMMLAB (HAlden huMan-Machine LABoratory) research simulator, located in Halden, Norway. Its aim was to develop an empirically based understanding of the performance, strengths, and weaknesses of the methods. The empirical basis was provided by experiments performed on the Halden simulator, using real crews in accident situations. It is expected that the results of this work will provide the technical basis for the development of improved HRA guidance, and, if necessary, improved HRA methods. To this end, the study was designed to incorporate the following elements:

- The involvement of multiple methods and analysis teams.
- Predictive analysis using the methods without knowledge of the experimental results.
- The collection and analysis of the experimental data.
- The comparison of the experimental results with the analysis predictions.

1.2 Overview of the Study Design

1.2.1 Four High-Level Study Tasks

The Empirical Study focused on the HRA of the control room personnel actions required in response to PRA initiating events. This focus was motivated by the widespread use of HRA methods for PRA within the industry, as well as by the significant research and development efforts on HRA methods addressing the issue of errors of commission and decision-making performance, as surveyed, for instance, in [2]. An overview of the study, consisting of the four high-level tasks listed below, is presented in Figure 1-1.

Task 1. Definition of the scenarios and human failure events (HFEs) to be analyzed and compilation of the inputs for the HRA teams.

Task 2. Analysis of the HFEs with HRA methods, which predictions of crew performance.

Task 3. Production of the empirical or reference data for the comparison, starting from the collection of raw data in simulator experiments conducted in HAMMLAB and followed by analysis of the data.

Task 4. Comparison of the predicted outcomes against the empirical data (observed outcomes).

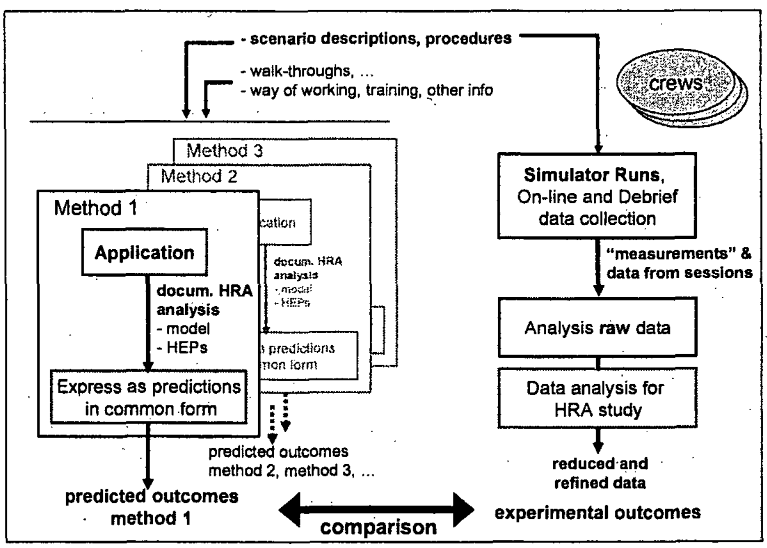

Figure 1-1. Overview of the HRA Empirical Study

Task 1 is the compilation of the inputs for the HRA analysts. As shown at the top of Fig. 1-1, these inputs include not only the descriptions of the scenarios and of the HFEs to be analyzed but also information on the relevant procedures, the training of the operators, their work method, the human-system interface, and other aspects of the performance context. The performance of the predictive HRAs (Task 2) is shown on the left. The production of the empirical data, Task 3 (right-hand side of Fig. 1-1), consisted of three subtasks: (1) performing the simulator experiment, in which the operator crews responded to the scenarios while observations and other data were collected, (2) a first data analysis stage aimed at interpreting the data with the goal of producing an integrated understanding of the performance of the individual crews, and (3) an HRA-oriented data analysis, which aggregated the set of crew performances in order to characterize the overall performance level relative to each HFE and the positive and negative drivers of performance. Task 4 was the comparison between the predicted and the empirical outcomes. It required that the predicted outcomes were expressed in a way that was compatible with the analytical level of the empirical outcomes.

1.2.2 Study Organization, Participants, and Roles

The description of the Empirical Study high-level tasks shows that there was a simulator study embedded within the overall study (Task 3). There were therefore four sets of study participants overall:

- **Halden experimental staff** (Tasks 1, 3): The simulator sessions were conducted in the OECD Halden Reactor Project's HAMMLAB research simulator facility. The HAMMLAB staff was largely responsible for analyzing the experimental data.
- **HRA teams** (Task 2): Each team applied an HRA method to obtain predictions for the HFEs in the scenarios defined for the study. Organizations representing industry, regulators, and the research community participated.
- **Operator crews** (Task 3): A set of licensed reactor operator crews responded to a series of scenarios in the HAMMLAB simulator. Each crew responded to four scenarios, which consisted of a base and a "complex" variant of two scenario types.
- **Assessment and evaluation group** (Overall organization and Task 4): This group had the overall responsibility for organizing and implementing the study. In the early stages of the study, it prepared the information package (analysis input) for the HRA teams and answered their subsequent requests for additional information and questions concerning ambiguities in the instructions and assumptions. After the HRA teams delivered their analyses, the group reviewed and summarized the predicted outcomes before performing the actual comparison.

For a number of its tasks, the assessment and evaluation group worked closely with the Halden staff, especially when preparing the information package, answering operational questions regarding the simulations, and preparing for the comparisons. To avoid biasing the comparison, a "blind" study protocol was used. The assessment and evaluation group did not receive any information about the actual crew performances in HAMMLAB until after the predicted outcomes had been summarized and reviewed with the HRA teams. Similarly, the Halden staff's data analysis (to produce the reference data) was performed without knowledge of the HRA predictions.

1.2.3 Phases of the Empirical Study

The Empirical Study is been executed in three phases, as shown in Table 1-1. The focus of Phase 1 was to test the study methodology. In Phase 1, the HRA teams performed HRAs of nine HFEs in a first set of scenarios, two variants of steam generator tube rupture (SGTR) scenarios. However, data analysis and a qualitative comparison were performed for two HFEs only. The remaining HFEs and the quantitative comparison will be performed in Phase 2. The two pilot phases were designed to allow the study participants (Halden, the assessment/evaluation group, and the HRA teams) to review the study methodology and the initial results, and, in particular, to allow the HRA teams to provide feedback on the methodology. A workshop on the first pilot phase was held in October 2007.

Table 1-1. Phases of the Empirical Study

Phase 1 (2007- 2008) Pilot study	- used data from first set of scenarios (two SGTR variants) - established the methodology and reached some preliminary results on HRA methods - this report
Phase 2 (2008 - 2009)	- data analysis and comparison of remaining HFEs in SGTR scenarios - overall study results for the SGTR scenarios - pilot study of the methodology for comparing and evaluating the quantitative results of HRA (the failure probabilities)
Phase 3 (2008-2010)	- second set of scenarios (two loss of feedwater variants) - to be reported in 2009

Phases 2 and 3 can overlap in time because the HRA teams are performing predictive analyses for the loss of feedwater (LOFW) scenarios while the SGTR data and predictions are being analyzed. In general, the experimental data analysis and the assessment and comparison of the predictions are the critical, most time-consuming, tasks for the study schedule.

This report presents the methodology and initial results from Phase 1 of the study. In Phase 1, the comparison is limited to the qualitative aspects. The comparison and evaluation of the HRA quantitative predictions as a measure of the difficulty of the HFEs will be reported at the conclusion of the second pilot phase in a separate report.

1.3 Study Design: The Overall Methodology

1.3.1 Scenario Design

Human performance levels in nuclear power plant operations are generally high because of the strong safety emphasis in the design of the control room, including the availability of safety systems as barriers to errors, the highly skilled personnel, the comprehensive analysis of potential accident scenarios, the support provided to the operators by abnormal and emergency procedures, and the extensive, ongoing training programs. As a result, operator crews are expected to make few errors in most scenarios; moreover, because accident scenarios generally develop relatively slowly, allowing the operators to receive feedback from the plant, unrecovered errors are expected to be even less likely (the probability of HFEs in PRAs, in general, accounts for recovery factors, such as detecting an omission). In PRAs, the post-initiator operator actions, assessed with HRA methods, are frequently required in postulated scenarios that exceed the limits of the basic nuclear power plant design; for instance, PRA scenarios often include multiple equipment failures. The infrequency of such scenarios is one of the reasons why simulator data is needed and used. Designing and setting up PRA-based scenarios in simulator studies is challenging because they must postulate component and system failures in combinations that lead to the required actions of interest while remaining plausible to the operators; the assumptions made in regarding equipment unavailability can appear particularly unrealistic.

Moreover, several scenarios are needed to provide a varied set of situations for the crews. A base and complex variant of each scenario (SGTR and LOFW) were designed. This was particularly useful in the context of the Empirical Study because the two variants include similar or related tasks that differ in terms of their performance context. The base variant can provide a baseline for performance, while the effects of the differences in the complex variant can be analyzed and compared with the base case. This comparison provides a more complete understanding of the crew actions than would be possible by examining unrelated scenarios individually. It also allows an evaluation of whether the HRA methods are sensitive to such scenario differences and whether their predictions are adjusted accordingly. In addition, the scenarios (2 SGTR and 2 LOFW) were presented to the crews in a semi-randomized order, which allowed control for order effects, e.g., learning and other potential biases

Full-power PRA scenarios have been used in the Empirical Study. Although the performance of HRA methods in other reactor modes such as low-power and shutdown are of interest, full-power scenarios were used for this study mainly because most HRA methods have been developed for control room actions modeled in full power internal event PRAs. Therefore, it is natural to test the methods with human actions they were intended to model. Future studies could include scenarios from other plant modes, such as shutdown.

1.3.2 Predictive HRAs

In principle, the HRAs needed for the Empirical Study are no different from such analyses performed for a PRA. However, the study did not address the whole spectrum of HRA tasks typically performed as part of a PRA. The following sections discuss methodology of the study, including the scope and limitations of the study.

1.3.2.1 Inputs to HRA Teams

To perform HRA within a PRA, the analysts need to be familiar with the background, training, and experience of the performers (the crews) and the performance conditions (e.g., human-system interface and job aids, such as procedures). However, in the Empirical Study, opportunities for all HRA teams to perform familiarization tasks, such as a plant visit, observation of the crews, walk-throughs of the tasks, and interviews with crews or training personnel, were limited. As a substitute, the familiarization package compiled by the assessment group documented as much of this as possible; additionally, the HRA teams requested and received further information in a question-and-answer process.

1.3.2.2 Interaction of HRA and Accident Sequence Modeling

At a higher level, HRA methods have the same purpose (or aims) due to the role of the HRA within the PRA: 1) identification of the HFEs to be included in the PRA accident sequence model, 2) qualitative analysis of the HFEs, and 3) quantification of the probability of these HFEs.

In a PRA, the definition of the accident sequence models and the identification of the associated HFEs within these models is performed with inputs from the HRA, in an interactive or iterative process. This identification analysis task is not addressed in the current Empirical Study, as the HFEs were predefined for the HRA analysts, to ensure that the HRA teams would produce predictions for identically defined HFEs. A different study design and methodology would be required to address HFE identification.

It should be noted that defining the HFEs for the HRA teams does not eliminate the qualitative analyses to be performed, since the HFEs were defined on a functional level, that is, "fails to perform X within Y minutes." As noted by Kirwan in *A Guide to Practical Human Reliability Assessment* (p. 318) [3], "targeted task analyses" should be performed in support of the HRA. This process identifies the main failure modes and the plant- and scenario-specific influences on human performance. Requirement HLR-HR-G of ASME RA-S-2002 [4] lists a number of these influences. The most important influences or factors are sometimes referred to as the factors "driving" performance, or the "driving factors" of performance. Comparing the specific factors identified as driving factors by the HRA teams for the defined HFEs with those observed in HAMMLAB was the main focus of the comparison.

1.3.2.3 Documentation of HRAs

There are differences in the underlying models, the number of performance shaping factors (PSFs), the definition of their scope, and the terminology used in different HRA methods. In addition, the documentation of HRA in PRA is typically oriented to tracing how the performance condition information obtained in the qualitative analysis has been incorporated into the estimation of the failure probability rather than into predicting specific outcomes in terms of behaviors and actions. To address the terminological differences, as well as to provide predicted outcomes that could be compared with the outcomes obtained in the simulator study, the HRA teams were asked to deliver their predictions in three parts:

- Form A, see Appendix B (for each HFE): An "open form" questionnaire where the teams reported 1) the human error probability (HEP), 2) the driving factors, and 3) the "operational expressions."
- A "normal" documentation of their HRAs and quantification, as in a PRA.
- Form B, see Appendices C and D (for each HFE): A "closed form" questionnaire, based on the taxonomy from the Human Event Repository and Analysis (HERA) [2] system [5], with adaptations for the study.

---

[2] The Human Event Repository and Analysis (HERA) system (NUREG/CR-6903, Vol. 1, 2006) [5] is a taxonomy and database designed to capture human performance at a fine level of detail. It was originally designed to chronicle human activities at nuclear power plants in the pre, during, and post initiator phases. A HERA user builds a timeline around an overall event, whereby individual human activities and plant state changes are treated as subevents. Human activities are considered human success (HS) or human error (XHE). For each human activity recorded in HERA, there is a supplemental worksheet, which further decomposes

In Form A, the "open form" refers to the fact that the HRA teams were asked to answer open-ended "essay-type" questions, using the terminology of the applied method. Specifically, the teams were asked to identify the important failure and success performance drivers related to the HFE, as well as to discuss these in terms of the expected crew behaviors and responses. These behaviors and responses are the means by which the driving factors express themselves in nuclear power plant operation-specific and in scenario-specific terms, and are therefore referred to as "operational expressions"; they were particularly helpful in unambiguously interpreting the statements made related to the driving factors.

In many cases, the methods themselves do not require the analysts to produce the results exactly the way they were requested in this study. Although asking the analysts to document their results in the way described above may have required the teams to go beyond the specifics of the method to some extent, it can be argued that it was not much more than what is normally expected when performing an HRA. Certainly a major goal of an HRA is to identify when the crews' response will be good (that is, anticipation of high success rates) and when and why potential problems might arise in crew performance: that is, the application of any of the HRA methods should be based on a reasonable understanding of what can occur in the various accident scenarios. Thus, it was expected that, in general, HRA teams should be able to provide the needed information and that this information could be compared to the simulator results and the results of the other methods.

Form B was an attempt to "standardize" the predictions in a common terminology, with a predefined taxonomy that would be more comprehensive than the terminology of the individual methods. There may have been issues with mapping the method terminology into the common terminology due to the fact that the taxonomy was not specifically designed for reporting HRA results.

1.3.3 Empirical Data Collection and Analysis

The methodology for the simulator study and associated data collection is largely derived from the set of methods used in earlier Halden studies of human performance and human-system interface evaluation, as well as HRA issues and factors. These methods, and the collected data, are described in chapter 2.

It is important to stress that the empirical data were produced by the crews of a reference plant performing in the HAMMLAB simulator. In addition to the general limitations of simulator studies, there were also some differences between the HAMMLAB simulator and the home plant. Specifically there were some differences in the control room interface and in the crew staffing. However, the home plant and the reference plant in the HAMMLAB simulator were similar enough to ensure that the crews' performance would be similar so that the data generated would be appropriate for evaluating HRA methods. One of the measures taken to control the differences was training of the participating crews, highlighting the differences in the plant systems as well as in the control room interface, to ensure that crews were feeling

the activity along a number of performance dimensions, such as performance shaping factors (PSFs). Each PSF features a detailed checklist of exemplars--specific types of behavior that one might expect to accompany that PSF. These tiers of information may be seen as increasing in objectivity the further one drills into the taxonomy. The detailed checklist helps determine whether a particular PSF is a driving factor in the overall event outcome.

comfortable with these differences and that observed performance difficulties would not be attributed to these factors. In addition, during the collection and analysis of the crew performance data, the HAMMLAB staff examined whether reference and home plant differences could have impacted crew performance. It is noted, however, that neither the reference data nor the HRA predictions should be considered as representative of crew actual performance in their home plant.

As noted above, the data analysis to obtain the empirical (reference) data for the comparison was performed in two stages. In the first stage, the aim was to interpret the data so as to produce an integrated understanding of the individual crew performances. In the second, HRA-oriented data analysis stage, the set of crew performances was aggregated in order to characterize the overall performance level related to each HFE and to identify the positive and negative drivers of performance. The approach to the HRA-oriented data analysis was one of the methodological developments of the Empirical Study, and is discussed in chapter 2.

1.3.4 Comparison Methodology

The outcomes predicted in the HRAs performed by the teams were compared with the outcomes obtained from the HAMMLAB experiments. Analytical predictions were compared with experimental outcomes for each of the following (the elements of response Form A):

- The level of difficulty associated with the operator actions of interest (with the HFEs). For the HRA predictions, the level of difficulty is represented by the HEP.
- The factors that most influence the crews' performance in these scenarios ("driving factors").
- The reason for the difficulties (or ease) with which the crews perform the tasks associated with each HFE, and how these difficulties are expressed in operational and scenario-specific terms ("operational expressions").

In addition, several other criteria were evaluated:

- The insights given by the HRA method for error reduction.
- Sensitivity issues, such as the impact of qualitative choices on the HEP.
- Guidance and traceability.

The design of the study methodology and experimental plans anticipated that the HAMMLAB experiments would not support the derivation of HFE failure probabilities from the experimental data. Though large for a simulator study, the number of sessions and crews (sample size) remains small in relation to the crews' expected levels of performance. As a result, the comparative analyses in the first pilot phase intentionally focus on the qualitative insights rather than on the quantitative results obtained with the methods, that is, on the methods' ability to predict the tendencies of behavior and performance in the scenarios.

The comparison of the quantitative HRA predictions will be addressed in the second pilot phase. In the present report, the ranking of the HFEs as predicted by the HRAs will be compared to the relative difficulty of the operator actions observed in the empirical data. In addition, the "accuracy" of the predicted HEPs from a given method may be evaluated both individually and as a group (conservative vs. optimistic tendencies) against reference HEPs, if these can be derived on the basis of the empirical observations. On another level, the HEPs from the different methods may be compared with each other, although the empirical data would inform such a comparison only to a limited degree.

To provide additional support for assessing the validity of the HRA methods, the HRA teams were provided with a common template for representing the results of their analysis (Form B). Several aspects of the HERA database taxonomy were included in Form B for the HRA teams to fill out. In particular, the wide range of PSFs and categories for describing their impact on performance was included from HERA. Each HFE is treated as a subevent in HERA for the purposes of the present study. The goal was to provide a common terminology in which the different teams could represent their results and support their ability to describe the expected consequences of a given scenario. Although the information from Form A turned out to be the main focus of the HRA team results analysis in the pilot study (focusing on method specific PSF terminology and operational expressions), the HERA form supported comparisons of the results of the methods (which often use different terminology) with the Halden data. In other words, in some cases the Form B results were used to help interpret the Form A, method-specific, results. Additionally, on the experimental side, the HERA forms supported a common analysis of all crews by different analysts, and, on the HRA method side, these forms provided the assessment group with a better basis for the summaries of the various methods and an improved understanding of the reasoning in the analyses.

1.4 Specifics of the Pilot Study

The first pilot phase of the Empirical Study was aimed primarily at establishing the methodology. As noted, some of the key aspects that were tested included the information package and interactions with the HRA teams, the HRA-oriented data analysis, and the approach for the qualitative comparison. This section discusses specific aspects of the pilot and of the implementation of the methodology in the pilot.

1.4.1 HRA Methods Represented in the Pilot Study

Twelve HRA teams agreed to participate in the pilot study. These teams and the HRA methods they used are listed in Table 1-2 (two teams used SPAR-H, and one team (NRI) used two methods. Note also that the "NRC staff and consultants" teams are three different teams). Summaries of each method and its references are provided in Chapter 3, together with the comparisons.

Table 1-2. HRA Methods in the Pilot Study[3]

Method: HRA Team					
ASEP/THERP	NRC staff and consultants, USA	CBDT	EPRI (Scientech), USA	HEART	Vattenfall & Ringhals NPP, Sweden
THERP with Bayesian Enhancement	VTT, Finland	Decision Trees + ASEP	NRI, Czech Rep.	KHRA	KAERI, Korea
ATHEANA	NRC staff and consultants, USA	MERMOS	EDF, France	CREAM	NRI, Czech Rep.
SPAR-H	NRC staff and consultants, USA Idaho National Laboratory, USA	PANAME	IRSN, France	CESA	PSI, Switzerland

1.4.2 Implementation of the Methodology in the Pilot Study

The methodology of the Empirical Study was implemented in the pilot study, as described in this overview. However, it should be noted that the simulator sessions and associated data collection were performed and completed prior to the predictive analyses by the HRA teams, the reason being that the Empirical Study was piggybacked onto a simulation study performed in HAMMLAB, "Performance Shaping Factors and Masking" [6], for which the simulator sessions and data collection took place from October to December 2006. The purpose of this study was to use PRA-relevant scenarios to study the effects of masking and other PSFs for HRA. Fourteen crews of three licensed pressurized water reactor (PWR) operators participated in the study, each crew responding to two versions (a *base case*—a familiar, routinely practiced case; and a *complex case*—a less familiar, more challenging case) of two scenarios, an SGTR and a total LOFW. Thus, each crew completed a total of four scenarios, in addition to training on the simulator.

The whole study was carried in a "blind" fashion. As with most simulator studies, the operator crews had no prior knowledge of the scenarios. The assessment and evaluation group compiled the "information package" for the HRA teams without any knowledge of crew performance results. Also, this group summarized and reviewed the HRA submittals and provided them to the HRA teams for their review prior to obtaining information on the simulator experiment results. The HRA teams completed their HRAs without any knowledge of crew performance results or the results of other HRA teams. The HAMMLAB staff that collected, analyzed and complied the crew data did not have any prior knowledge of HRA predictions. The information was compartmentalized to ensure that knowledge of the predictions would not bias the interpretation of the experimental data, or vice-versa.

[3] In addition to the listed teams performing analyses with HRA methods, three teams have utilized the data as input for their simulation models in order to test the applicability of these methods in such a setting: QUEST-HP (Risø, Denmark), Microsaint (Alion, USA) and IDAC (University of Maryland, USA). Also, one team has used the data to test a selection algorithm (Politecnico di Milano, Italy).

1.4.3 Review of Methodology and Results in a Workshop

The first phase of the Empirical Study was a pilot study, intended to test the methodology's adequacy in dealing with various benchmarking issues, and to evaluate its implementation—for instance, for the data analysis and comparison tasks. Consequently, the details of the comprehensive study methodology, the results for each study task (intermediate results), and the overall comparison results for the two HFEs were documented in a draft report. In a workshop hosted by the U.S. NRC in October 2007, with the participation of the HRA teams, the assessment and evaluation group, and the Halden staff, this material was presented and reviewed to obtain feedback and input for further methodological improvement.

1.4.4 Inter-Analyst Team and Inter-Method Consistency

Although it remains a goal for the future, the present pilot study did not intend to address the issues of inter-analyst team consistency when using the same method and inter-method consistency, instead focusing on the quality of the information provided by each of the methods (when applied by knowledgeable users).

Despite HRA "good practices," such as the recent NUREG-1792, there can be significant variability in the way HRA methods are applied, and it can be difficult to add provisions to ensure consistency in the method application. The consistency of results obtained from different analysts was not examined in the present study, as there were not multiple teams applying each method. Finally, given the exploratory goal of this pilot study, we did not apply any measure to ensure that the HRA teams were equally experienced in the application of their chosen method, especially since, with one exception, each HRA team only applied one method. Nevertheless, each team included experts in the methods that it applied, and, in some cases, the developers of the methods.

1.4.5 Details of the Information Package Given to the HRA Teams

Each HRA team received an information package, which included the following items, to be used as the basis for the application of an HRA method (or several):

1. Overview (of the information package) and instructions for the HRA teams.
2. Administrative information and agreement forms.
3. Study outline.
4. HAMMLAB information.
5. Scenario description and HFE definitions.
6. Characterization of the crews and their work practices and training.
7. Procedures used in HAMMLAB.
8. Forms for the HRA team responses.

More generally, the package provided information about the organization of the study, the general performance conditions (e.g., information about the interface, the work practices of the crews, the procedures), information about the specific scenarios simulated in HAMMLAB, and forms for the HRA team responses.

It was not possible to provide a complete set of HAMMLAB procedures in English, due to availability and proprietary issues. Consequently, the procedures included in the package were limited to those expected to be used in the scenario variants for this phase, although the study organizers recognized that information in other procedures may have an influence on the crews' performance in the scenario.

The HRA teams had the opportunity to request clarifications or additional information while analyzing the scenarios. To ensure a common understanding of the scenarios and predictions and consistent assumptions among the HRA teams, all questions and corresponding answers were provided to all teams.

1.4.6 Experimental Method and Measures

The study used the data from the PSF/Masking experiment, which had an extensive data collection in the fall of 2006 [6]. Thus, the design of the scenarios and the details of the data collection were decided in this project. Chapter 2 provides a description of the design and the experimental measures of this study, an extract of which is given below.

Fourteen crews with licensed PWR operators participated in the study, each crew consisting of a Shift Supervisor, a Reactor Operator, and an Assisting Reactor Operator. The HAMMLAB PWR simulator, called FRESH, is a full scope simulator of a French plant (CP0 series), with a computerized human-machine interface. The HAMMLAB PWR procedures are based on the procedures used at the participating operators' home plant, and have been adapted to the simulated PWR and the HAMMLAB interface. The participating operators' home plant uses the Emergency Response Guidelines (ERGs) developed by the Westinghouse Owners Group.

The home plant has conventional control rooms with panels and alarm tiles. The HAMMLAB PWR simulator is based on digital instrumentation and control. Given the few differences in the home plant's systems/equipment and those in the Halden PWR simulator, the simulator does not precisely reproduce the actual plant (i.e., the power operated relief valves (PORVs) are different). Therefore, prior to participating in the experimental scenarios, the crews were trained on the use of the screen-based interface and on the differences between their actual plant and the simulator.

The data collection included:

- Crew interview: After each scenario, the crew participated in an interview focusing sequentially on phases of the scenario.
- Operators' PSF ratings: After each scenario interview was complete, the operators individually rated several PSFs for all scenario phases.
- Operator Background Questionnaire.

- Observer PSF ratings and comments: An observer sitting in the control room rated four PSF items for each scenario phase and provided free text comments for the same phases.
- Operator Performance Rating System (OPAS) and performance rating: Under each scenario run, a process expert filled in the OPAS from the gallery by checking the completion of a set of predefined crews' actions and detections. He/she also rated the crews' overall performance in the scenarios phases.
- Observer comments: Under each scenario run, a process expert verbally commented on interesting aspects of the crews' activity and process development.
- Logs: All of the crew's simulator activities were logged, as well as the simulator events.
- Audio/videos: Two fixed cameras behind the operators and two head-mounted cameras on the shift supervisor and reactor operator were employed. All operators were equipped with wireless microphones.
- The detailed performance measures comprised extensive information about the various phases of the scenario. These phases correspond to the defined HFEs. The various experimental measures, including extensive data collection on influencing PSFs and narratives about crew behavior, enabled the Halden team to prepare detailed descriptions of what the crews did, when they did it, and why. This constituted a good basis for qualitative comparisons with the HRA method predictions for each HFE.

1.5 Overview of the Remainder of this Report

Chapter 2 documents the HAMMLAB simulator experiment that comprises the basis for the empirical data and the methodology and results of the data analysis.

Chapter 3 presents the comparisons of the outcomes predicted by each HRA teams with the empirical data. The assessments for each method are based off these comparisons.

Chapter 4 presents the overall insights regarding both the HRA methods, and the methodology of the study. It also includes an outlook of the follow-on phases of the Empirical Study.

Appendix A provides summaries of observed crew performances for a select number of crews related to the HFEs analyzed in this phase.

Appendices B and C include the forms that were given to the HRA teams to report their results.

It is noted that this work has also been published as a Halden report: HWR-844 rev 2, October 2009 [48].

2. EMPIRICAL DATA: HALDEN DATA COLLECTION AND ANALYSIS

2.1 Introduction and Overview

This chapter describes the data collection performed at the HAMMLAB facility and the derivation of the empirical (reference) data.

This chapter:

- Describes the methodology for the simulator study.
- Provides details about the participating crews of licensed reactor operators.
- Describes the experimental scenarios and human failure events (HFEs).
- Discusses the methodology used for the data analysis, integration, and aggregation.
- Presents the intermediate and final results of the Halden data analysis.

The empirical data, which are compared to the outcomes predicted by the human reliability analysis (HRA) teams, describe the aggregated performance of all crews on the HFEs in the scenarios. In the Halden data analysis, data on the individual crew performances was first analyzed to arrive at an integral understanding of each crew's performance. In a second stage, the integrated summary data on the individual crew level were analyzed and combined to describe the performances of the aggregated crews.

In the pilot study methodology, the aggregated performance of the crews in the base case scenario and the complex variant of the scenario is described in two ways, which correspond to the two ways in which the HRA teams were asked to report their predictions. These are namely:

- Performance expressed in operational terms ("operational expression").
- Main drivers of performance, or driving PSFs (performance shaping factors).

2.2 Methodology

The scenarios used for the experiments were generated so as to represent a relatively realistic accident progression. This means that the scenario unfolds from the initiating event according to how the crew handles the scenario. The experimental interventions after the initiating event are the implementation of planned malfunctions. There is a degree of freedom for the crews to impact the scenario's development, even in the presence of a rather comprehensive set of operating procedures. The crews' timing of operations, chosen strategy, and possible performance problems (on tasks earlier in the scenario) all influence the development of the scenario.

Due to these inherent complexities, we expect performance issues and PSF issues that are not fully foreseen and that are not satisfactorily captured by predefined measures. Therefore, the planned analysis of the experiment consists of two analysis levels:

- The first level focuses on integrating and combining the raw data (diverse measures, observations, video recordings, etc) into performance and measures and PSFs, mainly on the individual crew level.
- A second, higher level focuses on the relationship among the PSFs and the relationship between PSFs and human performance. These relationships are analyzed first at an individual crew level and then at an aggregated level.

The analysis levels are further described in the sections below.

The expected frequencies of task failures modeled in the probabilistic risk assessment (PRA) scenario events are generally low, and the number of simulator runs is limited by the availability of resources and crews. We therefore decided to study performance under relatively adverse conditions, where significant performance problems can be expected. These are referred to as *complex scenarios*. As a baseline against which performance in the complex scenarios can be compared, the study included scenarios with relatively advantageous conditions, where good performance can be expected. The latter are referred to as *base case* scenarios.

2.3 Simulator Study Design and Experimental Procedure

The data presented in this section were obtained from an experiment run in the context of a Halden project on the topic of "PSFs and Masking" in [6] and as noted in the previous chapter. In this experiment, crew performance in a "masking" condition (e.g., missing or misleading indicators of plant state) was compared to a base case for two scenarios, steam generator tube rupture (SGTR) and loss of feedwater (LOFW). For the current purpose, only data from the SGTR scenarios will be presented.

2.3.1 Scenario Presentation Order

All 14 crews ran all four scenarios of the experiment:

- SGTR base
- SGTR complex
- LOFW base
- LOFW complex

To control the confounding effects caused by learning due to the order of presentation of treatment level (base case or complex case, that is, degree of complexity or "treatment" or manipulation of the independent variable in the experiment) and scenario type, the experimental presentation order is organized by a combination of theoretical and combinatorial considerations. This includes, for example, excluding combinations with consecutive presentations of the same scenario type on the same day, and avoiding a contiguous scenario type between day one and day two. It was also assumed that there was symmetrical learning between scenario types, that is, that learning related to increased simulator experience only.

2.3.2 Participants' Daily Schedule

During the seven-week data collection period of the study, two crews per week participated in the experiment. Each crew stayed in Halden for three days, starting either on Monday or on Wednesday.

2.3.3 Operator Training in HAMMLAB

To account for the differences between the crews' home plant control room and the Halden pressurized water reactor (PWR) simulator control room, the crews were trained to use the screen-based interface and on the differences between their home plant and the simulator. The training included:

- Interface training (1 hour).
- A presentation on the differences between the HAMMLAB PWR simulator and the actual plant (1 hour).
- Participation in simulator exercises in non-experimental scenarios to learn system/equipment differences (1 hour).
- Participation in training scenarios (non-experimental scenarios) where the crew operates as a team, following procedures (5 hours).

The purpose was to ensure that the crews' performances were not influenced by their unfamiliarity with HAMMLAB.

2.4 Characterization of Crews, Work Practices, and Training

The description below illustrates the normal crew organization at the crews' home plant. In the current experiment, there was only one participating shift supervisor, one reactor operator, and one assisting reactor operator. As there were no major problems or activities that required a balance-of-plant operator (called turbine operator at the home plant), the lack of such an operator in the simulations is deemed to have had no significant effect on crew performance. The assisting operator did the initial checks for turbine trip, then acted as an assisting reactor operator. Interactions with the field operator(s), the safety engineer, and plant management were simulated in a role-play, with an operations expert at the gallery in HAMMLAB acting out each role via phone. The crew was supposed to interact with its organizational environment the same way it would in the plant or in a training simulator session. Substituting an operations expert for balance-of-plant and field personnel in the control room gallery is a familiar process in the training simulator at the home plant.

2.4.1 Crew Organization

The different units at the actual plant can exchange personnel, but, as they have dedicated training simulators, there are differences between the control rooms. In every crew, each of which is responsible for one reactor, there is a shift supervisor, a reactor operator, an assisting reactor operator, a balance-of-plant operator, and at least three field operators.

Shift supervisor (SS): Overviews the situation and calls for meetings when needed. Calls the safety engineer. Monitors critical safety functions. Must be consulted if a procedure step is omitted. Can help with alarms if asked.

Reactor operator (RO): Reads the emergency procedures. Reacts to reactor alarms.

Assisting reactor operator (ARO): The "arms and eyes" of the reactor operator. Performs most actions in emergency procedures, under orders from the reactor operator. Monitors steam generators and controls auxiliary feedwater (AFW) flow.

Balance-of-plant operator (normally called turbine operator (TO) at the plant and in HAMMLAB): Responsible for turbine and electrical systems. Reacts to turbine and electrical alarms.

Field operator (FO): Performs local actions, under orders from the operators.

In an emergency situation, the shift supervisor will call an on-duty safety engineer, who will call the emergency organization for technical support.

2.4.2 Crew Experience

As would be expected, it is sometimes the case that a few of the crews will have a relatively inexperienced crew member. Similarly, SSs will have varying degrees of experience. Table 2-1 presents a summary of the participating crews' experience and years on the job.

Table 2-1. Experience of Participating Crews

	Number of operators	Years Mean	Minimum	Maximum
Total years working at NPP	34	21.2	4	30
SSs working as SS at home plant	14	7.8	1	25
ROs working as RO at home plant	14	4.3	1	15
ROs working as RO and ARO at home plant	14	7.3	1	24
AROs working as ARO at home plant	12	7.7	0.3	25
AROs working as TO at home plant	5	8.2	4	18

Note that five operators who worked as AROs in the experiment worked as TOs at the home plant, although in the emergency scenarios they functioned as AROs (see Crew Organization above). Of those five, only two did not have any experience as an ARO.

2.4.3 Leadership Styles, Team Orientation, Crew Dynamics, and Communication Style

The SSs have the same initial training, but vary in their leadership styles: for instance, some are more democratic, while others are more autocratic. There are no clearly stated goals as to how the SSs should behave in that regard. In the initial training, they are trained to maintain an overview of the situation, and to call for meetings[4] when necessary; they are also trained to encourage democracy, that is, to always let the crew members speak first in meetings to avoid over-influencing them. They are also, however, taught to make decisions by themselves if there is no time for consultation.

The operators usually work independently, but are encouraged to communicate with each other as much as possible. Starting up major or important systems, or other actions that may affect the other operators, must be communicated. The reactor operator and the assisting reactor operator are exceptions to this rule; they usually work together, although they can also work independently. The reactor operator can, for example, continue alone in the emergency procedures while the assisting operator performs other tasks, such as controlling AFW flow or communicating with field operators.

In terms of communication protocol, all orders should be repeated by the recipient and should contain object and action. All crews are trained to communicate like this, though some feel uncomfortable with this level of formality and omit parts of the state-and-repeat protocol; they might, for example, give a more colloquial answer, such as *Yes* or *OK*, instead of repeating the order. As noted, however, the operators are trained in communication strategies. When the assisting operator is asked to read a value, he/she should answer with the appropriate value and trend, even if the question could be answered with a *Yes* or a *No*.

2.4.4 Use of and Adherence to Procedures

The RO reads the emergency procedures. Crews could hurry when necessary but should never read so quickly that thoroughness of the work is compromised, or so that the reading becomes incomprehensible to other crew members. They are taught that it is generally better to do something slowly and get it right, rather than to do it quickly and get it wrong. The pace of the reading varies slightly among the crews.

In terms of acting in advance, there are no clear indications as to whether or not this is allowed. The normal practice is to follow the procedure, but we believe that the operators do not feel that they are forbidden from anticipating procedurally-guided actions at need. If the crew feels that they are performing the wrong procedure, they have the option to start over in E-0 (safety systems verification and diagnosis procedure). The RO is allowed to make deviations if this is approved by the SS. Normally, if there is a need to deviate from the procedure, the RO and the SS would discuss it first.

[4] It is standard practice at the home plant to call short discussion meetings to determine an appropriate course of action quickly in the face of a plant upset condition. These meetings may result in switching to a different set of emergency operating procedures, if sufficient information is present to necessitate a switch. The practice of holding meetings is part of the operational culture of the plant and the regulatory framework for allowable actions within the 30-minute rule. Such meetings are not standard practice at U.S. plants, but are common in some other countries.

2.4.5 Criteria for Having Crew Meetings During Accident Scenarios

The SS—and, to some extent, the rest of the crew—are trained to use specific meeting practices for different types of meetings, those held for a quick overview of the situation and those held to plan or to make a decision when there is a problem. Any crew member can call for a meeting, and is encouraged to do so, but it is the responsibility of the SS to initiate a meeting when it is needed. Meeting location and frequency vary considerably, depending on the SS. Brief meetings are the most frequently used; these should be kept very short and aim to update everyone on the situation, form a common strategy, and initiate important actions. This meeting should be used when the situation is unclear and stressful, but is often held when things have calmed down a bit. Some crews take a brief meeting when they transfer from one procedure to another.

2.4.6 Scenario-Relevant Training

The theoretical training follows a cyclic program of six years, with each subject repeated every third or sixth year. The actual training focuses on SGTR procedures, E-2 (secondary break), and different functional restoration (FR) procedures (e.g., FR-H1) every sixth year. However, training for all major emergency procedures, like E-1, E-2, and E-3, is normally held every year in the simulator. E-0 training is held a minimum of 10 times a year. In the interviews after the scenarios investigated in this study, we asked the crews if there were some parts of the scenario in which they had not been trained. Most crews answered that they had been trained in all events, though not necessarily in the same combination as in the SGTR complex scenario, and that they were very familiar with the SGTR base scenario.

At the home plant, training for an SGTR scenario is normally held twice every year in the simulator. The crews have one week of simulator training in autumn, using one unit's simulator, and then they train again in the spring for the same scenarios in the other unit's simulator.

2.5 Scenario Description

2.5.1 Operating Procedures

The HAMMLAB PWR emergency operating procedures (EOPs) were based on the emergency response guidelines (ERGs) developed by the Westinghouse Owners Group. Below is a short summary of the procedures used in the experiment:

- E-0 "Reactor trip or Safety injection": E-0 is the safety systems verification and diagnosis procedure that should be applied when the reactor has tripped, when that safety injection has been initiated, or when there is a need for a reactor trip or a safety injection.
- E-3 "Tube rupture in one or several steam generators": E-3 is the SGTR event procedure for handling tube rupture. E-0 and several other procedures contain steps for transferring to E-3.
- ES-1.1 "Safety Injection Termination."

- E-2 "Isolation of steam generator with secondary break."
- FR-H5 "Response to steam generator low level."

2.5.2 SGTR Base Case Scenario

In this scenario, an SGTR is initiated in steam generator (SG) number 1 (SG1), which is sufficient to cause nearly immediate alarms of secondary radiation and other abnormal indications/alarms, such as SG1 abnormal level and lowering pressurizer level. Conditions, while continually degrading, are not enough to cause an immediate automatic scram. About three minutes after the tube rupture is initiated, the large screen display will indicate lowering pressurizer pressure and level, increased charging flow (as it attempts to make up for the loss of reactor coolant from the tube break), increasing SG1 level, and a slight imbalance in feedwater flow to the SGs. If the crew also calls up the radiation monitoring display screen, they will see higher radiation indications associated with SG1. It is expected that at this point, or as conditions continue to deteriorate over the next few minutes, the crew will manually scram the reactor. Even if they do not, an automatic scram will eventually occur due to low pressurizer pressure or some other trip setting. Whatever the case (manual or auto scram), the crew is then expected to enter the E-0 procedure.

About 10 minutes after entering E-0 (if the crew has not been delayed based on responses to the steps in the E-0 procedure), the crew will typically be reaching step 19, which is the first E-0 step for which radiation indications of an SGTR necessitate a transfer to procedure E-3 (the SGTR procedure). At this point, secondary radiation is high (as it has been virtually from the beginning), and the SG1 level becomes elevated as compared to the other SGs once the level indications are restored following the scram, but it takes a while longer before SG pressures divert. Post-trip, auxiliary feedwater system (AFWS) input feed imbalances may also exist among the SGs. While it is expected that the crew may enter E-3 at this point, it is noted that a couple of steps later, in E-0, there is another step calling for a transition to E-3, based on an SG-level-checking step (if any SG level is rising uncontrollably, go to E-3).

If/when the crew enters E-3, the scenario proceeds in response to the crew's actions, with no failures or other complicating factors induced by the simulation design: that is, the plant response will be based on the crew's actions in procedure E-3. In general, the crew is expected to perform four primary tasks corresponding to the HFEs defined for the base SGTR scenario, including (a) identifying and isolating a ruptured SG, (b) quickly cooling down the reactor coolant system (RCS) by dumping steam, (c) quickly depressurizing the RCS using the pressurizer sprays or a pressurizer power operated relief valve (PORV) to expedite the depressurization, and (d) stopping safety injection (SI) upon indication that the SI termination criteria are met. Note that for the present report, only the first task is analyzed.

2.5.3 SGTR Complex Case Scenario

This scenario is similar to the SGTR base scenario except for five very significant differences, two of which are relevant to the present analysis:

(a) The event starts off with a major steamline break with a nearly coincident SGTR in SG1, which will cause an immediate, automatic scram, and the expectation that the crew will enter the E-0 procedure.

(b) Auto closure (as expected) of the main steamline isolation valves (MSIVs) in response to the steamline break but along with failure of any remaining secondary radiation indications (neither immediately known nor expected by the crew) as part of the simulation design.

The steamline break "drives" the plant response early in the scenario, when the initial plant behavior resembles that expected in response to a significant steamline break, with quick closure of the MSIVs. This fact, along with the failure of all secondary radiation indications/alarms, is expected to "mask," at least initially, the nearly coincident occurrence of the SGTR in SG1. This should make it considerably more difficult for the crew to diagnose the existence of the SGTR, especially in response to E-0 step 19, which concerns elevated radiation indications.

2.5.4 SGTR HFE Definitions and Event Tree

Figure 2-1 represents a typical PRA event tree for an SGTR event. It is presented here to provide an overall PRA context for the HFEs to be evaluated (note that this report only addresses HFE 1). Its sequence end states (outcomes) refer to whether the reactor core is safe in the long run, or if there is core damage (CD). Those paths through the event tree and the relevant human failure events (HFEs) of interest for the current study are set in bold. All other sequences on the event tree, and those system successes or failures or operator actions associated with refueling water storage tank (RWST) refill, were not simulated.

As a model of an accident sequence, the event tree represents in general terms the way the operators are trained to respond to an SGTR event with the E-3 procedure. However, in a PRA, the success criteria for the events are typically determined by avoiding irreversible changes to the plant state that affect the likelihood of core damage. For this exercise, the training staff expectations on the operator responses were considered in determining the success criteria. These expectations are reflected in the crews' training. In applying the procedures, the operators are also trained to focus on more intermediate and detailed goals that are particularly relevant to an SGTR event. They are taught to define "success" as "timely operator intervention in order to limit the radiological releases and prevent steam generator (SG) overfill" (a quote from a basis document for the procedures); they also learn to terminate primary-to-secondary leakage expeditiously. They want to limit the radiological releases that are, in part, a function of the time it takes before the rupture is mitigated, and they do not want to overfill the ruptured SG, since this could cause an SG pressure relief valve to open (thereby allowing more release), or, worse yet, cause a main steamline break or leak (also allowing more release as well as further complicating the shutdown).

The operators' more pertinent goal of limiting the radiological release is achieved by following the E-3 procedure. For the HFE analyzed in this report, the relevant tasks are to identify and isolate the ruptured SG; because of the overall goal of limiting radiological release, the operators are trained to perform these actions quickly and efficiently, using similarly efficient procedures.

Further, the operators are taught that failing at any of these tasks brings undesirable consequences. For instance, if the affected SG is not identified and isolated, releases will remain high, an outcome to be avoided.

The operators are trained to recognize such undesirable consequences, and to perform their tasks expeditiously and correctly, as specified in the procedures. They are also taught that, in order to limit the release, all tasks should be completed before the ruptured SG overfills. While they do not think of the task in terms of clock time, they are aware of the need to work with enough urgency to meet the overall goal; thus, when they simulate an SGTR event in their training, there is some level of expectation regarding typical response times to perform the various tasks. The HFE success-failure definitions are based on these temporal expectations, as well as on the expected accomplishments for each task. While the threshold times allotted to each task are not exact, they do represent times by which the operators could be viewed as being slower than expected, since the overall goal could then be jeopardized. Based on these considerations, HFE 1 was defined as follows:

HFE #1, *"Failure of the crew to identify and isolate the ruptured SG."* Success requires that the crew:

- Enters procedure E-3 (preferably from E-0 step 19).
- Has closed/isolated all steam outlet paths from the ruptured SG (SG1).
- Has stopped all feed to the ruptured SG as long as the ruptured SG level is at least 10%, as indicated in the narrow range SG level indications (to ensure that the SG U-tubes will remain covered).

The crew is expected to take at least a few minutes before the plant trip to observe and evaluate the initial indications of the tube rupture, about eight to ten minutes to enter and get to E-0 step 19, five minutes to actually enter E-3 and perform the initial isolations/stoppages, and an additional few minutes for reasonably acceptable variability among crew responses. Based on these expectations, it was assumed that once the tube rupture occurs and triggers the event, failure to successfully perform the above within 20 minutes in the base case (HFE 1A) or 25 minutes in the complex case (HFE 1B) would constitute "failure" (as this would be a slower response than expected/desired).

Note that the isolation manipulations involve the following and would typically take less than three minutes to do:

Control room actions. These are all expected of the crew, and are a part of the HFE:

- Verify steam dump to atmosphere valve set point is at 70.5 bar.
- Verify blow down isolated.

- Verify main feedwater isolation.
- Close steam valve to turbine-driven AFW pump.
- Close main steamline isolation valve and its bypass valve.
- Stop AFW when level is greater than 10%.

Local actions. The crew should at least make a phone call for these actions which are part of this HFE.

- Verify atmosphere valve's steam dump closed.
- Verify lock steam valve to turbine-driven AFW pump closed.
- Verify steam traps closed.

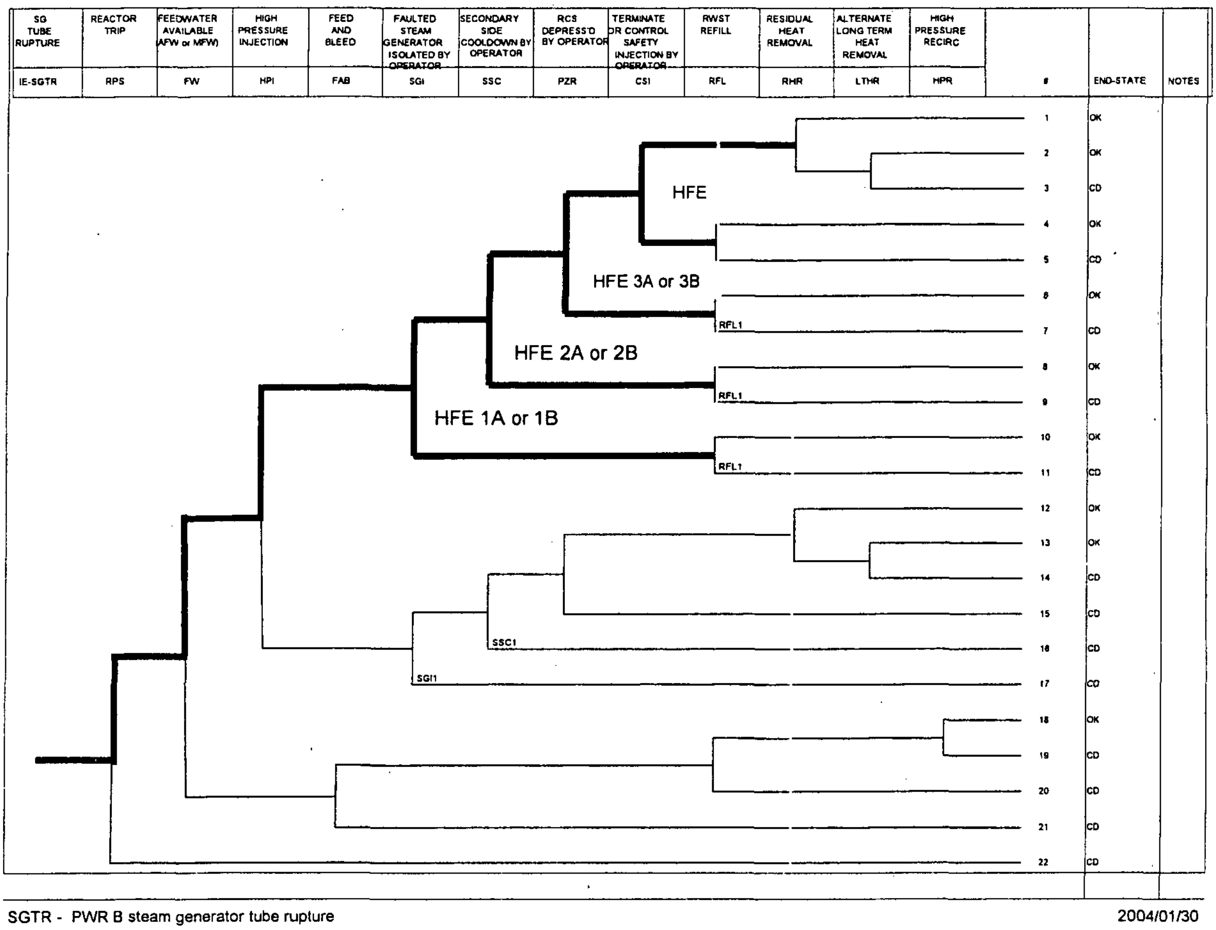

Figure 2-1. Event Tree for SGTR Scenario

2.6 Simulator Data Analysis Methodology

The experience from previous HAMMLAB studies and the test analyses performed for the current experiment have shown that a thorough qualitative analysis is needed to derive sufficient insight into drivers and explanations of crew performance. In this section we describe the integration and aggregation process used for generating the higher-level data representations. These representations constitute the building blocks for the results presented in the following sections, namely:

- Crew performance descriptions.
- Operational expression of crew performance.
- Identification of PSF drivers.

As noted in the previous chapter, the basis for the analysis process is a diverse pool of qualitative and quantitative data collected during the experimental runs:

- *Operator Background Questionnaire*: Capturing years of experience in the current position and other job-related information.
- *Logs*: All crew activities and plant states are recorded in the simulator logging system.
- *Audio/video recordings*: Two fixed cameras are placed behind the operators, and two head-mounted cameras are attached to the shift supervisor and the reactor operator. All operators are equipped with wireless microphones.
- *Observer PSF ratings and comments*: An observer located in the control room rates four PSF items (two on complexity and two on teamwork) for each scenario phase and provides free text comments for those phases.
- *OPAS and performance rating*: During each scenario run, a nuclear power plant operations expert fills in the operator performance assessment system (OPAS) from the observation gallery by checking the completion of a set of predefined crew actions and detections. He/she also rates the crew's overall performance for each phase of the scenario.
- *Observer comments*: For each scenario run, a process expert verbally comments on interesting aspects of crew activities and process development.
- *Crew interview*: After each scenario, the crews participate in an interview focusing sequentially on phases of the scenario.
- *Operator PSF ratings*: After the interview, operators individually rate several PSFs for all scenario phases.

The next four sections describe the different phases of the analysis process, namely initial data screening, selection of crews for in-depth analysis, generation of operational stories of crew behavior, and PSF rating. Section 2.7 presents results relating to HFE performance and operational aspects. Methodology and results for the PSF driver identification are discussed in detail in Section 2.8.

2.6.1 Initial Data Screening

During the first stage of analysis, attention was focused on the quantifiable data, namely expert and observer performance ratings, crew PSF ratings, OPAS scores, and performance figures generated from simulator log files (e.g., task performance time).

2.6.2 Selection of Crews for In-Depth Analysis

To derive sufficient insight into drivers and explanations of crew performance, a subset of runs was selected for in-depth qualitative analysis. The selection was aimed at identifying a mixture of crews at both ends of the performance spectrum, and was performed by a panel of experts, mainly drawing on the HFE criterion, performance time.

The selection led to a set of nine crews, who were subsequently analyzed in detail.

- Three base case crews, consisting of:
  - Two crews succeeding on the HFE 1A performance time criterion ("fastest crews")
  - One crew failing the HFE 1A performance time criterion ("slowest crew")
- Six complex case crews, consisting of:
  - Three crews succeeding on the HFE 1B performance time criterion ("fastest crews")
  - Three crews failing the HFE 1A performance time criterion ("slowest crews")

2.6.3 Operational Stories and Influencing Factor Identification

This stage of the analysis was based on recorded crew communication, recorded expert comments, simulator logs, and crew interviews. The core of the analysis process was the detailed review of the video recordings of the scenario phases corresponding to HFE 1. These reviews were structured so as to be useful and relevant for comparison to the HRA submissions.

- Analysts viewed the video and transcribed key communications and events. They also commented on salient aspects of crew performance.
- Immediately after the viewing, they completed a simplified version of the Human Event Repository and Analysis (HERA) system worksheets (c.f. NUREG/CR-6903, volume 1, [5]) in order to record the PSF details identified during the video review. In completing HERA, the analysts also drew on additional data sources, such as the crew interview, crew PSF questionnaire, and observer comments.
- Finally, the analysts summarized the crew performance in a crew summary, highlighting performance characteristics, drivers, and key problems. The contents of a crew summary are outlined in Table 2-2.

Table 2-2. Structure of a Crew Summary (Individual Crew Performance)

HFE:
Narrative (Identification phase) - timeline of key crew behaviors, communications, operator actions - short free-form description of salient aspects of crew performance
Narrative (Isolation phase) - timeline of key crew behaviors, communications, operator actions - short free-form description of salient aspects of crew performance
Summary of most influencing factors affecting performance (individual crew)
Summary of the (a) observed difficulties (or ease) of various tasks within performance and (b) why the task was easy or difficult

The summaries for the selected crews can be found in Appendix A. A typical crew summary is one to two pages in length, while the timelines are extracts from a narrative built from the analysis of the video recording (DVD story).

The summary of the most influential factors at the individual crew level was based on the HERA taxonomy, whose factors are listed in the last column of Table 2-11 in Section 2.8.1. The PSFs were categorized as shown in Table 2-3. The approach to identifying the influencing factors is described next.

Table 2-3. Categorization of Influencing Factors at Individual Crew Level

Categorization of factor	Description
Direct negative influence	The factor has a negative impact that is observable in the operator's performance and can be linked to delay or failure of the HFE.
Negative influence present	The factor is rated negatively, but there is no clear evidence that it significantly affected this crew's performance.
Neutral	The factor does not appear to affect performance (positive rating but no observable impact, neutral rating).
Positive influence	The factor is rated positively, and there is an observable, positive impact on the performance of the HFE.

2.6.4 Identification of Influencing Factors, Modified HERA Form

A particularly important aspect of the analysis process was the detailed identification and description of observed PSFs. For the present purpose, we briefly outline the identification method used during the video review. The results of the PSF integration and aggregation are presented in Section 2.8.

As mentioned, the PSF analysis process of the simulator data was designed around the HERA system. However, modifications were necessary to make HERA fit the constraints and requirements of a simulator study, as noted below:

- The first step in this adaptation process was a screening phase. In this phase, a team of process experts, HERA experts, and experimenters discussed all the HERA PSF items (PSF details) and organized them into three categories[5]:
  - *Constant PSF details.* These items—for example, procedures—are "constant" across crews. It should be emphasized that, while these factors are the same for all crews, (1) they may evolve during the scenario, and (2) they may not impact all crews in the same way.
  - *Variable PSF details.* These were items where variability was expected between crews and within runs. Examples are "Team Dynamics" and "Work practices."
  - *Not applicable.* A number of PSF details, often related to maintenance work or balance-of-plant activities, were considered irrelevant for the current analysis and were therefore removed from the working list of PSF details.
- Next, the constant PSF details were rated by an expert panel consisting of process experts, HERA experts, and Human Factors experts.
- For the variable PSF details, working definitions and operationalizations were produced. This step was important, as without scenario-specific guidance on what constitutes the "presence" of a PSF detail, low inter-rater reliability for PSF assignment during DVD review could be expected.
- A number of scenario-specific items were added to the list of HERA PSF details. These additional items related to specific hypotheses about crew performance problems that had been noticed during the observation of the experimental runs or in the initial data screening. Such items addressed, for instance, specific aspects of supervisor behavior, or ways in which crew meetings were conducted.
- An electronic version of the modified HERA form was produced. This form included the variable PSF items, added items, and working definitions. The form was completed by the analysts immediately after viewing the DVD recordings of a particular run.
- Once all selected runs had been analyzed, data from all runs were aggregated, and the constant PSFs were added. This data set represents the basis from which the higher-level PSF aggregations and summaries in the following sections were produced.

[5] Constant and variable PSFs are an extension of *static* and *dynamic* PSFs, as described in Boring (2007).

2.7 Overall Performance of Individual Crews and Aggregated Operational Stories

Note that the naming of crews A-N in the following sections is randomized, and does not represent the sequence in which they participated in the data collection.

An aggregated operational story was generated from the individual crew summaries (c.f. Appendix A) to obtain a narrative that can be compared to the HRA submissions. Note that these stories, presented in Sections 2.7.3 and 2.7.4, are based on all 14 crews, not just the ones selected for in-depth analysis.

2.7.1 Overall Performance of Individual Crews on HFE 1A (SGTR Base Case)

Table 2-4 shows the performance time for the scenario section corresponding to HFE 1A, and the SG level at the time of isolation. Performance times range from 10 to 21 minutes, with a median of 15 minutes. Only one crew (N) failed the HFE-1A criterion of 20 minutes.

Table 2-4. SGTR Base (HFE 1A) Performance (Grey = Selected for Further Analysis)

Crew	HFE-1A* Criterion 20min	SG level NR at isolation
M	0:10:23	20
H	0:11:59	10
L	0:13:06	6
B	0:13:19	21
A	0:13:33	17
I	0:13:37	31
E	0:14:22	40
K	0:15:09	39
D	0:16:34	55
J	0:17:38	44
G	0:18:38	39
F	0:18:45	73
C	0:18:53	57
N	0:21:29	75

* Time from start of leakage

2.7.2 Overall Performance of Individual Crews on HFE 1B (SGTR Complex Case)

The performance on the complex SGTR scenario is more diverse than the performance in the base case. The range is 20 to 45 minutes, with a median of about 26 minutes. Half the crews do not meet the HFE 1B timing criterion of 25 minutes. Three of the crews (F, J, and E) miss the criterion by five minutes or more.

Table 2-5. SGTR Complex (HFE 1B) Performance (Grey = Selected for Further Analysis)

Crew	HFE-1B Criterion 25min*	SG level NR at isolation
L	0:19:59	78
B	0:21:10	100
I	0:21:36	70
M	0:22:12	81
G	0:23:39	88
N	0:24:37	86
H	0:24:43	91
K	0:26:39	64
D	0:27:14	100
A	0:28:01	100
C	0:28:57	99
F	0:30:16	100
J	0:32:08	100
E	0:45:27	98

* Time from start of leakage

2.7.3 Aggregated Operational Story on HFE 1A (SGTR Base Case, All Crews)

The crews' progression through the emergency operating procedures (E-0 and E-3) in the SGTR base case is documented in Tables 2-6 and 2-7. Right after the start of the steam generator leakage, all crews detected the radiation alarms, and, one to three minutes later, they manually tripped the reactor. It took one minute on average to start E-0 after the manual trip. Four crews manually started the Safety Injection before entering E-0 (half a minute on average before entering E-0), and therefore used less time at step 4 (determining if SI was required).

Twelve out of fourteen crews stopped AFW to the ruptured SG before E-3 (typically when reading the note in E-0, step 12), five minutes on average after reactor trip and four minutes after the start of E-0.

Following the start of E-0, it typically took six to seven minutes for the crews to enter E-0 step 19 ("Check if SGs are not ruptured"). The identification here is based on checking radiation from:

- SG sampling,
- condenser air ejector, and
- steamlines.

All 14 crews transferred from this procedure step to procedure E-3, based on the radiation indications.

Switching procedures typically requires a consultation between RO and SS. At this stage, although the situation was straightforward, several crews decided to hold a meeting to assess the status and develop a strategy.

All crews entered E-3 with the specific guidance of procedure E-0 step 19. It took an average seven minutes to transfer to procedure E-3 after the start of E-0 (compared to an average of almost 21 minutes in the complex case).

The isolation took, on average, five minutes (six in the complex case). Some crews had problems (e.g., missing a local action) in E-3 step 3, due to some simulator/home plant differences[6] and to the complex build-up of this procedure step (eight points, combining local and control room actions); however, the majority of crews performed this step well in the base case.

Table 2-6. Crews' Procedure Progressions in Base Case
Times From Beginning of Simulation

Crew	Start E-0 (at reactor trip)	Enter step 19 in E-0 (identification)	Enter E-3* (all crews entered from E-0 step 19)	Isolate faulted SG according to E-3, step 3 a-h
A	00:05:31	00:12:12	00:12:21	00:16:44
B	00:04:42	00:09:46	00:10:49	00:16:28
C	00:06:05	00:13:30	00:16:57	00:22:35
D	00:04:43	00:14:07	00:14:15	00:19:44
E	00:04:55	00:12:28	00:12:30	00:17:35
F	00:07:31	00:12:21	00:13:34	00:21:58
G	00:10:54	00:15:31	00:16:31	00:22:11
H	00:06:50	00:12:02	00:12:27	00:15:34
I	00:06:06	00:11:42	00:12:44	00:16:49
J	00:07:34	00:12:47	00:13:28	00:21:15
K	00:06:30	00:11:58	00:12:38	00:18:20
L	00:04:54	00:11:13	00:11:18	00:16:15
M	00:05:03	00:11:01	00:11:03	00:13:36
N	00:05:30	00:16:25	00:19:23	00:24:44

[6] The turbine-driven auxiliary feedwater (AFW) pump steam supply valve should, according to procedure E-3 step 3, be closed from the control room if possible (sometimes there is a signal (ATWT) that prevents this). Then a field operator needs to locally ensure that the valve doesn't open again (automatically on ATWT signal). At the crew's home plant, this action can only be done locally by manually closing a valve (versus closing air to AFW pump steam supply valve in the simulated plant).

Table 2-7. Crews' Performance Times for Selected Activities in Base Case

Crew	Detect Activity Alarm*	Manual Reactor trip	Communicate SGTR	Start E-0 (at reactor trip)	Manual Start of SI when level prz < 10%	Stop auxiliary feedwater to faulted SG before E-3	Enter step 19	Transfer to procedure E-3
A	00:00:00	00:01:39	00:01:26	00:01:52	00:02:37	00:03:42	00:08:33	00:08:42
B	00:00:00	00:00:41	00:02:43	00:01:28	00:01:11	00:04:30	00:06:32	00:07:35
C	00:00:00	00:01:24	00:01:27	00:02:05	00:04:04	00:08:04	00:09:30	00:12:57
D	00:00:00	00:00:40	00:00:36	00:01:28	00:03:22	00:06:25	00:10:52	00:11:00
E	00:00:00	00:00:31	00:02:45	00:00:45	00:03:15	00:05:07	00:08:18	00:08:20
F	00:00:00	00:03:58	00:03:52	00:04:10	00:04:31	-	00:09:00	00:10:13
G	00:00:00	-	00:09:12	00:07:06	00:06:15	00:09:26	00:11:43	00:12:43
H	00:00:00	00:02:52	-	00:03:07	00:03:45	00:06:40	00:08:19	00:08:44
I	00:00:00	00:02:08	-	00:02:44	00:04:26	00:06:57	00:08:20	00:09:22
J	00:00:00	00:03:27	00:01:38	00:03:47	-	-	00:09:00	00:09:41
K	00:00:00	00:02:15	00:02:21	00:03:06	00:02:50	00:10:08	00:08:34	00:09:14
L	00:00:00	00:01:23	00:00:43	00:01:42	00:01:25	00:05:11	00:08:01	00:08:06
M	00:00:00	00:01:16	00:03:58	00:01:40	00:03:18	00:05:11	00:07:38	00:07:40
N	00:00:00	00:01:45	00:00:57	00:02:12	00:04:22	00:08:23	00:13:07	00:16:05

***Time of detection is almost coincident with start of leakage (one to two seconds later)**
- Missing data in OPAS

2.7.4 Aggregated Operational Story on HFE 1B (SGTR Complex Case, All Crews)

After the start of E-0, it typically took a couple of minutes for the crews to enter E-0 step 9 ("Check if Main Steamline should be isolated," see Table 2-9). This is an important step in the complex case, as it should remind the crews that the steamlines are isolated. At about nine minutes after the start of E-0, most crews entered step 19, "Check if SGs are not ruptured" (Table 2-7). The identification is based on checking radiation from:

- SG sampling,
- condenser air ejector, and
- steamlines.

Most, if not all, crews did not consider the fact that the steamline isolation would prevent radiation from being measured in the steamlines. Two to three minutes later, the crews arrived at step 21, "Check if SI should be terminated" (Table 2-7). This was an important decision point in the complex scenario, as:

- Level increase in SG1 (caused by the tube leakage) was by this point clearly diverging from the other two SGs.
- Procedural subpoint (c) requires the crew to assess whether the RCS pressure is stable/increasing or not. In the latter case (pressure decreasing), the crew should continue to step 22; otherwise, they would have to transfer to ES-1.1 ("SI termination").

Changing procedures typically requires a consultation between RO and SS. At this point, conflicting bits of information had to be considered:

- The trend in RCS pressure (it was often just beginning to decrease when RO first assessed it) and the cooling down, which could have explained it.
- The status of the secondary system (steamline isolation, FW and AFW to SGs and related pumps, and valves status).
- The increasing level in SG1 without radiation indication.
- The communication with the chemical department and FOs about the conditions in the turbine hall (e.g., presence of steam).

The following procedure progressions were observed in the complex case (Table 2-8):

- Six crews entered ES-1.1, "SI termination." Of these:
  - Two read the foldout page carefully and transferred to E-3 on SG1 level.
  - Four eventually returned to E-0 and transferred to E-3 from step 19. Of these four, three did so based on the SG level, one (the slow crew) on radiation. Those crews basing the transfer to E-3 on SG level divergence were making a knowledge-based transfer, since step 19 does not include SG level as a criterion.
- Four crews went directly to E-3 from E-0 step 21. One crew confirmed the decision by looking ahead at the foldout in ES-1.1, while the other three appeared to make a knowledge-based decision.
- Two transferred to E-3 from E-0 steps 24 and 25.
- One crew transferred to E-3 from E-0 step 19 based on radiation. This crew was delayed in E-0 due to their early manual steamline break identification and isolation, and needed to manually start SI at step 4.
- One crew entered E-2 from E-0 step 14 (second loop), then transferred to E-3 from E-2 step 7.

The majority of the crews entered E-3 without specific guidance from a procedure point (five were guided or confirmed in their decision by finding an applicable transfer point).

It took an average of almost 21 minutes to transfer to procedure E-3 after the start of E-0 (compared to an average of seven minutes in the base case), while the isolation generally took six minutes (five in the base case). Many crews had problems (e.g., missing a local action) in E-3 step 3, due to some simulator/home plant differences and to the complex build-up of this procedure step (eight points combining local and control room actions).

Table 2-8. Procedure Progressions and Basis for Transfer to E-3 in Complex Scenario

Crew	Point of transfer to E-3	Basis for transfer to E-3
A	E-0 step 21 – ES-1.1 foldout page	**SG level**
B	E-0 step 24	**SG level**
C	E-0 step 21	Knowledge-based (level)
D	E-0 step 24-25	Knowledge-based (level)
E	E-0 step 21 – ES-1.1 – E-0 step 19	**SG1 gamma levels 1 and 2** (slow crew)
F	E-0 step 21 – ES-1.1 – E-0 step 19	Knowledge-based (level)
G	E-0 step 21	Knowledge-based (level)
H	E-0 step 21 – ES-1.1 – FR-H5 – E-0 step 19	Knowledge-based (level)
I	E-0 step 21 – ES-1.1 – E-0 step 19	Knowledge-based (level)
J	E-0 (second loop) step 14 – E-2 step 7	Knowledge-based (the transfer point in E-2 step 7 does not mention SG level)
K	E-0 step 19	**Gamma radiation** (The crew manually trips the reactor as they identify steamline break. The crew gets some delay because of these actions and the need to manually start SI at step 4).
L	E-0 step 21	**Knowledge-based (level) + ES-1.1 foldout**
M	E-0 step 21 – ES-1.1 foldout page	**SG level**
N	E-0 step 21	Knowledge based (level)

Bold: Decision guided or confirmed by procedure transfer point

Table 2-9. Details of Procedure Progression for Selected Crews (Complex Case) Times From Start of Simulation

Crew	Start E-0	E-0 Step 9	E-0 Step 19	E-0 Step 21	Start ES-1.1	E-0 restarted		
C	4:10	6:00-RO reads the step to check criteria for isolation, even though he has just received information that steamlines are isolated. ARO says that they have steamline isolation. RO now understands the information. Still no consideration of a steamline break. (No one questions why they have steamline isolation.)	11:00-Some confusion as to whether the sampling valves from SGs are open (difference Home-HAMMLAB). This is quickly solved. AFW to SG1 reduced.	13:45–Crew closes main FW and AFW to SG1, and ensures that feedwater valves and pumps are closed. Crew cannot explain SG1 level without radiation and lack of level in PRZ. Long discussions and unstructured meetings on whether to transfer to ES-1.1 (favored by SS) or E-3 (suggested by RO). At about 20:30 SS decides **to go to E-3**.				33:15–E-3 started at 24:20. Lack of focus on quickly isolating: SG1 NR level 72% when E-3 started, 99% when ended.
F	5:42		11:20-The checks in step 19 are not reported properly since ARO also looks at containment activity in step 20.	13:15-RO interprets RCS pressure as stable. It seems like RO takes the value at the given time instead of the trend, which is decreasing. RO: We **go to ES-1.1.**	13:30–They take two meetings and at 19:20 the SS orders a new diagnosis by going **to E-0.**	At 20:45-Crew is at step 12. SGTR mentioned but not acted upon. At 25:00 RO is at Step 14, right column c. Trying to perform a Steamline Isolation (they have automatic Steamline Isolation from the beginning of the scenario). Shortly after the isolation, they get reports back from the FO that there has been steam in the turbine building, and it is clearing away. Together with an increasing PRZ pressure, this is interpreted as a secondary break that has now been isolated. At 29:00 crew is at step 19. At 30:40 SS orders **transfer to E-3**. RO is hesitant (no radiation).	36:10-	
I	4:54		9:22	11:56 transfer to ES-1.1.	13:00–RO starts ES-1.1, opens the foldout page but does not read it carefully.	17:30-Crew restarts E-0. SS now leads the procedure work and quickly checks some selected steps in E-0. SS orders RO to do step 18 and 19 ("Check if SGs are not ruptured"), but they still have no activity. At 20:00 SS says that they could get water from somewhere else and they close the normal feedwater bypass valve to make sure they have no flow to SG1. At 20:30 SS decides to **go to E-3**, and orders the transfer.	26:53–E-3 started at 24:00	
J	4:25	6:20–Crew checks criteria to stop SI.	10:30–They take some time to check if sampling is open.	13:15-RO and ARO conclude that the RCS pressure is decreasing and continue in procedure E-0. At 13:20 the crew is in step 24. ARO does not answer to step b. Right column not read, SG level explained as FW break. Small RO-SS consultation, FO sent to check for indications of FW leakage.		18:50–Step 14, right column point c, RO wants to actuate steamline isolation. At 21:00 crew **transfers to E-2** from step 14. Step 18 not checked for confirming decision. 27:00–from E-2 step 7 **transfer to E-3**.	36:40	
L	4:25		11:00	14:50–ARO communicates large level increase in SG1. Step c, slightly decreasing RCS pressure interpreted as stable. RO wants to go to ES-1.1 while ARO suggests an SGTR, and RO agrees. Crew orders sampling and checks that all AFW has been closed to SG1. At 18:36, SS orders **transfer to E-3.** At 19:13, RO checks foldout page of ES-1.1, which confirms decision to enter E-3.				24:36–E-3 started at 19:35.
M	4:33	6:00–ARO communicates SL isolation.	9:55–Crew controls radiation from steamline radiation.	11:50–RO concludes RCS pressure decreasing and wants to go to step 22. SS says, "we are cooling down...have to decrease the AFW flow." At 13:17, RO suggests ES-1.1. SS: "No, meeting." Crew stuck in step, meeting is loosely held. At 17:42 SS decides to go to ES-1.1.	20:50–SS reads foldout page. Discussion with RO. At 23:50 **transfer to E-3.**			26:50–Good RO-ARO task allocation.

2.8 Identification of Driving Factors–Aggregated Across All Crews

This section covers the final phase of the analysis of the simulator reference data, namely the identification of driving factors or PSFs.

2.8.1 Performance Shaping Factors (PSFs) Used in the Empirical Study

2.8.1.1 Background to PSF Selection

For a number of reasons, the PSFs used during quantification are different for the various HRA methods. In deriving a set of driving PSFs, one needs to account for:

- Which PSFs were used in the methods.
- The factors used in task analyses.

In HRA practice, task analyses are sometimes performed solely to address the performance shaping factors included in the specific HRA quantification method used in the PRA. On the other hand, a task analysis that examines all of the PSFs identified in the NRC's Good Practices (NUREG-1792) [7] has a broader scope. As is the case with the Technique for Human Error Rate Prediction (THERP), the method does not provide an explicit model of the quantitative impact of all of the PSFs identified in the THERP Handbook; as a result, the factors treated in the methods may differ from the broader set considered in the qualitative analysis.

Table 2-10 shows the PSFs used in selected HRA methods as well as those identified (for post-initiator actions) in the HRA Good Practices (NUREG-1792), and matches these with the PSFs selected for the data aggregation and comparison in the Empirical Study. It is worth noting that some PSF taxonomies used in other methods (not shown in this table) differ significantly from these taxonomies.

Table 2-11 shows the Empirical Study PSFs and those PSFs used in analyses of operational experience and events. It can be seen that some factors used in retrospective analyses arise because an event is a specific performance instance, whereas HRAs need to consider a range of possible performances. In this table, the PSFs from NUREG-1792 are also presented.

It is important to consider a number of the PSFs' characteristics:

- Scenario complexity vs. execution complexity: Scenario complexity is the difficulty of understanding the scenario (detection/diagnosis) and how to respond appropriately to the scenario. Execution complexity has to do with the difficulty of the task, such as the number of task steps, the possible non-linear response of the system, the need to coordinate aspects of the response, the timing of the individual steps, etc. While many methods combine these two types of complexity, they have been treated distinctly in the Empirical Study because each covers distinct areas.
- Work Processes, Communication, and Team Dynamics are factors that, in a retrospective analysis, are frequently identified as drivers or contributing factors to an event. While elements of these factors may be shared across all crews, these factors

are often related to crew-to-crew variability. This crew-to-crew variability is not treated as such in a predictive analysis; instead, as noted in the PSF working definitions (see Table 2-12), these factors reflect the potential sensitivity of task success to the quality of crew performance with respect to these factors. The HRA methods that do treat these factors tend to be the newer methods.

- Additionally, some HRA methods address PSFs that are more relevant for "local" actions (actions outside the control room). Since this study is performed in the control room, these are not relevant, and have not been used; examples include performance conditions, such as lighting, noise, and vibration.

Table 2-10. PSFs in HRA Methods Vs. PSFs Used in Empirical Study

Selected HRA Methods and HRA Good Practice			
PSFs in SPAR-H (NUREG/CR-6883)	**PSFs in PLG SLIM**	**PSFs in HRA Good Practices (NUREG-1792)***	**PSFs used in aggregation/ comparison**
Available time	adequacy of time	Time available and time required	Adequacy of time
Stress/stressors	Stress	Workload, time pressure, stress	Stress
	Preceding and concurrent actions [task load and focus of attention]		(not used)
Complexity [both scenario and execution complexity]		Complexity of required diagnosis and response [combines scenario and execution complexity]	Scenario complexity
	Task complexity [number and difficulty of sub-steps, communication and coordination requirements]		Execution complexity
Experience / training	Training and experience	Training and experience	Training
			Experience
Procedures	Procedural guidance	Procedures and administrative controls	Procedural guidance
Ergonomics/HMI [includes scenario-specific availability of information]	HMI and indications of conditions	Availability and clarity of instrumentation (cues)	HMI and indications of conditions
		Ergonomic quality of HSI	
		Accessibility and operability of equipment	
Fitness for duty			(not used)
Work processes			Work processes
		Team/crew dynamics and crew characteristics	Team dynamics
	["Communication" partly covered by SLIM PSF 'task complexity']	Communications	Communication
		- Available staffing and resources - Special tools - Special fitness needs	(not used)
		Environment	(not used)
		Consideration of scenario diversions and deviations	(not used)

* Refer to NUREG-1792, Table 5-1, for the full names of these PSFs. The labels shown here are shortened.

Table 2-11. PSFs Used in Empirical Study Compared to Retrospective Factors

	Predictive	*Retrospective*
PSFs used in aggregation/ comparison	**PSFs in HRA Good Practices (NUREG-1792)**	**PSFs in HERA (NUREG/CR-6903)**
Adequacy of time	Time available and time required	Available Time
Stress	Workload, time pressure, stress	Stress & Stressors
Scenario complexity Execution complexity	Complexity of required diagnosis and response [combines scenario and execution complexity]	Complexity
Training and experience	Training and experience	Experience & Training
Procedural guidance	Procedures and administrative controls	Procedures & Reference Documents
HMI and indications of conditions	Availability and clarity of instrumentation (cues) Ergonomic quality of HIS Accessibility and operability of equipment	Ergonomics & HMI
		Environment
Work processes		Work Processes
Team dynamics	Team/crew dynamics and crew characteristics	Team Dynamics/Characteristics
Communication	Communications	Communication
	- Available staffing and resources - Special tools - Special fitness needs	Fitness for Duty/Fatigue

It should be noted that the integration of performance in terms of main PSF drivers does not try to use a single "orthogonal" set of PSFs. Some of the PSFs are partly overlapping and would be inappropriate to use in a statistical analysis, where independence of factors is assumed. Non-recognition of the overlap of factors risks double-counting the influence of specific factors, though this form of double-counting is allowed because:

- The choice of PSF may depend on a method, and we want to be able to match as many as possible, that is, not to give preference to one set of PSFs.
- Some PSFs seem to be defined in terms of an interaction; for instance, scenario complexity is raised by lack of training and/or unhelpful procedures.

As long as a consensus set of PSFs and PSF definitions does not exist across HRA methods, we have to avail ourselves of PSFs commonly found in HRA, even at the risk of double-counting. It is worth stressing that this "main driver" story is not a performance model (i.e., convertible into a human error probability (HEP) or even a performance ranking). Instead, it is an explanation of the observed performances in terms of factors that may be familiar to the HRA community, given that this community has different ways of expressing the same thing.

2.8.1.2 PSFs Used for Comparison and Data Analysis

The qualitative predictions consist of the factors that are identified as important to performance. Nearly all HRA methods use such performance factors, or PSFs. There are significant differences among the sets of PSFs used in different HRA methods. In addition, while the factors may be labeled differently in the methods, one may generally say that the PSFs correspond to those factors that would be used in a task analysis to characterize the performance conditions for a task[7].

While the analyses performed by the teams used the PSFs of the respective methods, a single set of PSFs was needed in the Empirical Study. The selected PSFs affected:

- The analysis of the Halden data.
- The "normalization" of the HRA team predictions, that is, expressing the predicted outcomes in a "common language."

2.8.1.3 PSFs Used in First Pilot Phase–Working Definitions

Table 2-12 lists the PSFs used in the first pilot phase and their working definitions, which are largely based on the definitions provided in Appendix B of NUREG-1792 (Good Practices). In a number of cases, distinct factors are used for the different components of a factor treated together in some methods; for instance, the factors "training" and "experience" are treated separately, although some methods address these together under "training and experience."

[7]It is worth noting that there is no single methodology recognized for task analyses. This is partly because the performance conditions and the sensitivity of task performance to a given performance condition are specific to the application domain.

Table 2-12. Performance Shaping Factors–Working Definitions For First Pilot Phase

Factor	Working definition
Adequacy of time	Adequacy of time relates to the difference between available time and the required time. Available time is estimated based on an expected evolution of the scenario, which determines when the action modeled by the HFE can no longer be effective in reaching the success criteria. The required time is an estimate of the time needed by the crews to perform the cognitive and execution components of the task. Adequacy of time affects the assessment of the HEP simply because there may not be enough time to finish the required actions, as well as to check the performance of the action, and to detect and correct errors.
Time pressure	Time pressure refers to the crews' subjective perception that there is a limited amount or shortage of time in which to accomplish the required tasks. In many methods (and in NUREG-1792), time pressure is addressed as a component of or contributor to stress. The crew's perception of the available time can differ from the time actually available in the scenario. Consequently, the crews may experience or report time pressure when the adequacy of time is good; conversely, they may not feel time pressure, although the adequacy of time is poor.
Stress	Stress refers to the deleterious effects caused by high workload, perceived time pressure, urgency, or perceived threat to performance.
Scenario complexity	Scenario complexity is the difficulty of situation assessment and diagnosis. It is related to "masking," diagnosis complexity, the need to decipher numerous indications and alarms, and the ambiguity associated with assessing the situation. In many PSF frameworks, this factor relates to the indications of conditions (availability of cues, ease of perceiving these cues, or the difficulty of interpreting these indications).
Indications of conditions	Indications of conditions refers to the availability and clarity of key indications and/or alarms. This is affected by the availability of instrumentation, and, given that the instrumentation is available, the salience of cues, signal-to-noise, and ambiguity of cues. In some cases, it also refers to the availability of system feedback for execution. This factor is often treated in "scenario complexity," although the latter has a larger scope.
Execution complexity	Execution complexity captures the difficulty of performance (implementation) of the task (not including situation assessment, diagnosis, etc). Execution complexity is influenced by the number of steps to be performed, whether the task is associated with a single variable or with multiple variables, and whether special sequencing or coordination of multiple performers is required.

Factor	Working definition
Training	Training details the crews' degree of familiarity with the scenario and the actions that can be expected based on their training. It includes both theoretical (classroom) knowledge and practice (e.g., in a training simulator). This factor should consider not only the amount or general quality of training, but also the applicability of the training in a specific scenario, that is, how helpful it will be in the scenario. In rare cases, the training may even be counterproductive. In predictive analysis, training as a factor is frequently combined with experience.
Exper-ience	Experience is the personnel's familiarity with the task being analyzed. Although correlated, it is not strictly equivalent to the crew's amount of experience (e.g., number of years in position). Like training, in rare cases, experience may be counterproductive.
Procedural guidance	Procedural guidance refers to the support provided by the procedure for performing the situation assessment (decision making) and execution of the specific task being analyzed. In the context of the scenario of interest, steps *that are ambiguous, unclear, or not detailed, and situations where the* procedure is unclear, can contribute to a poor rating for this factor.
Human-machine interface	Human-machine interface broadly encompasses ergonomics, including the presentation and labeling of process parameters, the availability of feedback following an action on a component or system, and the interface for acting on components or systems.
Work processes	Work processes refer to the way of working and the mechanics of work, such as the care taken in reading procedures, and, more generally, in performing individual work. In a predictive analysis, this factor indicates how sensitive the task may be to work practices. In analyzing actual performance, this factor is rated poor if work is not thorough, decisions are made without review, or there is poor general handling of procedures. Note that in fast-moving scenarios, "good" work processes may have a negative effect on task success, as thorough work may proceed too slowly to address the situation at hand.

Factor	Working definition
Communi-cation	In a predictive analysis, communication refers to: a) The impact of the environment, such as noise and the hardware used for communication (e.g., an intercom), on task success. b) The "communication requirements" of the task. These requirements may be viewed as contributing to scenario complexity or task complexity (depending on whether the communication is about situation assessment or about procedure). In actual performance analysis, this factor refers to the use of communication protocols, such as repeat-back, and to the quality of communication, such as failure to provide explicit feedback.
Team dynamics	Team dynamics relate to the management of the team, the adequacy of interactions, the sharing of information, proactive communication, or the treatment of suggestions. In a predictive analysis, this factor represents the requirements of the task in terms of good team dynamics, such as the sensitivity of a task to the quality of team dynamics.

2.8.2 Rating of Driving Factors

In an HRA, the rating of a PSF refers to the quality of a PSF—in other words, how positive or negative is the factor contributing to the PSF under consideration. In addition, some methods weight the PSFs; the PSF weight refers to the expected performance's sensitivity to this PSF. Note, however, that some methods have a fixed PSF weighting. For instance, when the PSF rating has a specified multiplicative impact on a nominal HEP, this implies that the PSF affects the HEP of all tasks to the same degree.

In the Empirical Study, the "main drivers" of performance, whether positively or negatively rated, are generally factors that are considered to have significant weight.

The Empirical Study intends to shed light on:

- The HRA methods' ability to accurately identify and rate the performance conditions. Do the methods predict which factors will affect performance?
- The HRA methods' reflection of the PSFs in terms of their impact on performance and on the HEP. Do the methods overestimate or underestimate the impact of a poor PSF on performance, or can the PSFs be addressed in the estimation of the HEPs?

Consequently, a qualitative scale is needed for the rating of the driving factors. This section discusses the qualitative scale that is used, shown in Table 2-13.

Table 2-13. Driving Factor Rating Scale

POSITIVE			NOMINAL			NEGATIVE
very good	good	somewhat good	nominal/ average	somewhat poor	poor	very poor
			nominal	somewhat high	high	very high

The PSFs as defined in the Empirical Study are rated on a seven-item scale from "very good" to "very poor."

It is worth noting that, in some HRA methods and analyses, one-sided scales ranging from nominal to "degraded" or "very poor" are used. In some cases, this is because the methods do not take credit for "better than nominal" conditions and consider only factors that would degrade nominal performance. Fortunately, a rating on a one-sided scale is often converted to a two-sided scale rating on a related PSF.

For instance, the PSF "misleading cues," which can be viewed as a one-sided scale from "not misleading" to "very misleading," can typically be expressed as a very negative rating on a two-sided scale, that is, misleading cues can be expressed as indications of conditions–very poor.

It would be optimal if "anchors" were defined for the rating scale, that is, descriptions of the performance conditions for which each rating would apply (what would be considered very poor, poor, good, etc). For instance, this exists explicitly for very few methods, aside from the numerical Pickard, Lowe, and Garrick (PLG) Success Likelihood Index Methodology Scale (SLIM) scale. In most methods (e.g., THERP, SPAR-H), a description is given of various issues to consider in rating a particular PSF, but no anchored rating scale is provided. In this work (the Empirical Study), we do not attempt to derive an anchored rating scale.

The ratings of the driving factors or driving PSFs thus combine the PSF rating (how good or how bad) with the strength of the factor's influence (how important). This represents the overall significance, or relative strength, of each driving factor.

2.8.3 Methodology for the Identification of Driving Factors

As noted in Chapter 1, one of the dimensions of the comparison of predicted outcomes with the empirical, observed outcome consists of the factors that "most influence" performance. This section presents the methodology for identifying these "driving PSFs" from the observations of the crew performances.

The performance of the individual crews in a given scenario can be expected to vary. A number of performance criteria may be used to characterize this variability. In binary terms, one of these criteria is meeting the success criteria for the HFE. However, a number of other, more continuous performance criteria are also applicable, such as margin with respect to physical parameter limits, level of situation awareness, and performance time. Furthermore, the performances will be diverse in terms of the PSFs that have been observed to drive the performances of the various crews.

The selection of the PSFs was presented earlier. Additionally, the identification of the PSF main drivers was based on several principles:

- *Considering the driving PSFs in both positive and negative terms.* Recognizing that some factors could have positive and negative aspects in connection with the different elements of a task. This has been addressed by assessing an overall rating for the PSF while noting any aspect where the PSF had the opposite effect. In other words, the PSF could be rated negatively overall, but it could still have positive elements.
- *Contrasting teams at both ends of the performance spectrum.* To increase the contrast among the performances, the driving PSF identification focused on both ends of the performance spectrum, which, in accordance with the HFE definition, was defined in terms of performance time. In practice, two to three crews were used for each end of the performance spectrum. Subsequently, information from the remaining crews was used to confirm and/or extend the tendencies identified from the analysis of those selected crews.
- *Comparing the factors in the base case scenario with the factors in the complex case scenario.* Factors that impacted all crews in the same way in a scenario variant tended not to be highlighted when analyzing the observations of the crews, since the observations tended to focus primarily on the differences among the crews. Contrasting these factors across the scenarios helped to identify the factors that are different or more problematic in the complex case.

Step 1: Summarizing main factors influencing the fastest and slowest crews (per case)

To support the identification of PSF main drivers, the identified influencing factors were represented in a 2x2 matrix for the base and complex cases, respectively. Table 2-14 shows the structure of the influencing factor matrix. This matrix provides an overview of the factors in the fastest and slowest performers so that the comparison and contrast can be made. (Tables 2-15 and 2-16 show these matrices for the SGTR base case and complex cases, HFE 1A and 1B, respectively.)

Table 2-14. Structure of the Influencing Factor Matrix ("2x2" Matrix)

	HFE success/fastest (2-3 crews)	HFE failure/slowest (2-3 crews)
positive factors observed		
negative factors observed		

Step 2: Contrasting fastest and slowest performers (per case)

The negative factors in the fastest and slowest performing crews were compared. Factors common to both the fastest and the slowest performers provide a first-level characterization of the PSFs. In addition, the factors unique to the slowest performers point to additional PSFs that appeared to have a negative influence on performance.

The positive factors in the fastest and the slowest performers were also compared. Here, the factors unique to the fastest performers point to PSFs that could have a positive influence if present.

Step 3: Interaction of positive and negative factors (per case)

This step focuses on the work of the fastest performers, with the aim of learning how they were able to "compensate" for the negative factors in order to achieve their level of performance. What factors were present for these crews, that may not have been present for the slowest performers? This step allows a lower-level understanding of the factors important to performance in the scenario.

Step 4: Contrasting base and complex case scenario performance (comparison across cases)

When the PSFs in the base case vs. the complex case scenario are compared, factors that are unique to the complex case or that have a different effect in the complex case are highlighted. Additionally, in the base case, where performance would generally tend to be better, it highlights the positive factors or the lack of negative factors that make the base case less difficult. The positive PSFs for the base case are largely identified by the lack of corresponding negative PSFs in the complex scenario. The matrix in this case is a combination of the previously used matrices (Tables 2-15 and 2-16).

It is worth noting that this step would have to be modified if, for causes unforeseen in the scenario design, the performance related to an HFE in the base case is in fact worse than in the complex case scenario. (This was not the case for HFE 1A/HFE 1B.)

As a combination of the PSF rating (how good or how bad) and the strength of the influence (how important), the final driving factor rating, such as "very good" or "somewhat poor," represents the overall significance or relative strength of each driving factor.

2.8.4 Influencing Factor Matrices for HFE 1A and HFE 1B

Each of the crew operational summaries discussed earlier in Section 2.6.3 and included in Appendix A summarizes the actions and performance difficulties in nuclear power plant (NPP) operational terms for an individual crew and identifies the performance influences for this crew.

In step 1 of driving factor identification, the performance influences underlying the individual crew performance are first summarized for the fastest and slowest performers. Tables 2-15 and 2-16 contrast the performance influences for fast vs. slow crews for HFE 1A in the base case and HFE 1B in the complex case, respectively.

Table 2-15. Influencing Factor Matrix (HFE 1A–SGTR Base Case)

HFE 1A	Two fast crews (H, M)	One slow crew (N)
positive	**Procedural guidance**. Good match of the procedures to the scenario (all). **Scenario complexity/indications of conditions**. Low complexity, indications are clear (all). **Work processes.** Good RO-ARO coordination and communication, good procedure work (H). **Team dynamics**. SS decisive (H). SS keeps good overview of process and crew's work progress (M). **Training**. Crew easily identifies tube rupture based on their training and works ahead of the procedures (high degree of familiarity with the procedure, which means training is good) (all).	**Procedural guidance** (as for fast crews). Good match of the procedures to the scenario (all). **Scenario complexity/indications of conditions** (as for fast crews). Low complexity, indications are clear (all). **Training**. Indications were detected early. No specific difficulty was observed with the diagnosis or the isolation (N).
negative	**No direct negative influences.** No "negative factor present" identified (no negatively rated factor without observable impact) (all).	**Training**. ARO did not use the large overhead screen efficiently and used time to navigate to the appropriate screen at the workstations. SS does not focus on overfilling the SG. SS did not focus on speeding up the work (unaware of scenario dynamics). SS interrupts sometimes with less important things (N). **Work processes**. Crew follows good practices (meeting at procedure transfer, thorough checks and verifications). "Whole crew were clearly updated and coordinated on the situation and chosen strategy," but thoroughness and unwarranted attention to detail slows them down in this scenario (N). **Team dynamics**. RO waits for ARO and does not work independently (also slows down crew) (N).

The influencing factor matrix is the first step in the analysis of the driving factors, summarizing the observed influences on performance. As shown in Tables 2-15 and 2-16, the influencing factors are listed as positive or negative; together with specific elements of the crew operational summaries (cf. Appendix A). Some influencing factors may be placed in both the upper row (positive influences) and the lower row (negative influences). In Table 2-15, for example, for the slow crew, the factor "training" was assessed as positive based on the way the crew managed the diagnosis and the isolation, but also appears among the negative influences for this crew in terms of the SS's prioritization of tasks and lack of attention to the dynamics of the scenario. On the whole, it can be seen that for all crews most of the factors in the base case were assessed positively. The negative factors that contribute to the performance of crew N appear to be specific to this crew; as a result, these factors are not assessed as driving factors when the performance of all crews is considered in the aggregate.

A comparison of these influencing factor matrices with the more extensive crew operational summaries in Appendix A shows that the specific observations have not in all cases been associated with the same performance shaping factor. One reason for this is that the crew operational summaries were derived using the HERA taxonomy (as discussed in 2.6.4), while the influencing factor matrices were based on the PSF working definitions (discussed in Section 2.8.1). The working definitions and operationalizations in these two parts of the data analysis are being improved and harmonized based on this pilot study. The differences reflect, to some extent, the different viewpoints of retrospective and predictive analysis.

As noted, the analysis based on the contrast of fastest and slowest performers is step 2 of the driving factor analysis. To arrive at the driving factors discussed in the next section, step 3 of this analysis additionally considers how the influences listed in this table may have interacted; step 4 then contrasts the driving factors in the base and the complex cases. In all of these steps, the actual observations associated with each driving factor are "carried along" to ensure that the definitions and scope of the driving factors remain consistent at each stage of the analysis.

Table 2-16. Influencing Factor Matrix (HFE 1B–SGTR Complex Case)

HFE 1B	Three fast crews (I, L, M)	Three slow crews (C, F, J)
positive	**Indications of conditions**. Diverse indications of SGTR are available and detected (all). **Training**. Crew investigates alternative cause to SG1 level increase by isolating all feedwater (L). **Experience**. SS decisive and good at prioritizing (I). **Work practices**. RO checks for alternative causes (L). **Team dynamics**. RO works independently, SS steps in and takes over when there are problems (I). RO is proactive and communicative (I). SS has good overview (I). Good teamwork, communication, and coordination (L).	**Indications of conditions**. Crew uses diverse indications to conclude SGTR (C). Diverse indications are available and detected (but not fully understood) (F, J). **Training**. Use of (attention to) redundant (diverse) indications (C).
negative	**Procedural guidance**. Transfer to E-3 is indirect when indications of radiation are lacking (E-0 step 19 only mentions activity) (all). **Scenario complexity**. Crew does not initially understand that there is a secondary break (I). **Training**. Crew misses the transfer based on SG level from ES-1.1 to E-3 on the foldout page (I). Crew fails to act on (try to understand) SG1 level rising with AFW closed (M). Crew detects SG level mismatch but fails to diagnose SGTR immediately, due to lack of radiation indication. E-3 transfer is reached because of SS decision to enter ES1.1 to stop SI, which leads to the fold-out page (M). **Team dynamics**. SS repeatedly takes over subordinates' tasks (too involved in details), with no attributable negative impacts (I). SS takes over RO tasks, develops strategy without consultation (M). **Work processes**. On E-3 step 3, omitted sub-task while trying to deal with mixed CR and local actions listed in the task (L). ES-1.1 transfer to E-3 on high SG level missed on reading foldout page (due to inattention during reading) (I).	**Procedural guidance** (as for fast crews). Transfer to E-3 is indirect when indications of radiation are lacking (E-0 step 19 only mentions activity) (all). Poor layout of E-3 step 3 (mix of CR and local actions) (all). Crew SGTR diagnosis does not result in transfer to E-3 as a result of transfer conditions specified in procedures (C). **Training/experience**. Crew spends 10 mins. in E-0 step 19. SS hesitant, very cautious, indecisive (C). Crew hesitates to conclude SGTR from available indications due to missing radiation indication. Crew difficulty in using SGTR level as an indication of SGTR (F, J). Crew fails to arrive at a coherent picture of the plant condition (F). At E-0 step 25, crew appears to rule out SGTR due to absent radiation indication and again on exiting E-2 (J). Crew neglects to review basis for transfer and to evaluate chosen strategy at E-0 to E-2 and E-2 to E-3 (J). Early secondary break hypothesis, evidence of confirmation bias interpreting indications, failure to review assessment (J). CR crew assigns a procedure with mainly CR tasks to a field operator (J). **Work processes**. Some observations of less-than-adequate procedure reading (C). Inadequate performance of E-0 step 19, omission of reporting on result of check (F). On transfer out of E-0, SG-level information reported in meeting is not assessed, meeting does not lead to a plan or clear orders (F). Second meeting is unstructured, inconclusive (F). Tendency to follow procedures literally without assessing situation. Procedure transfer/entry conditions not reviewed on transfer (J). **Work processes/Communication/team dynamics.** SS gives few direct orders and orders are not clear (F). **Team dynamics**. Unstructured meeting, crew fails to reach decisions, spends 10 minutes in E-0 step 19 (C). SS fails to "lead," "distracting" crew during isolation, commenting that SGTR diagnosis is not definite (C). SS has difficulty giving clear directions (C). SS weak in maintaining overview and situation assessment; neglects to follow up on RO suggestions (J).

2.8.5 Driving Factors for Performance on HFE1A (Base Case)

The PSF drivers for HFE 1A in the SGTR base case scenario are summarized in Table 2-17. The qualitative basis underlying their classification as positive or negative factors is discussed after the table.

As noted, although they are included in Tables 2-15 and 2-16, the driving factor ratings were not considered in the comparison to the HRA method prediction, which considered a) whether the HRA identified a given PSF as an important influence and b) the predicted direction of the influence (positive or negative, without regard for the specific positive or negative rating).

Table 2-17. Summary of PSF Drivers–Identified for HFE 1A

	Base case (HFE 1A)
Positive Driving Factors	HMI and indications of conditions–very good Training and experience–good to very good Adequacy of time–good Procedural guidance–good [*–]
Negative Driving Factors	Execution complexity–somewhat high

[*–] While overall effect is positive, this PSF had a secondary negative influence

It is worth noting that the positive PSFs were to some extent also identified by noting their absence as a negative influence in the base case scenario (as compared to the complex scenario).

Positive Driving PSFs for Base Case, HFE 1A

Adequacy of time-good. Average performance time was 15:32 in a 20-minute time window. Only 1 of 14 crews failed to meet the assumed time window for success, and its performance time was only 1min 29s after the time window. It should be stressed that fast performance is not being judged as being equivalent to good performance (i.e., good management of available time by the crew is best—that is, it results in deliberate, unrushed performance—when there is no urgency). In this light, the time was adequate for all crews.

Training and experience–good to very good. Training for SGTR scenarios is normally held twice a year in the simulator.

Procedural guidance–good. For the base case SGTR, the procedural guidance is good in terms of SGTR diagnosis and transfer to E-3. It relies on the radiation indication as a primary cue for the transfer to E-3, and this cue is available in this scenario.

HMI and indications of conditions–very good. The primary cues, as well as redundant cues, are available for diagnosing SGTR.

Team dynamics: These are listed as positive for Crews M and H. In both cases, however, these seem to be positive crew characteristics rather than PSFs. In PSF terms: requirement for effective teamwork/work processes–somewhat high (i.e., negative, and tending to contribute to failure). HRA teams would not be expected to address this.

Negative driving PSFs for base case, HFE 1A

Execution complexity–somewhat high. This refers to the complexity in the performance of E-3 step 3, which includes a long list of actions and contains a mix of control room and local actions, requiring coordination and prioritization.

Procedural guidance–includes secondary negative influence (overall rating is "good"). Although the procedural guidance is rated "good" for the base case, the somewhat high "execution complexity," discussed in the previous paragraph, is caused by the content of E-3 step 3. For instance, clearly separating the control room from the local actions would reduce the execution complexity, which is something to consider when rating the Procedural Guidance; nevertheless, the Procedural Guidance on the whole supports task success, so this "negative rating/influence" is considered secondary.

The positive rating (overall) for Procedural Guidance is discussed in the "Positive driving PSFs" section.

2.8.6 Driving Factors for Performance on HFE 1B (Complex Case)

The PSF drivers for HFE 1A in the SGTR base case scenario are summarized in Table 2-18. The qualitative basis underlying their classification as positive or negative factors is discussed after the table.

As noted, the driving factor ratings were not considered in the comparison to the HRA method prediction, which considered a) whether the HRA identified a given PSF as an important influence and b) the predicted direction of the influence (positive or negative, without regard for the specific positive or negative rating).

Table 2-18 Summary of PSF Drivers–identified for HFE 1B

	Complex case (HFE 1B)
Positive Driving Factors	(none)
Negative Driving Factors	Scenario complexity–high Indications of plant conditions–somewhat poor to poor [*+] Procedural guidance–poor Training–somewhat poor [*+] Execution complexity–somewhat high Adequacy of time–somewhat poor Work processes–high [requirements]

[*+] While overall effect is negative, this PSF had a secondary positive influence

Positive Driving PSFs for Complex Case, HFE 1B

The items listed in the 2x2 table (Table 2-16) under *success/positive*, such as Team Dynamics and Communication, were covered in the above discussion of *Work Processes*, a negative driving PSF. The only other PSFs contributing to success seem to be *Ergonomics and HMI* (indications of plant conditions) and *Training and Experience.*

Indications of plant conditions–includes secondary positive influence (overall rating is "somewhat poor to poor"): It is worth noting that there are diverse indications available and that the crews seem to have diagnosed the SGTR on the basis of these indications (particularly SG level mismatch). Although there was ambiguity early in the scenario, the crews would have been "dead in the water" without the SG level indications.

The negative rating (overall) is discussed in the "Negative driving PSFs" section below.

Training and experience–includes secondary positive influence (overall rating is "somewhat poor"): Training and the knowledge base of the operators contributed to the fact that the crews transferred to E-3 (within the time window or within the subsequent 10 minutes).

The negative rating (overall) is discussed in the "Negative driving PSFs" section next.

Negative driving PSFs for complex case, HFE 1B

Scenario complexity–high: This refers to the masking effect, specifically, the lack of one of the main indicators (radiation indications) used for diagnosis of SGTR and for transfer to E-3. This complexity affected all crews, and interacts with the lack of specific guidance and training on reconciling the set of [available] indications with an SGTR within the defined time window.

Indications of plant conditions–somewhat poor to poor: This is an alternate way to refer to the effect of the missing radiation indications (alternate to "scenario complexity"). It is somewhat poor because several diverse indications are available (SG level mismatch, SG flow mismatch), and poor because radiation level, one of the most valuable indications for SGTR, is missing.

Procedural guidance–poor: The post-trip procedural guidance (E-0) is not very helpful in this scenario because the scenario is strongly but not solely reliant on radiation level. E-0 step 19 in particular would be the usual transfer point to E-3, but relies solely on radiation level. Other E-0 steps could potentially get the crew to E-3, but were not very effective in this exercise. The main difficulties with procedural guidance in the complex case focused on E-0 and the transfers to E-3. The indications that SG1 was ruptured helped to guide this transition. Therefore, the identification of the ruptured steam generator was accomplished prior to the transition. However, once in E-3, it is observed that the crews (in base and complex cases) may be experiencing some delays due to the execution complexity associated with performing E-3 Step 3. This step includes a long list of actions and contains a mix of control room and local actions, and thus requires coordination and prioritization in order to be efficiently performed. This aspect of the procedural guidance could be improved (i.e., the organization of this step could make this more efficient). Note that this observation is also made under "execution complexity."

Training–somewhat poor: This refers to the potential effects of training on this scenario or related scenarios. Note that it is also important to consider whether there might be training for unrelated scenarios with similar cues and indications.

The overall training for SGTR is good, but there is a lack of training on more troublesome variants of SGTR scenarios, such as this one. In particular, training that would make the operators less reliant on the radiation indications would have helped. It is also worth noting that, given the problem with the procedures, it was ultimately their knowledge base (from training) that contributed to their success.

The operators' apparent concern about adhering to procedures, as well as their search for procedural ways to get to E-3, which seemed to slow them, suggests that they had diagnosed the SGTR. Many were aware of the diverging SG levels, and several mentioned it; however, they seem to have viewed the lack of radiation indications as conflicting information, and were confused by this information. A number of crews also looked into other procedures/documents for other criteria/steps for transferring into E-3, with some success. This supports "poor" procedural guidance and only "somewhat poor" training.

Some crews had trouble transferring to E-3 based on knowledge and training, given that they are trained to "decide beyond verbatim procedure adherence." The absence of the radiation cues seemed to slow and confuse them, as most crews did not seem to notice that steamline isolation would prevent radiation from being measured in the steamlines (demonstrating a lack of knowledge or a failure to use knowledge). At a meta-level (unrelated to specific scenarios), the crews do not appear to have training on handling a conflict between their conclusions and the stated procedure.

As a final note, it is important to interpret this rating in the light of the assumed/defined time window of 25 minutes. Many crews managed to respond within the given time, and most "failures" (as defined in the study) were actually fairly close. The latter implies that their training was not great for a "quick" response (within the expected time frame), but overall their training was not bad, since they all solved the problem.

Execution complexity–somewhat high: Successful performance requires effective coordination among the crew members. Because the indications in the scenario and the procedural guidance do not match well, effective communication and coordination among the crew members is needed to avoid looping and other such problems.

The execution of E-3 step 3 is fairly complex; it includes a long list of control room and local actions, and thus requires coordination and prioritization in order to be efficiently performed. Note that this observation is also made under procedural guidance, since performance might be improved by changing this step in the procedures (i.e., the organization of this step could make this more efficient).

Adequacy of time–somewhat poor: The adequacy of time is somewhat negative, because the crews will have to reach consensus (or the leader will need to gather enough initiative) to enter E-3 without a real transfer; and, in fact, many poor performance outcomes originated from an inability to decide to enter E-3. There was, however, no evidence that the crews' performance was negatively influenced by time pressure; this factor may constrain the likelihood of success, but it does not directly influence the crews (unless time pressure hindered their decision making processes, which is unlikely in this case).

Work processes–high: "Successful performance requires effective work processes from the crew." Specifically, the evidence shows that crews need to use the appropriate format for conducting meetings (as taught in training) in order to reach the correct decision, though not necessarily to understand the scenario.

In a predictive analysis, it is very difficult for an analyst to rate this PSF. One can imagine an analysis that finds that "typical work processes" may hinder successful task performance; for instance, if the crew is trained to strive for consensus and to resolve disagreements, this will have a negative impact on tasks with limited time. One may also view this PSF as "sensitivity to work processes," which is referred to as "communication/coordination requirements." The treatment of this PSF is not fully resolved in the pilot study.

Crew dynamics, in terms of a strong SS taking charge (and in other respects), seem to have an effect in some cases, but it varies from crew to crew. The HRA teams were not expected to address this. Note that communication and work processes are closely related, as *work processes* is a broader term that includes such factors as communication and care with procedure reading.

2.9 Discussion

2.9.1 Issues on Empirical Information Aggregation and Formatting

Most of the effort put into developing the data analysis methodology has been in finding a presentation format compatible with outputs obtainable from HRA applications. The experimental results were reported in three formats:

- response times for identification/isolation and ruptured SG levels,
- aggregated operational stories for the two scenario variants, and
- aggregated driving PSFs (based on driving factors' summaries for "fastest" and "slowest" performing crews and operational stories).

These formats were chosen to allow the comparison of HRA method predictions to observed simulator performance. The response times were needed to assess the performance of the HFEs of the study. The aggregated stories were written to summarize the performance of 14 different crews (in the two scenarios) into single operational expressions, which are the typical level of representation of analyses (as a discretization of all possible scenario variants). The same goes for the summary of the driving PSFs, which could be considered the PSFs of the aggregated stories, as opposed to the various configurations of context in the individual scenario runs.

The next sections discuss a number of methodological issues related to the aggregation process that will be addressed in future phases of the Empirical Study.

2.9.1.1 Failure vs. Performance

The performance of the HFEs in the pilot study was operationalized in terms of the crews' completion time and the ruptured SG level at isolation (the lower the better). When the crews were evaluated on the performance of the HFE, a strong emphasis was given to time, with the "best" crews being the fastest to isolate, the "worst" being the slowest. This is a consequence of defining the HFE on a time criterion, although the time criterion has a strong relation to several functional goals, including the PRA-relevant one of avoiding overfilling the steam generators.

On the fine-grained level of a simulator trial, however, the speed of action can only be one of several indicators of good performance, and one that can never be isolated from the other indicators. For instance, a crew can act very quickly when a shift supervisor takes an extremely active role, decides strategies without consultation, and orders the crew to perform steps from procedures, although the latter is a reactor operator responsibility. This performance would not be optimal in terms of other indicators, as such behavior would not be considered apropos to the shift supervisor function and the training received, and would reduce the possibility for second checks, with one person centralizing all diagnosis and planning functions. It might also disrupt successive team collaboration, as the reactor operator might feel displaced from his/her functions, and thus assume either a passive or an antagonistic position.

In such cases, there is a disjunction between PSFs for the HFE, those that influence the speed of actions and quick success, and the factors that are normally considered positive influences on performance (like good consultations) but which could in the same cases slow down the performance of the HFE. In other words, there is a mismatch between the categorization required by the PRA's HFE representation and the one implicit in the reliability models (the models of human performance) of the HRA methods, as evinced in the pilot study: the PSF profile of the second fastest crew has many similarities to the profiles of the slow-performing crews in the complex scenario.

In general, there seems to be some trade-off between optimizing the comparison of the HFEs (in terms of success and failure) and the comparison the methods' ability to describe the influences on performance. For example, the approach of optimizing the comparison of the identified influences (and, hence, the validity of the reliability models) would have consequences on both the criteria for crew selection (fastest and slowest crews), and on the PSFs profiling (the "best/worse" crews' PSF profiles would have more consistency).

2.9.1.2 The Difficult Treatment of "Variable" PSFs

The derivation of the driving PSFs for the base and complex scenarios is based on the driving factors identified during the DVD reviews and summarized in the crew summaries. The DVD reviews of individual scenario runs (as well as their final aggregation and evaluation) were influenced by the HERA terminology and other HRA-specific documents. This has been challenging for the experimenters because such classifications and their definitions are not observational tools, and because they incorporate context-performance models not necessarily meant for fine-level analysis of crew behavior and interaction.

Further, a distinction was made between "constant PSFs" and "variable PSFs." Constant PSFs were considered the same for all crews:

- Scenario descriptions
- The nature of the simulated plant responses, procedures, and interface
- The plant-specific work practices of the participating crews

These PSFs could in part be "predicted" before the actual runs, although the data collected in the simulator experiment allowed them to be characterized more accurately and in more detail, that is, parts of the procedure or specific interfaces within the control room could impact different operator tasks and scenario phases to varying degrees.

Variable PSFs are those factors that are not supposed to be the same for all crews, and which had to be evaluated for each crew after the scenario run. Among the variable PSFs identified are factors that relate to crew attributes or characteristics (e.g., experience, accuracy of procedure reading, leadership style). Many of these were classified under "work practices," "crew dynamics," and "communication."

Both constant and variable PSFs can have a dynamic nature in that they could be evaluated only as a result of their interaction with other context factors and of the interaction of these factors with crew behavior. For instance, stress levels can vary across crews: a late identification could create high stress levels during isolation for a crew with little experience in working together, but not for a more experienced one.

This classification prompts two related challenges. The first is that the variable, crew-interaction PSFs are not in the scope of most current HRA methods. Most methods do not incorporate reliability models of crew interaction and functioning (they model the crew as a second level of information processing at best), and, most importantly, do not attempt to address crew-to-crew variability. Some more recent methods, such as A Technique for Human Event Analysis (ATHEANA) and Methode d'Evaluation de la Realisacion des Missions Operateur la Sureté (MERMOS), do consider crew interactions as a "factor"; however, they do not specify an explicit way to incorporate these factors.

The second challenge is that there is little guidance in most HRA methods (and in taxonomies such as HERA) on how to determine the presence and appropriate level of constant (systematic) crew characteristics PSFs (e.g., work practices, differences in experience and cohesion). It is therefore hard to compare simulator results with predictions on such factors.

In summary, more work is needed on the treatment of variable and crew-level PSFs within the Empirical Study. In comparing simulator results with HRA predictions, these must be considered because they contribute significantly to the differences in the performances of the crews. The factors "team dynamics," "work processes" and "communication" particularly come to mind. In contrast, determining the influence of the crew-independent PSFs may have been more straightforward because these PSFs are more similar to the factors emphasized in most HRA methods.

2.9.1.3 The Interaction Between PSFs: Factor Models vs. Models of Observed Influences

For the identification of the driver factors from the crew summaries, the scenario events analyzed were evaluated against a list of PSFs. For each PSF, the presence, sign, and effect was determined, and the manifestation was described in operational terms. For instance, in one crew summary, "communication" was rated as "negative influence present" and described in the following way: "While working on the isolation, RO and ARO talk past each other, and the orders to the field operator are initially not what they intended."

This format is consistent with the modeling of performance and PSFs in many HRA methods, where the assessment problem can be formulated as follows:

$$Pf(T_i) = f(w_{i1}v(F_1), \ldots, w_{in}v(F_n), e_i) \qquad (1)$$

where $Pf(T_i)$ is the probability of failure of task T_i in a particular event sequence, F_1, ..., and F_n are PSFs that influence human performance of the given task, $v(F_1)$, ..., and $v(F_n)$ are their quantified values, w_{i1}, ..., and w_{in} are weighting coefficients representing the influence of each PSF in task T_i, and e_i is an error term representing model and data uncertainty. f represents the function that yields the probability estimate, which together with the parameters of the expression above could be called the reliability model (i.e., a model of human performance). Different HRA methods incorporate different reliability models.

For many HRA methods, the reliability model or function *f* is a linear additive function:

$$Pf(T_i) = w_{i1}v(F_1) + w_{i2}v(F_2), \ldots , + w_{in}v(F_n) + e_i \qquad (2)$$

This type of model treats the PSFs as orthogonal, direct influences on the probability of task failure.

Mapping this kind of modeling to empirical data is generally not a simple task. In the first place, the assignments and ratings cannot be strictly done "one-by-one," as the PSFs are not orthogonal. Thus, when writing crew summaries, it was not always straightforward to identify the main influencing factor(s). In addition, the categorization in terms of "presence" and "directness" does not exhaust the range of possible interactions. To meet these challenges, we deliberately did not use an orthogonal model for the representation of the PSFs in the empirical data. This enabled the comparison to evaluate whether the HRA methods identified the drivers, independent of their specific definitions. In the pilot study, the issue of double counting was not considered to be a fundamental issue in the comparison because it focused on this qualitative aspect. Furthermore, patterns and instances of factor interactions were indicated and commented on in the derivation of the drivers (Section 2.8), as well as by referring to operational details when writing the comparisons. These strategies, and other issues relating to the derivation of the driving factors, are thoroughly discussed in the next section.

2.9.1.4 Interpretation of Driving Factors

It is worth noting that the PSFs identified as main drivers are not intended to represent a model of performance. The assessment group is aware that some of the PSFs used for the identification of main drivers in some cases double-count some effects. As an example, consider the PSFs "scenario complexity" and "HMI and indications of conditions." Those methods that use "scenario complexity" as a PSF take into consideration factors such as masked plant cues and poor indications which are also considered in the "HMI and indications of conditions" PSF. The double-counting was deliberate so as to be able to match the factors as they are referred to in a broad range of methods.

Further remarks concerning PSF ratings identified as main drivers:

- It is not always straightforward to use a single rating for a PSF, and a range is sometimes preferred, such as "poor to very poor." This is because the task may have subtasks for which the PSF will be different (e.g., detection vs. situation assessment/response planning vs. execution/implementation), or because the factor's strength of impact varied among the crews.
- In a few cases, the same PSF may have a positive rating as well as a negative rating. This should not be taken to represent a huge uncertainty; for example, we have seen that procedural guidance may be very good for execution but poor for

diagnosis/decision. In such cases, we assess the overall effect of the PSF but document both sets of effects, and report the overall effect.

For example, in HFE 1B, the PSF "training and experience" was given an overall rating of "somewhat poor," but had positive influences within the performance as well. In this case, the PSF is included in the "Negative driving PSF" section as well as in the "Positive driving PSF" section. In the first case, it is shown as "Training and experience – somewhat poor," with a corresponding discussion of the negative aspects; in the "Positive Driving PSFs" section, it is shown as "Training and experience (somewhat poor)" (with the rating in parentheses indicating the overall effect), with a discussion of the positive aspects.

- PSFs may be present without having a clear link to performance. These will not appear in the 1A/1B integration under positive or negative PSFs.

3. COMPARISON OF METHODS TO DATA

3.1 Comparison Methodology

In the first phase of the pilot study, the comparison of predicted and experimental outcomes considered:

1. The main drivers of performance (driving factors).
2. How the difficulty (or ease) of performance for the tasks associated with each human failure event (HFE) would be expressed operationally.

The comparison was conducted in two phases:

- A "blind" review in which the results of the analyses by the human reliability analysis (HRA) teams were summarized by the assessment group without considering the crew performances in the simulated scenario.
- The actual comparison of individual analyses with the observed results from crews in the HAMMLAB (HAlden huMan-Machine LABoratory).

Each phase featured an iterative process in which assessment group members reviewed and summarized the individual submitted analyses, and that summary was in turn reviewed and verified by the team that completed the HRA. In this manner, the analysis process has attempted to ensure that the characterization utilized in the comparison accurately represents the intent of the HRA teams who completed the analyses according to specific HRA methods. At the same time, these summaries aimed for a uniform, less HRA-dependent representation of the predicted outcomes.

3.1.1 Summarizing the HRAs—the HRA Submittals and Predicted Outcomes

The method-specific analysis submissions typically included three types of information, Forms A and B and specific documentation on the analysis from the method. As noted earlier, Form A (see Appendix B) represents high-level summary information, with a particular emphasis on identifying the main drivers in terms of performance shaping factors (PSFs), causal factors, other influence characterizations explicitly identified through the HRA method being used, and the potential fault types that might be expected of crews. Form B (see Appendix C) provides detailed information standardized according to the Human Event Repository and Analysis (HERA) taxonomy. Finally, each analysis team provided supplemental material specific to each method. This latter material included task analytic reviews of operating procedures, analysis worksheets specific to the HRA method, and documentation of assumptions made by the HRA team.

Each submission packet was reviewed by a team of at least two assessors, each of whom had experience with the HRA method associated with the submission. All information provided by the HRA teams was reviewed independently, and a consensus reached on the main findings from the analysis submission. The HRA teams' Form A summaries served as the basis of the present comparison; however, all parts of the submission were used in the comparison.

The decision to focus primarily on Form A was motivated by the desire for a high-level initial comparison. Form A represented a straightforward way to describe the PSFs, causal factors, and other influence characterizations explicitly identified through the HRA method being used. The HRA teams were to identify factors relevant to the success and/or failure of the HFE, with particular focus on the factors that may drive the crews to fail. The discussion was to reflect the basis for the human error probability (HEP) obtained for the HFE, while staying in keeping with the "factors" or characterizations explicitly identified as important in the application of the HRA method. The terminology of the HRA method was to be used.

At its core, each HRA method attempts to capture those factors that affected performance. This was the primary basis for the present comparison of HRA methods to the data—the extent to which the HRA method accurately and completely predicted those factors that shaped the crew performances observed in the HAMMLAB simulator.

Form B presents a more detailed framework for comparison. While Form A is open-ended in the sense that HRA teams could formulate performance contributors in the method's own manner, Form B provides a predefined structure to which the HRA teams had to conform in their analyses. The completed Form B provided insights into the overall analyses, especially in the way individual teams framed their analyses in a standard language. Form B was used in some cases to map the open-ended responses in Form A to a common language for comparison. Further analyses using Form B are still in progress, and are not presented in this document.

The assessment teams completed an initial review and summary of the analysis submissions for each HRA team prior to seeing the results of the HAMMLAB simulator study. The assessment team solicited additional clarifying information as needed and provided a summary, which the analysis team commented on and corrected. Subsequently, the assessment team reviewed the main drivers and operational story, which was obtained from the simulator study and detailed in Chapter 2 of this report. This information served as the basis for the comparison between the empirical findings and the analyses. The comparison information was appended to the initial summary, and is presented in subsequent sections of this chapter. The assessment teams added a new discussion comparing the method's drivers and operational findings to the study's actual findings. Following each comparison, new sections were added to capture insights from the HRA method and the analysis. These sections covered any insights for error reduction that might be included in the analysis submission, a discussion of how the identified performance drivers impacted the computation of an HEP, and a discussion of the quality of guidance provided by the method and its traceability.

3.1.2 Criteria for the Comparison of HRA Submittals vs. Empirical Data

In short, the following sections are provided for the comparison of each method:

- A short overview of the method(s) used.
- A list of the main references for the method(s).

- A summary and assessment of the influencing factors that most affect performance for both HFE 1A and 1B, separately, including:
  - Negative influences as identified by the HRA teams (summarized by the assessment team).
  - A comparison of the negative influences on operating crew data (performed by the assessment team).
  - An assessment of the HRA team analysis of the negative influences (performed by the assessment team).
  - Positive or neutral influences as identified by the HRA teams (summarized by the assessment team).
  - A comparison of the positive influences on operating crew data (performed by the assessment team).
  - An assessment of the HRA team analysis of the positive influences (performed by the assessment team).
  - A summary of the operational description (qualitative analysis) of expected crew performance as described in the HRA team's analysis and summarized by the assessment team. The qualitative analysis was to discuss (a) the perceived difficulty or ease that the crew will have in performing the action of interest and (b) why the action should be easy or difficult, based on insights from using the HRA method (in operational or scenario-specific terms to the extent possible, i.e., operational description).
  - An assessment of the HRA team's operational description (performed by the assessment team).
- Additional comments by the assessment team on the HRA method and analysis, based on the comparison. (Note: The comments may be broken out separately for HFE 1A and HFE 1B if appropriate and an initial evaluation of the strengths and weaknesses of the method may be provided if identified.)
  - Insights for error reduction from the application of the method as applied to these scenarios (provided by the assessment team).
  - Insights on the impact of the PSF influences on the HEP; an assessment by the assessment team regarding the sensitivity of the HEP produced by the HRA method to the identified PSFs.
- Insights on guidance and traceability of the method; an assessment of the adequacy of the guidance and traceability of the method based on the comparison (performed by the assessment team).

Note that the comparison does not include HEP information for each HRA method's findings. The emphasis in this comparison is on comparing the HRA method's findings to actual crew performance data, not in comparing the quantitative data between HRA methods.

A final revision of the method comparison will follow after the comment period and workshop held in Washington, DC, October 23 to 25, 2007. It must be noted that the comparisons included herein are preliminary and have not been endorsed by the individual HRA methods developers or analysis teams.

3.2 Comparison Results

3.2.1 ASEP/THERP (NRC)

3.2.1.1 Short Overviews of the ASEP and THERP Methods

Note: As noted by their submittal, the NRC ASEP/THERP team started with the Accident Sequence Evaluation Program Human Reliability Analysis Procedure (ASEP) method, but within ASEP, where appropriate, the method allows use of Technique for Human Error Rate Prediction (THERP) to support the quantification. Thus, quantification was performed using values from ASEP in some cases, but in most cases from THERP.

Accident Sequence Evaluation Program Human Reliability Analysis Procedure (ASEP)

As described in NUREG/CR-4772, ASEP [10] is intended to be a less resource-intensive version of the THERP method described in NUREG/CR-1278 (THERP Handbook) [11], but it also extends THERP in several ways, particularly with respect to the treatment of pre-initiators. In contrast to THERP, ASEP is intended to be implemented by systems analysts who are not HRA specialists. Given the "short-cuts" in the method (compared to THERP), the quantification approach is purposely intended to provide somewhat more conservative estimates than if THERP were used directly. Like THERP, ASEP relies on a TRC for quantifying the probability of failure in the diagnosis portion of human actions, and uses a time-related PSF to address the impact of time on the response execution portion.

As a technique for estimating HEPs, ASEP addresses the quantification of both pre-accident and post-accident HFEs, and provides specific guidance for deriving both screening values and nominal values for both types of HFEs. The analyst essentially performs the quantification by first evaluating factors prescribed by the ASEP guidance and relevant to the HFE being addressed (e.g., whether a post-calibration test is supposed to be performed following the calibration of a component; the time available to perform a desired action following a plant challenge). The analyst then selects the appropriate HEP (with uncertainty bounds) based on tables and curves provided in ASEP that address a variety of these factors and combinations of factors that could influence the likelihood of the HFE. ASEP does not address HFEs that are directly associated with causing initiating events (such as a human error that results in a trip of a feedwater pump and a subsequent plant trip). Rather, it is based upon the THERP Handbook, but purposely simplifies some of the THERP guidance, such as the model for dependency. It is almost entirely self-contained; the user need not be familiar with the THERP Handbook [11] and is not required to use any of the THERP models or data.

Note that ASEP does not address most activities related to the HRA process (such as the identification of HFEs), and does not provide detailed guidance on how to model the HFEs. Thus, in using ASEP, it is assumed the HFEs have already been identified and modeled and only the quantification of the associated HEPs is required.

ASEP's ease-of-use and compatibility with the standard probabilistic risk assessment (PRA) framework are practical strengths. As the only moderately detailed systematic process for estimating pre-initiator HEPs, ASEP has been used extensively for this purpose. Similarly, its approach for estimating post-initiator HEPs has been widely used, although numerous other methods have gained favor in dealing with post-initiator HFEs. On the downside, ASEP only produces HEPs, does not aid in identifying HFEs, and is not always helpful in identifying the causes of errors. Based on its simplified approach with its treatment of only a subset of all possible PSFs that could affect the human actions of interest, ASEP's results are probably best categorized as providing HEPs that are likely to be conservative as long as the factors treated in the method are the most appropriate ones for addressing the HFE of concern. If, however, other PSFs are particularly relevant and could affect the HEPs, ASEP's estimates could then be inappropriate. When recognized, this may lead to the need for compensatory steps, such as use of other methods to check and modify the obtained values, or use of expert judgment to make appropriate adjustments.

Technique for Human Error Rate Prediction (THERP)

As described in "The Handbook for Human Reliability Analysis With Emphasis on Nuclear Power Plant Applications" (NUREG/CR-1278, THERP Handbook, [11]), THERP is a method for identifying, modeling, and quantifying HFEs in a PRA. At some 700 pages, the THERP Handbook [11] provides a comprehensive source of human reliability knowledge in the context of nuclear power plant (NPP) safety. With respect to modeling, THERP does not provide explicit guidance on how to model an HFE in a PRA. Nonetheless, its qualitative guidance can be useful in doing so. THERP provides some guidance on decomposition of non-diagnosis HFEs into lower-level errors, and identifies important PSFs through task analysis (one principal feature of a THERP analysis). This decomposition is graphically represented with HRA event trees. THERP focuses primarily on rule-based behavior, in which operators follow procedures; however, THERP also treats diagnosis HFEs via a time/reliability correlation (TRC). With respect to quantification, THERP contains a database of nominal HEPs. The analyst adjusts these nominal HEPs (which include uncertainty bounds) upward or downward to reflect plant-specific PSFs, resulting in basic HEPs. Finally, dependence among tasks is accounted for, producing conditional HEPs, and recovery factors are applied, producing a joint HEP.

There are two important companion volumes to NUREG/CR-1278. The first is NUREG/CR-2254 [12], which illustrates the application of THERP to HRA at a nuclear power plant. The second, referred to as ASEP and published in NUREG/CR-4772 [10], is primarily devoted to describing a simplified, less resource-intensive version of THERP. (ASEP is summarized above). However, in addition to describing ASEP, NUREG/CR-4772 [10] extends THERP by providing guidance for post-initiator HFEs beyond that provided in NUREG/CR-1278 [11]. Specifically, this guidance addresses the use of the then-new symptom-oriented emergency operating procedures (EOPs), and addresses emergency actions that have been memorized.

3.2.1.2 Main References to the ASEP and THERP Methods

[10] Swain, A.D.: "Accident Sequence Evaluation Program Human Reliability Analysis Procedure," NUREG/CR-4772/SAND86-1996, Sandia National Laboratories for the U.S. Nuclear Regulatory Commission, Washington, D.C., February 1987.

[11] Swain, A.D. and H.E. Guttmann: "Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications," NUREG/CR-1278/SAND80-0200, Sandia National Laboratories for the U.S. Nuclear Regulatory Commission, Washington, D.C., August 1983.

[12] Bell, J. and A. Swain: "A Procedure for Conducting a Human Reliability Analysis for Nuclear Power Plants," NUREG/CR-2254, U.S. Nuclear Regulatory Commission, Washington, D.C., May 1983.

3.2.1.3 NRC, ASEP/THERP Analysis of HFE 1A

Summary of Negative Influences Identified by NRC ASEP/THERP

For the Cognitive Part (Diagnosis):

Time available is the main driver of this HEP in NRC ASEP/THERP, and the team used ASEP Table 8-2 [10] to estimate the failure probability; they chose the lower bound because, although they thought that time would be the main driver of success, they examined how different crews would be trained to trip the reactor in different plants. They point out that in some plants the crews are instructed to take preemptive actions (in which case time would not be an issue) when they have sufficient confidence that there is indeed an SGTR, while in other plants the crews may take preemptive actions only when they are certain the (safety) system (SI) automatic actuation is imminent. They noted that still other plants have a hands-off approach and wait for the safety system to activate automatically. The team notes that all of these options would allow the crews to achieve reactor trip, but only preemptively tripping the reactor would allow success given the definition of success in this exercise, that is, SG isolation within 20 minutes. Therefore, in a way, they examined "crew variability from plant-to-plant in their analysis." Since the information package provided by Halden did not have information about crew training in this regard, the team made the assessment based on certain U.S. plants' operational experience, meaning that the crews would generally perform preemptive actions. Based on this fact, along with an assumption of good procedures, they decided it was appropriate to use the lower bound value from Table 8-2 [10]. Thus, time available was the main limiting factor, and its impact factored in through the ASEP TRC.

For the Execution Part:

The team assumed various types of failures, such as failure to read, to locate dials and to follow a procedural step. They also assumed step-by-step non-dynamic tasks (i.e., no complexity), low stress, and good training because SGTR is a design-based scenario.

Comparison to Empirical Data

Success for this action is defined as "completion within the time defined." It is therefore natural for NRC ASEP/THERP to first examine the time available to diagnose vs. time available to perform. Based on the information in the "information package," and based on an estimate of the time it would take to successfully isolate the ruptured steam generator once they had entered E-0 for plant trip, they estimated that there are seven minutes in which to decide to trip the reactor. Thus, they correctly use the method to estimate that diagnosis/decision task and time is the main limiting factor for this HFE. The choices the ASEP/THERP analysis made in terms of PSF levels for the diagnosis curve and for response execution were all positive. ASEP includes guidance for choosing values in Table 8-2 [10], and the team chose the lower bound, assuming that procedures and training are good. They examined procedural guidance and training, which are underlying considerations for choosing a value. Therefore, it seems that the team correctly identified a low probability event, essentially assuming no negative factors for the diagnosis besides time limitations. This assumption was supported by the results.

The Halden data suggests that execution complexity is a potential negative PSF for the execution task rather than the diagnosis: that is, a few crews had problems implementing E-3 due to "some simulator/home plant differences and to the complex build-up of this procedure step (eight points mixing local and control room actions). The level of execution complexity (based on assessment of the crew data) was not identified by the NRC ASEP/THERP analysis.

Assessment of Negative Influences Identified by NRC ASEP/THERP

The ASEP NRC team correctly determined that there are essentially no negative factors with respect to diagnosis, but they also addressed the potential effects of time limitation (i.e., the potential delay in tripping the reactor so that they wouldn't be able to complete the isolation within 20 minutes) in applying ASEP. They did not detect the potential problems with executing E-3.

Summary of Positive and Neutral Influences as Identified by NRC ASEP/THERP

In the context of ASEP, step-by-step non-dynamic tasks, low stress, and good training were assumed for response execution because "SGTR is a design-based scenario." The ASEP NRC team also chose the lower bound values from the diagnosis curve (Table 8-2 [10]), which took the position that procedures and training are good.

Comparison to Empirical Data

It seems that there was adequate time for all but one crew, although other crews certainly approached the time limit before responding. The NRC ASEP/THERP analysts included the potential impact of time limitations through use of the ASEP TRC. It is noted that the ASEP team correctly estimates time available for recognizing/deciding the need for the action. They credited most of the identified positive influences on the diagnosis, but failed to explicitly state that the indications of the event were identified as very good.

It should be noted that there are two types of time considerations in HRA: (a) "time" as a boundary condition which is estimated by HRA analysts to evaluate whether the crew will have enough time to diagnose and perform an action, and (b) "time" as a PSF which has to do with "the crews' awareness of time available when they perform an action," which can influence their performance positively or negatively.

It did not appear that the ASEP team examined whether the crews felt time pressure to accomplish HFE 1A, given the indications of temperature, pressure, etc.

Assessment of Positive and Neutral Influences Identified by NRC ASEP/THERP

Given the definition of this human action, it may be a reasonable assumption for the ASEP team not to identify time as a positive PSF. TRCs obviously factor time limits into deriving HEPs and the team credited procedures and training so as to obtain as low an HEP as possible, given the time limits. Thus, they treated time as positively as possible in the context of ASEP. The team correctly identified low stress and good training as positive PSFs, and took credit for good training and procedures by using the lower bound of the diagnosis curve. They do not explicitly talk about human-machine interface (HMI (very good indications)), but from the write-up it is implied that it is not negative.

Summary of Operational Description Provided in the NRI DT+ASEP Analysis

The main difficulty the crew will experience is in diagnosing the need to trip the reactor in time. If the crew waits for the reactor to automatically trip, they will not have enough time to successfully go through the E-0 and E-3 procedures and isolate the ruptured steam generator. Although they were only given seven minutes to realize that the reactor needs to be manually tripped, it was felt that the alarms and readings presented to the crew were sufficient to allow them to trip the reactor in time, so the HEP calculation remained small. Crews assumed to have had normal training would be expected to recognize the secondary side radiation alarm, changes in steam generator (SG) water levels, and reactor coolant system (RCS) pressure, and to identify the scenario as a steam generator tube rupture (SGTR) event. Given the training required of operator crews, the teams felt that the crew should easily move through the procedures necessary for dealing with this HFE: in other words, that the training would generally be good.

Assessment of Operational Description

The NRC ASEP/THERP qualitative assessment is generally consistent with what occurred in the scenario, and their assessment of possible delays seemed reasonable.

3.2.1.4 NRC, ASEP/THERP Analysis of HFE 1B

Summary of Negative Influences Identified by NRC ASEP/THERP

Note that the summary of negative influences was supported by information from Form B:

For the Cognitive Part

- In the situation assessment/interpretation: In E-0 step 21c, the crew could misinterpret the RCS trend in trying to decide to go to ES-1.1. The timing mismatch between procedural instruction and plant state is beyond ASEP's scope (see the ASEP qualitative assessment [operational description] below for more discussion on this); therefore, it was assumed at this step that there is a 50/50 chance of saying "yes" or "no" at step 21c. However, whether the crew answers "yes" or "no" at this point, they still have opportunities in the procedures to reach the appropriate conclusion in terms of deciding whether to go to E-3.

For the execution part

Note that in the ASEP/THERP analysis, HRA event trees were used, mixing diagnosis and execution. The NRC ASEP/THERP analysis essentially quantified the crew correctly, following steps in the procedures.

- Procedures: The design of the emergency operating procedure (EOP) does not adequately cover this type of scenario with multiple faults.
- Mismatch between the speed of the scenario and the pace of the operator following procedures could affect the outcome.
- Extreme or unusual conditions: Two initiating events plus key instrumentation failures.
- Complexity: The scenario could be out of the EOP scope.
- Experience and Training: The scenario relies on the crew's interpretation of plant state in order to follow correct procedure path. Operator training is important, since for this scenario the procedure instruction requires the operators to interpret the plant state (e.g., referring to RCS pressure trend at E-3, step 21c). The information provided is insufficient to judge whether an operator's experience/training is a positive or negative factor.
- Somewhat high stress level.
- Information fails to point directly to the problem: Failure of the secondary side radiation indication eliminates the most obvious symptom to transferring to E-3 from E-0.

Comparison to Empirical Data

The Halden summaries identified scenario complexity, indications of plant conditions, procedures, training, adequacy of time, execution complexity, and work processes as factors bearing negatively on performance. Although the terminology used in the ASEP/THERP analysis is different than that of the Halden PSF definitions and the NRC ASEP/THERP analysis appeared to address most of these factors as affecting "execution" of the procedures, the NRC

ASEP/THERP team correctly identified most of the PSFs observed in the empirical data: that is, their discussion addressed these factors. (Note, however, that Form B was used by the assessors to help with this interpretation).

Regarding available time, the ASEP team did not seem to see or use available time as a constraint on performance in this scenario, except in terms of recovery.

Assessment of Negative Influences Identified by NRC ASEP/THERP

In the Halden summary of driving PSFs for operating crews, the inadequacy of the procedures and training for the complex scenario, in conjunction with the missing radiation indications and several other factors, were determined to be very important. Based on the assessors' review of their analysis and with support from Form B, it appeared that the NRC ASEP/THERP team identified all of these factors as important. However, although all of these negative PSFs could be identified in Form B, the ASEP/THERP analysis did not appear to take all of them into consideration when deriving HEPs. HFE 1B was analyzed using THERP guidance for executing an action/procedure, and lower-bound values were frequently chosen based on "expert opinion" in estimating HEPs for the particular tasks of their task analysis. For some of the individual contributions to failure (e.g., by transfer to ES 1.1), the failure probability is the same for this complex case and for the simple case (HFE 1A).

The ASEP analysis predicted that on average the crews would fail 2% of the time; 7 out of 14 crews failed to meet the time criterion of 25 minutes, even though all but one finished within 32 minutes. Thus, although the NRC ASEP team correctly identified the most important negative PSFs, the estimation of their effects seems to be very optimistic with respect to the observations of crew performance. In fact, it did not appear that the ASEP/THERP analysis directly incorporated the effects of these factors. It seems that the quantitative results are driven by the evaluation of the positive PSFs, and especially by the assumption for recovery. In other words, even though negative influences were discussed, as discussed below, the drivers were generally assumed to be positive.

Summary of Positive and Neutral Influences Identified by NRC ASEP/THERP

Clear indication to enter EOPs; good procedure quality (i.e., clear step-by-step instruction for the crew to follow); training on the use of emergency procedures, which implies that the crews will progress through the procedures with little difficulty; and recovery (before transferring between emergency procedures) were thought to be positive influences in the ASEP/THERP analysis. It is assumed that the operating crews would meet to discuss the decision and the evidence present before initiating this transfer. The team meeting may stop the crew from incorrectly transferring between procedures, and allow them to recover from an incorrect decision (as the crew reviews the evidence and realizes any inadvertent mistakes made by a crew member).

Comparison to Empirical Data

The positive PSFs from the Halden data include indications of plant conditions, training, adequacy of time, and procedural guidance. The NRC ASEP team identified training on the use of the emergency procedures, which implies that the crews will progress through the procedures with little difficulty; thus, the training would be a positive influence. As noted in the discussion of negative influences, however, the procedures were not very effective, and the training did not seem to specifically help the crews with implementing the procedures. Nevertheless, their overall training (deemed a secondary positive influence in the data, since, more generally, the specific training on this scenario was poor) did seem to help them make the knowledge-based decision that eventually took most of the crews to E-3.

The ASEP/THERP analysis also refers to clear indications to enter the EOPs as positive factors, although these are not the available indications referred to in the Halden analysis. The Halden results referred to the fact that the crews had indications like steam generator level to support their diagnosis of plant conditions (secondary positive influence); thus, it does not appear that the ASEP analysis was identifying the same positive factors, so its predictions were not generally supported by the data. However, the NRC ASEP/THERP analysis also identified "recovery" as a potentially important factor, if an error did occur.

Assessment of Positive and Neutral Influences Identified by NRC ASEP/THERP

The ASEP analysis did not seem to do well in identifying the nature of the important positive factors (secondary positive influences); the positive factors they identified were not supported by the data, and were in fact generally identified as negative influences in the empirical data. However, their analysis of negative factors did touch on the aspects identified as negative, though it appears that they weighted what they saw as positive more heavily than what they saw as negative, which was not supported by the data.

Summary of Operational Description Provided in the NRC ASEP/THERP Analysis

This is a rare event that has two initiating events (main steamline break (MSLB) and SGTR), with failure of some key plant state indications (i.e., secondary side radiation). The crew's success or failure in correctly identifying and isolating the ruptured steam generator is largely dependent on the time it takes them to reach certain steps within the procedures. For instance, in E-0 step 21c, the procedure directs the crew to determine the trend of RCS pressure, which in this scenario is highly dependent on the time at which it is checked. The crew may determine that the pressure is decreasing (reference 9 minute complex scenario screen), or they may determine that it is stable (reference 12 minute complex scenario screen). The timing mismatch between procedural instruction and plant state is beyond ASEP's scope; therefore, the analysts made an assumption at this step, and stated that there is a 50/50 chance of saying "yes" or "no" at step 21c. However, this had little influence on the HEP, since either choice includes guidance to enter E-3. Other than this, the procedure instructions should direct the crew to the right procedure (E-3) with little difficulty. The analysts felt that, given the crew's training in the emergency procedures, they should experience little difficulty in carrying them out and successfully completing this HFE.

The most significant aspects of the HFE are the number of ways for the operators to get to E-3, and the indications needed (except for the secondary side radiation) to get them there.

Assessment of Operational Description

The NRC ASEP/THERP qualitative assessment is generally not consistent with Halden's operational summary. Although some crews did have problems with deciding whether or not to enter ES-1.1 due to the RCS pressure trend (this was a relatively difficult decision), almost all crews had trouble using the procedures (E-0 or ES-1.1 to get to E-3). Most crews diagnosed the SGTR based on the diverging SG levels, and made a knowledge-based judgment to go to E-3. They also discussed difficulties with the procedures, and some of the associated delays that the operating crews experienced.

3.2.1.5 Additional Comments on NRC ASEP/THERP Analysis

Comments on NRC ASEP/THERP Analysis of HFE 1A and HFE 1B

In Form B, the NRC ASEP analysis mostly identified the correct PSFs. However, the analysis identifies contributors to the HEP from failure mechanisms that were not observed.

They state that the HEP comprises the following HFEs:

1. Recognizing the need to go to E-0 (0.0005).
2. In E-0.3, failure to verify that the 6kV busses are energized (0.001).
3. In E-0.18, failure to read the SG pressure correctly and recover from transferring to E2 (0.001).
4. In E-0.20, incorrect transfer to E-1 (5.6E-03).
5. In E-0.21 failure to determine the trend of RCS and incorrectly transfer to ES1.1 before going to E3 (0.0006).
6. In E-0.24 failure to check SG levels (5E-03).
7. Execution failures in E-3.

While the analysts correctly identified the potential problems with E-0 step 21, this particular cause of failure plays a very minor role in the determination of the HEP. The method does not address the time delay in reaching a diagnosis, which would be addressed by the TRC portion of ASEP.

The analysis here appears to be an example of a reasonable qualitative analysis (with interpretation support from Form B) that is not reflected in the numeric results. In general, it could be argued that the connection between an ASEP/THERP quantification analysis and the analysis of driving PSFs and operational descriptions performed for this study was not a good match; however, a failure of the method to match up with the study's qualitative analysis does not necessarily imply that the HEPs obtained using THERP/ASEP will always be wrong. The low probability predicted for the base case was reasonable, even though the HEP for the complex scenario was apparently optimistic.

Insights for Error Reduction

It is not clear that the ASEP/THERP analysis will produce good insights for error reduction. The type of factors it explicitly addresses is more related to execution errors (e.g., slips), and only a couple of PSFs are considered when the diagnosis curves are used. Such insights might be obtained if a good analysis of the context and what the crews will experience is performed independently of the ASEP/THERP analysis.

Impact on HEP (Sensitivity to Driving Factors)

With this application, the HEP did not appear to be very sensitive to the factors identified in the analysis and represented in Form B.

Guidance and Traceability

Derivation of the HEP results in ASEP/THERP is theoretically traceable, if the method is used as described and reasons for choices are provided. However, attempting to include the kinds of factors the NRC ASEP/THERP team identified in Form B would take a special effort by the analysts to discuss how this was accomplished, and additional guidance would be needed to address such factors in ASEP/THERP. Furthermore, it is not clear that the guidance is adequate to address all the types of scenarios and conditions that will need to be addressed in PRA.

3.2.1.6 NRC ASEP/THERP Team Comments on the Original Comparison

No comments were provided.

3.2.2 CBDTM + THERP (EPRI)

3.2.2.1 Short Overview of the CBDT Method

Cause-Based Decision Tree (CBDT) Method

As documented in EPRI TR-100259 [13], the CBDT method is primarily intended for use in quantifying post-initiator human actions (e.g., actions determined by control room crews associated with emergency and abnormal operating procedures) that have been included in the logic models for an NPP PRA.

The CBDT method was originally intended as a supplement to the Human Cognitive Reliability/Operator Reliability Experiments (HCR/ORE) method (which is also documented in EPRI TR-100259 [13]), to serve as a check where the HCR/ORE approach produces very low probability values. Since HCR/ORE relies on a TRC approach, the CBDT method was at least initially intended to address actions with longer time frames, where "extrapolation using the lognormal curve (from the HCR/ORE TRC) could be extremely optimistic." For the longer time frame actions, it is assumed that other types of influences may become important and may not be adequately covered with the HCR/ORE TRC approach. In addition, the CBDT method is recommended in Electric Power Research Institute (EPRI) TR-100259, when the use of the

HCR/ ORE method may yield "very conservative human error probabilities." Thus, in its current form, the CBDT method's basic approach to quantification is time-independent, although time is considered in addressing the potential for self-recovery of an error or recovery by another crew member.

In more recent years, the CBDT method has come to be used as a "standalone" method, at least for quantifying HFEs with adequate time available. The method is described as an analytical approach, as opposed to the empirical approach represented by the use of the HCR/ORE TRC, and uses a series of decision trees to allow the analyst to consider a number of factors that could affect the reliability of crew response, including the quality of training, procedures, the man-machine interface, and so forth. The emphasis of the method on evaluating a relatively large set of causal factors that could influence the likelihood of success/failure of an action was a significant step in the improvement of HRA methods, which has led to its use as a primary method for quantifying post-initiator actions (e.g., in EPRI HRA Calculator® [14]).

The CBDT approach involves "the identification of situation-specific error-conducive factors"; thus, it focuses on potential failure mechanisms and their causes, evaluating the impact of a set of situational characteristics or factors in specific scenarios. It uses eight decision trees that estimate HEP values based on an assessment of the following eight general failure mechanisms and factors that could contribute to those failure mechanisms, which are related to the plant information-operator interface and the operator-procedure interface:

1. Relevant data/indications not available due to location, accuracy, reliability, or training related to their use.
2. Data not attended to due to workload, monitoring requirements, location, and inadequate alarms.
3. Data errors (data misread or miscommunicated) due to location on panel, quality of display, or nature of interpersonal communications.
4. Data misleading because cues do not match procedures, cue recognition training is inadequate, and so forth.
5. Steps in procedures missed as a result of procedure format (visibility and salience of instructions, use of concurrent procedures, use of place-keeping aids).
6. Misinterpretation of instructions as a result of a lack of instructional clarity (standardized vocabulary, completeness of information, training).
7. Error in interpreting logic as a result of instructional complexity (e.g., use of "not" statements, complex use of "and" and "or" terms).
8. Potential for deliberate violations as a result of various aspects, such as belief in instructional inadequacy or availability and consequences of alternatives.

A non-response probability (Pc) is calculated using the CBDT decision trees. In doing so, it is assumed that the effects of the various PSFs represented in the trees are independent, so the HEPs obtained from the various trees are summed together to obtain the initial probability for Pc. A recovery analysis, based on "revisitation" by either the individual performing the task or by another individual, is then performed. Time is a critical parameter in this case, and, with enough time, recovery is likely. The resulting failure probability for Pc is combined with the value obtained for failure in executing the response (Pe) to obtain the final HEP. It is noted in TR-100259 that the HEPs included in the CBDT decision trees for Pc were adapted from values given in THERP (NUREG-1278 [11]). An attachment to TR-100259 provides a brief discussion of the origin of the values and the assumptions used in modifying them for use in the decision trees.

The approach for estimating Pe is the same regardless of whether Pc is obtained with the HCR/ORE TRC or the CBDT method (EPRI TR-100259). Pe is essentially the probability of a manipulative slip (i.e., an unintended or inadvertent action, such as turning an incorrect switch or skipping a step in a procedure). It appears that only control room actions are addressed (i.e., guidance and data for quantifying local actions is not provided).

Although the CBDT method described in TR-100259 is primarily a post-initiator quantification process, it can also be seen as part of a "suite" of EPRI methods that generally try to cover the range of tasks associated with performing an HRA. In particular, SHARP1 [15] is cited as a general HRA framework that should be used in conjunction with CBDT (and HCR/ORE) to support accomplishment of various other aspects associated with performing an HRA in the context of a PRA (e.g., identification and definition of human actions). Furthermore, the CBDT method, along with the HCR/ORE approach, has been included in EPRI's recently developed "EPRI HRA Calculator®" as the primary methods for post-initiator quantification.

3.2.2.2 Main References to the CBDTM + THERP Method

[13] Parry, G. et al.: "An Approach to the Analysis of Operator Actions in PRA," EPRI TR-100259, Electric Power Research Institute, Palo Alto, CA, 1992.

[14] Julius, J., J. Grobbelaar, D. Spiegel, and F. Rahn: "The EPRI HRA Calculator® User's Manual, Version 3.0, Product ID #1008238, Electric Power Research Institute, Palo Alto, CA, May 2005.

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

3.2.2.3 EPRI, CBDTM + THERP Analysis of HFE 1A

<u>Summary of Negative Influences Identified by CBDTM + THERP</u>

For the Cognitive Part:

The method is based on identifying failure mechanisms rather than PSFs, so negative influences have to be inferred from the choices made on the branches of the decision trees; however, the branches of the decision trees reflect factors or conditions (e.g., HMI characteristics) that could lead to the failure mechanism, and are therefore similar to PSFs in most cases. The principal failure mechanism identified was missing E-0 step 19 because it is not distinct from other steps, so there is a potential for missing the step, though some recovery is given in light of the fact that step 25 would get the crew to the same place. However, no specific negative PSFs were identified.

For the Execution Part:

No specific negative influences were identified, with the possible exception of time. No credit is taken for recovery since the time available to perform the actions is short, based on the assumption that the reactor manual trip is performed after about six minutes. The evaluation did take into account there number of steps to be performed. The ex-control room actions were not considered critical to completion since they were verification steps, and failure to perform them did not contribute to the HEP evaluation. The applicable PSFs for the THERP evaluation are classified as optimal.

Comparison to Empirical Data

Only one crew failed to complete the action in the designated time frame, and they were only about a minute and half late. The main factor in the crews' completion time appeared to be the rate at which they progressed through the procedure, but it was also affected by the time at which they tripped the reactor and entered E-0. The crew that failed performed slowly, but otherwise had many elements of good performance: they were cautious and held meetings, but lacked the sense of urgency that might have helped them to finish in time and avoid the danger of overfilling the steam generator. Thus, aspects of crew-to-crew variability, which could fall into the category of work processes or some other crew-related category, would seem to be a potential contributor to crew performance. Regardless of their work processes or characteristics, however, all but one of the crews met the criterion for HFE 1A. Thus, work processes were not identified as a negative driving factor in the data for HFE 1A. In any case, the HRA methods—and the CBDT method is no exception—were not expected to be able to address these effects in this study. In any case, the HRA analysts recognized that there would be a delay in entering E-0, and accounted for this by not allowing self-recovery for failures during execution.

The execution complexity was empirically judged to be somewhat high due to the long list of actions, and included a mixture of control room and local actions, which required coordination and prioritization in order to be efficiently performed. The analysis accounted for the number of actions to be taken, so to this extent the execution complexity was reflected in the evaluation of the HEP. The PSFs used to evaluate the failure probabilities associated with the individual execution steps were considered nominal.

Assessment of Negative Influences Identified by CBDTM + THERP

The analysis addressed the complexity of the response somewhat, by virtue of taking into account the multiple steps. The primary cognitive failure mechanism identified was missing a step in the procedure. This was not observed as an issue in the simulator trials.

Summary of Positive or Neutral Influences Identified by CBDTM + THERP

For the Cognitive Part:

As evinced by the paths chosen through the decision trees, the analysts concluded that the information required to make the diagnosis is available, with all relevant indications available in control room; workload is low; it was assumed that there are no issues with communication; no indications are misleading; and the procedures are direct and easily interpreted.

For the Execution Part:

The task type was considered to be standard (step by step), and the stress was low.

Comparison to Empirical Data

Training and experience, procedural guidance, adequacy of time, and HMI and indication of conditions were all identified as being good or very good in the empirical data. The CBDT analysis identified information availability, procedural guidance and training as positive influences in the analysis. These results are consistent with the Halden observations. In addition, the CBDT analysis essentially assumed that adequate time would be available, which is also consistent with the results.

Assessment of Positive Factors and Neutral Influences Identified by CBTM + THERP

Although covered in somewhat different terminology, the CBDT analysis identified the main positive driving factors. The identification of information availability, procedural guidance, and training as positive influences in the analysis is consistent with the Halden observations.

Summary of Operational Description Based on Information Provided in the CBDTM + THERP Analysis

The analysts felt that there were clear initial cues to indicate an SGTR, and the operators would trip the reactor after some delay (in this case a rather precise 6 minutes and 33 seconds, based on the sample Halden print-out). The likelihood of not entering E-0 is considered negligible. The cognitive failure modeled is that of failing to transition into E-3 at E-0 step 19, with the failure mechanism being omission of the step in E-0. Some recovery is given, due to the fact that they would get a second chance at E-0 step 25. Consistent with the use of THERP for the execution contribution, the failures are due to errors of omission (EOOs) and errors of commission (EOCs) in following the specific steps of E-3. E-3 step 4b has been treated as a recovery of E-0 step 12, which allows auxiliary feedwater (AFW) to be stopped feeding the ruptured generator.

Assessment of Operational Description

Since the failure probabilities of the identified failure mechanisms are small, it is not surprising that there is no evidence as to whether they are or are not the most likely mechanisms. However, the general assessment that success is highly likely is consistent with the observations.

3.2.2.4 EPRI, CBDTM + THERP Analysis and Comparison, HFE 1B

<u>*Summary of Negative Influences Identified by CBDTM + THERP*</u>

For the Cognitive Part:

The method is based on identifying failure mechanisms rather than PSFs, so negative influences have to be inferred from the choices made on the branches of the decision trees. Despite this, the branches of the decision trees reflect factors or conditions (e.g., HMI characteristics) that could lead to the failure mechanism, and are therefore similar to PSFs in most cases. Besides the complication caused by the lack of radiation signal, no specific negative influences were explicitly identified. Since E-0 step 19 is not credited for cognition (because the radiation monitors would not be effective due to early steamline isolation), the analysis assumes that the recognition of the SGTR is performed using step 24. E-0 step 24 is not distinct from other steps, so there is a potential for missing it. No recovery is given for step 25, as its cues (radiation monitors) would not be effective due to early steamline isolation.

There is a minor contribution from increased failure of attention as compared with HFE 1A. This appears to be because the decision is made in E-0 step 24 rather than step 19, and it is regarded as a monitoring activity without an alarm. It is assumed that there is no time to recover if an initial diagnosis is not made.

For the Execution Part:

None. The PSFs are classified as optimal, though no credit is taken for recovery since the time available to perform the actions is short, assuming that it will take about 18 to 20 minutes to enter E-3, and given the masking effects of this scenario.

<u>*Comparison to Empirical Data*</u>

The implementation of the CBDT recognized the lack of a radiation signal, and was based on the premise that the operators would focus on differences in steam generator levels as the indication of an SGTR, which was consistent with data. The analysis focused on the availability and clarity of steam generator level indication and the procedural direction to use that indication (step 24). The analyst followed the path through the procedures with the available information. The qualitative discussion recognizes that there could be an "additional ~10 minutes to actually enter E-3...given the masking effects of this scenario," and thus recognizes the complexity introduced by the scenario. The precise cause for the delay (e.g., problems with procedures) was not identified in the analysis. The analysts accounted for this delay by recognizing that there would not be enough time to go back through E-0 as a recovery from an initial failure to recognize the SGTR, and this was true, to some extent; however, the delays were manifested in other ways by several crews. The negative influence from a lack of specific training on this complex scenario, and the complications this created for the transfer to E-3 (i.e., procedures were not as helpful as they should have been), was not identified by the analysis team.

The execution complexity was empirically judged to be somewhat high because of the long list of actions, which included a mixture of control room and local actions and thus required coordination and prioritization in order to be efficiently performed. The analysis did take into account the number of actions to be taken, so to this extent the execution complexity was reflected in the evaluation of the HEP. The PSFs used to evaluate the failure probabilities associated with the individual execution steps were considered nominal.

Assessment of Negative Influences Identified by CBDTM + THERP

The analysis identified the complication caused by the lack of radiation signal, and recognized that this could cause a delay in diagnosis. To account for this delay, the analysts did not allow recovery by recycling through E-0. The negative influence of insufficient training on the specific scenario was not identified. Although this question is asked on decision tree (d) for the failure mechanism "information misleading," the decision that "all cues were as stated" bypassed this question.

The analysis addressed the complexity of the response somewhat, by taking into account the multiple steps.

Summary of Positive or Neutral Influences Identified by CBDTM + THERP

For the Cognitive Part:

Based on the paths chosen through the decision trees, the following assumptions can be made: information required to make the diagnosis is available, and all relevant indications are available in the control room; workload is low; there are no issues with communication; no indications are misleading; and the procedures are direct and easily interpreted.

For the Execution Part:

The type of task was assessed as standard (step by step), and the stress is low.

Comparison to Empirical Data

Although it was not mentioned in the summary of the observations (but was identified as a secondary positive influence), the necessary information (e.g., SG level, AFW flows, pressurizer (PRZ) level, etc) was all available, and led to a correct diagnosis. The analysis assumed a low workload, which could be equated to low stress, as observed at the simulator. In contrast with the assumptions made in the analysis, however, some crews appeared to have communication problems. The questions about training were bypassed in the decision trees, though the qualitative discussion assumed that training on the procedures would be conducted at least once biannually.

Assessment of Positive Factors or Neutral Influences Identified by CBDTM + THERP

The availability of the information needed to identify the SGTR was recognized; the training and procedural directions were inappropriately identified as positive influences, although the general SGTR training was identified as a secondary positive influence in the crew performance evaluations, which was recognized to some extent in the analysis.

Summary of Operational Description Based on Information Provided in the CBDTM + THERP Analysis

The HFE is modeled on the assumption that steam generator level will be used to detect the occurrence of an SGTR, as the radiation indications are not available. The likelihood of not entering E-0 is considered negligible. The cognitive failure modeled is that of failing to transition into E-3 at E-0 step 24, with the major failure mechanism being the omission of the step in E-0, with some contribution from failure of attention due to the fact that recognition of the symptoms (uncontrolled level increase in one SG) is considered to be a monitoring activity with no backup alarm. No recovery is credited, possibly because there is no time to iterate E-0 from the beginning. Consistent with the use of THERP for the execution contribution, the failures are due to EOOs and EOCs in following the specific steps of E-3.

Assessment of Operational Description

The assumption that the transition would be made at E-0 step 24 was not borne out in practice, with only two crews transitioning directly from this step; however, steam generator level was a critical factor in influencing most crews' decisions, though they got there by a number of means, including recycling through E-0. Different crews used different approaches to end up in E-3, and, since this method is not designed to address this feature, this aspect is not addressed in the operational description, which is more in the nature of the "expected" response.

3.2.2.5 Additional Comments on CBDTM + THERP Analysis

The method is based on the identification of failure mechanisms rather than PSFs, and, as with many HRA methods, it is based on the "average" scenario and conditions. Crew-to-crew variability has a significant impact on the empirical data. Furthermore, both HFE 1A and HFE 1B are time-critical; for HFE 1A, this is because the time of entry into E-0 varies from crew to crew, while for HFE 1B, the time it takes to fill the SG as measured by the wide and narrow range level monitors is short compared to the time taken to identify the path through the procedures. In the EPRI suite of methods, time-critical HFEs would be addressed by the HCR/ORE correlation. However, the decision had been made to employ the CBDT for this analysis.

Insights for Error Reduction

In this specific case, it is not clear that the method will provide significant insights for error reduction in HFE 1A. The specific error mechanisms identified are related to missing procedural steps.

Because of the specific choices made in the decision trees, some opportunities for error reduction, such as the addition of level differences in step 19 as an indicator to transfer to E-3, would not have been identified.

Impact on HEP (Sensitivity to Driving Factors)

For this model, the driving factors are specific failure mechanisms, such as "missing a step in a procedure." The probabilities assigned to these drive the results.

See the short overview of the method in Section 3.2.2.1.

Guidance and Traceability

The evaluation method is explicit in the choice of paths through the decision trees and in the identification of the specific errors in execution. The method results are traceable as long as there is good documentation as to why the particular branches on the decision trees are chosen.

The results of this comparison should provide some indications of additional clarification and guidance that would enable the use of the CBDT method to address the complex scenario to be improved. For example, one way to address the issue of the availability of indications in the CBDT is to make sure that a clear distinction is made between cues that can be identified as primary (in this case the radiation signals) and those that are secondary (SG level), and use decision tree (a) to address the primary cue rather than the cue that was eventually used. In this case, the primary cue indicator is available in the control room, but the indication is inaccurate because it has been isolated. This was the key factor in this scenario.

3.2.2.6 CBDTM + THERP Team Comments on the Original Comparison

The typical EPRI approach is to first evaluate the human error probability using the CBDT method for the cognitive portion and THERP for the execution portion. For many human failure events, such as HFE 1A, the time window is long relative to the amount of time required to complete the action, which is sufficient. For some human failure events, such as HFE 1B, the amount of time required to complete the action is nearly as long as the time window available (in this case due to delays). When this happens, the EPRI approach is to complement the CBDT method with the HCR/ORE method. For this pilot evaluation of HFE 1B, only the CBDT was applied. In future evaluations, the overall EPRI approach of CBDTM, supplemented with HCR/ORE, will be applied.

3.2.3 SPAR-H (INL)

3.2.3.1 Short Overview of the SPAR-H Method

The Standardized Plant Analysis Risk-Human Reliability Analysis (SPAR-H) method [16] was developed as a simple-to-use approach for risk analysts to compute HEPs. One way in which SPAR-H achieves simplicity is through the use of eight PSFs. A PSF is an aspect of the human's individual characteristics, environment, organization, or task that could degrade or improve human performance, thus respectively increasing or decreasing the likelihood of human error.

Many early HRA methods focused on the error likelihood of particular scenarios, whereby the risk analyst would map novel scenarios back to predefined scenarios to extract an HEP. This scenario-based HRA approach proved inflexible in application, and was prone to mismatches. A different approach emerged in SPAR-H and other HRA methods, in which the risk analyst focused not on mapping whole scenarios but on mapping the applicable PSFs within those scenarios. The advent of PSFs brought greater generalizability of HRA and greater inter-analyst reliability through simplified HEP estimation processes. With simplicity, however, there can come a certain degree of ambiguity in the way that PSFs are assigned.

SPAR-H features eight PSFs: Available Time, Stress and Stressors, Complexity, Experience and Training, Procedures, Ergonomics and Human-Machine Interface, Fitness for Duty, and Work Processes. Each PSF features different levels that correspond to multipliers on the nominal HEP. When there is a clear indication that the PSF has no significant contribution to the overall HEP, the multiplier is 1.0, indicating no change to the nominal HEP (i.e., nominal HEP x 1.0 = nominal HEP). Similarly, when there is not enough available information about the PSF level assignment, a multiplier of 1.0 is used. Negative influences are those that increase the likelihood of error. For example, "High stress" increases the likelihood of error over "Nominal stress," and has a corresponding multiplier of 2.0. Negative influences always have a multiplier value greater than 1.0 (e.g., nominal HEP x 2.0 > nominal HEP). Positive influences are those that decrease the likelihood of error. For example, "High Level of Training" decreases the likelihood of error below "Nominal Training," and has a corresponding multiplier of 0.5. Positive influences always have a multiplier value less than 1.0 (e.g., nominal HEP x 0.5 < nominal HEP). While all PSFs have nominal and negative influences, several of the PSFs do not feature positive influences.

SPAR-H is documented extensively in [16].

3.2.3.2 Main References to the SPAR-H Method

[16] Gertman, D., Blackman, H., Marble, J., Byers, J., Haney, L., and Smith, C. (2005): "The SPAR-H Human Reliability Analysis Method," NUREG/CR-6883. Washington, D.C., U.S. Nuclear Regulatory Commission.

3.2.3.3 INL, SPAR-H Analysis of HFE 1A

<u>*Summary of Negative Influences Identified by INL SPAR-H*</u>

For the Cognitive Part (Diagnosis):

Stress and complexity were identified as having a negative influence. An SGTR would produce elevated stress levels in the crew, and there would be a number of variables to consider and track while responding to the event. If the crew is overwhelmed by the complexity or stress, failure is more likely.

For the Execution Part:

Stress was identified as having a negative influence. Complexity was not thought to be a factor anymore, since "once the SGTR and ruptured steam generator has been diagnosed, the situation is no longer complex and the crew's actions are guided by procedures."

<u>Comparison to Empirical Data</u>

There was no evidence that stress or complexity affected the cognitive performance of this HFE, so the predictions were not supported. Only one crew failed to complete the action in the designated time frame, and they were only about a minute and a half late. The main influence on the crews' completion time appeared to be the rate at which they progressed through the procedure, but it was also affected by the time at which they tripped the reactor and entered E-0. The crew that failed performed slowly, but otherwise had many elements of good performance: they were cautious and held meetings, but lacked the sense of urgency that might have helped them to finish in time and avoid the danger of overfilling the steam generator. Thus, aspects of crew–to-crew variability, which could fall into the category of work processes or some other crew-related category, would seem to be a potential contributor to crew performance. Regardless of their work processes or characteristics, however, all but one of the crews met the criterion for HFE 1A, so work processes were not identified as a negative driving factor in the.data for HFE 1A. In any case, the HRA methods were not expected to be able to address these effects in this study.

With respect to the execution part, there was evidence that some crews had problems in implementing E-3 due to "some simulator/home plant differences and to the complex build-up of this procedure step (eight steps mixing local and control room actions)."

<u>Assessment of Negative Influences Identified by INL SPAR-H</u>

For simulator exercises, stress does not generally appear to be an important PSF. This may be different for real events, and additional guidance in SPAR-H for conditions under which stress might be relevant would be helpful.

Since this scenario is consistent with the crews' standard training on SGTR and the indications fit well with the procedures, the scenario did not seem to be cognitively complex to the crews, which was contrary to what was suggested by the SPAR-H analysis. However, the judgment of complexity was the HRA team's judgment, based on their analysis of the scenario and the presence of multiple paths through the procedures. It is difficult to say whether such judgments could be improved with better guidance for estimating PSF levels in these types of scenarios. Also, the SPAR-H analysis did not detect that some crews might have difficulty with executing some of the steps of E-3.

Summary of Positive and Neutral Influences as Identified by INL SPAR-H

For the Cognitive Part:

Procedures and HMI were identified as good. Diagnostic/symptom-oriented procedures and a well-designed interface were identified as positive contributors to performance. The SPAR-H analysis of this HFE identified available time, experience and training, fitness for duty, and work processes as nominal (generally good conditions in the context of SPAR-H).

For the Execution Part:

The HMI was identified as good. Procedures were in this case assumed to be nominal.

Comparison to Empirical Data

As suggested above by the SPAR-H analysis, the relationship between the procedures and the indications did seem to be a strong positive influence in the crews' performance. Thus, these predictions were supported by the data. The INL SPAR-H assumption that available time and experience and training were nominal (which means generally good conditions in SPAR-H) was not inconsistent with the data (see discussion in next section).

Assessment of Positive and Neutral Influences Identified by INL SPAR-H

Treating available time, experience and training, fitness for duty, and work processes as nominal is not inconsistent with the results. However, given that crews are trained in this basic scenario twice a year, it is not clear why training would not be seen as a more positive influence, as it was judged by the assessment team. The SPAR-H team noted that "crews are trained on SGTR scenarios twice a year, so we expect them to understand the situation and what their response should be." Again, judgments relating to the levels of the PSFs can be difficult, and could increase the potential for analyst-to-analyst variability in these types of scenarios.

Summary of Operational Description Provided in the INL SPAR-H Analysis

In their operational description, the INL SPAR-H team noted that "in this scenario, we predict that crews will be able to identify and isolate the ruptured steam generator fairly easily, if they act promptly and follow their procedures. There is sufficient time to accomplish this task if they respond quickly to the initial alarms, and their procedures are diagnostic and symptom-oriented and will direct crews to take the appropriate actions. Crews are trained on SGTR scenarios twice a year, so we expect them to understand the situation and what their response should be."

The SPAR-H analysis presumed nominal time because there is sufficient time for crews to identify and isolate the ruptured steam generator. However, the team noted that this assumes that crews will act promptly and appropriately to the initial alarms. They argued that the key to the success or failure of this HFE is whether or not the crew initiates a manual reactor scram. "If they do not, the reactor will scram automatically in 11 minutes. Once the reactor has tripped, the crew is required to enter procedure E-0. From there, it will take approximately 10 minutes for the crew to transfer to E-3. Given that failure criteria is met if the crew takes longer than 20 minutes to isolate the ruptured steam generator, a crew will automatically fail this HFE if they wait for the reactor to automatically scram." In terms of the SPAR-H analysis, if the crew waits for the reactor to automatically scram, they no longer have enough time to complete the remaining tasks, and the probability of failure increases to 1.0.

The SPAR-H analysis also presumed nominal work processes because the analysts "had no reason to predict otherwise." However, they noted that if work processes are poor for a crew, the probability of failure would be higher: "for example, if a crew does not follow their procedures, the HEP quadruples. If the crew takes too long to take the appropriate steps (due for example, to poor coordination, communication, or command and control), then the probability of failure increases. As discussed above, if the crew waits long enough that the reactor automatically scrams, failure is certain. If they are merely delayed in entering E-0, available time drops to barely adequate and the HEP increases to 0.238. Crews who have a culture of lax procedural adherence, crews who have a slow response time or who take too much time in crew meetings or discussing the plan of action, or crews with poor command and control would be likely to have more difficulty in succeeding on this HFE."

Assessment of Operational Description

The SPAR-H qualitative assessment or operational description is generally consistent with what occurred in the scenario. Their assessment of the factors that could lead to delays seemed reasonable, but such delays were generally not manifested by the crews. One crew did proceed relatively slowly and cautiously, but they were only 1.5 minutes late with respect to the time criterion.

3.2.3.4 INL SPAR-H Analysis of HFE 1B

Summary of Negative Influences Identified by INL SPAR-H

For the Cognitive Part:

Stress, complexity, and HMI were identified as negative influences on this HFE. The situation was said to be highly complex, with multiple equipment and indication failures, and the steamline break and automatic reactor trip would produce elevated stress levels in the crew. The PSF with the strongest impact on this HFE was said to be the misleading indicators. The team noted that, "due to the main steam line break, all primary indications of the steam generator tube rupture are masked. There is sufficient time for the crews to identify and isolate the ruptured steam generator, *if* they promptly identify it. Given the fact that the SGTR is masked, this is unlikely."

For the Execution Part:

Stress was identified as having a negative influence.

Comparison to Empirical Data

The INL SPAR-H analysis straightforwardly identified two of the more important negative influences identified in the analysis of operating crew performance, including complexity and the impact of the misleading indicators (which is covered under the SPAR-H PSF of HMI). In the

SPAR-H analysis, the impact of misleading indications had the greatest impact on the HEP estimate (multiplier of 50), followed by complexity (multiplier of 5), and stress (multiplier of 2).

In the Halden summary of the operating crew results, stress was not identified as an important factor, although at least a few crews noted that (as usual in these situations) they had felt some sense of time pressure. The Halden crew summaries also identified procedures, training, work processes/team dynamics, and execution complexity as factors bearing negatively on performance, but these factors were treated as nominal (generally good) in the INL SPAR-H analysis. The effects of factors like team dynamics and work processes varied among the crews, with some having characteristics that facilitated performance and others having negative effects. The INL SPAR-H analysis did note that work processes could have a negative impact if they were poor, but they had no basis for evaluating this factor.

In addition, the INL SPAR-H summary of the driving PSFs for the operating crews noted that the time allowed to complete the action in this study constrained the likelihood of success. In the SPAR-H analysis, although they did not include time available as a negative PSF (it was treated as nominal in the SPAR-H PSF evaluation since time to complete the action was estimated at 20 minutes and 25 minutes was available), they noted that, "given that success criteria for this HFE are time-limited, we predict that crews are most likely to fail this HFE. They may identify and isolate the ruptured steam generator, but not within the required time." It is unclear why available time was not weighted (treated other than nominally) in the INL SPAR-H analysis.

Assessment of Negative Influences Identified by INL SPAR-H

In the Halden summary of driving PSFs for operating crews, the inadequacy of the procedures and training for the complex scenario, in conjunction with the missing radiation indications and several other factors, were thought to be very important. The INL SPAR-H analysis treated training and procedures as nominal, so they did not directly influence the calculation of the predicted HEP (multipliers of 1.0). However, the SPAR-H documentation noted that, "due to the masking of the steam line break and loss of secondary radiation indications, the procedures will not assist diagnosis." Thus, the analysts were aware of the limitations of the procedures for this scenario, and it is not clear why procedures were treated as nominal in the INL SPAR-H analysis. Maybe the effects of procedures and training were assumed to be covered to some extent by the weighting of the missing indications and complexity, but this was not discussed. In summary, they decided not to weight training, procedures, or available time as important negative driving factors (which was inconsistent with the data), but the SPAR-H analysis did identify two of the important driving factors that influenced those crews who could contribute to the HEP. The analysis predicted that on average the crews would fail 80% of the time, and 7 out of 14 did fail to meet the time criterion of 25 minutes, even though all but one got it done within 32 minutes.

Summary of Positive and Neutral Influences Identified by INL-SPAR-H

For the Cognitive Part:

The SPAR-H analysis identified available time, experience and training, procedures, fitness for duty, and work processes as nominal (generally good) for this HFE.

For the Execution Part:

The INL SPAR-H analysis noted that once (and if) the diagnosis of the SGTR is made, the diagnostic and symptom-oriented procedures and the well-designed HAMMLAB control systems interface should assist the crews in taking the appropriate steps.

<u>Comparison to Empirical Data</u>

Although the empirical data suggested that training and indications had an overall negative influence on crew performance, they were also identified as having a secondary positive influence. The operating crews' general training apparently resulted in a good knowledge base, and, along with the availability of other information (e.g., SG level), they supported the eventual, though late, diagnosis. All crews eventually succeeded due to a knowledge-based diagnosis, using SG-level disparity and the appropriate post-diagnosis procedures. Thus, while the lack of scenario-specific training appeared to hamper the crews and their normal training on SGTR may have made them overly reliant on radiation indications, their knowledge base eventually allowed them to solve the problem (but not within the time allowed in some cases). It is this knowledge base, along with some strong "crew characteristics or work processes" for some crews, that allowed 7 out of 14 crews to succeed within the time frame.

The INL SPAR-H analysis identified available time, experience and training, procedures, fitness for duty, and work processes as nominal for this HFE. Their treatment of training and experience as nominal (generally good) is consistent with the secondary positive influence noted above, but overall the lack of training on the specific scenario was a negative influence (training–somewhat poor). The analysts' assumption that available time and procedures were nominal seems inconsistent with other comments they made, and treating them as nominal was inconsistent with the results (adequacy of time–somewhat poor, procedure guidance-poor). The SPAR-H analysis also noted that, once the SGTR was diagnosed, the procedures for carrying out the actions and the HMI should positively support the crews. This assertion was consistent with some of the crew results, but some crews had problems in implementing E-3. In addition, for the empirical data, it was judged that execution complexity was moderately high. The SPAR-H analysis did not identify the problems with E-3, or that it would be complex for some crews.

<u>Assessment of Positive and Neutral Influences Identified by INL SPAR-H</u>

The INL team states that, "given that success criteria for this HFE are time related; we predict that crews are most likely to fail this HFE. They may identify and isolate the ruptured steam generator, but not within the required time." However, the SPAR-H analysis doesn't include available time as a negative factor in the analysis. In the HRA worksheet, it is stated that "time available was 25 minutes. Time required was approximately 20 minutes." Their treatment of time available and procedures as nominal was inconsistent not only with the empirical data, but also with statements they made in their analysis.

Deciding which factors to weight negatively and which to leave as nominal seems like a relatively difficult process in SPAR-H, since in some instances they can have both characteristics. Guidance for how to balance these aspects in SPAR-H is not provided.

Overall, it seems like the analysts' understanding of the scenario events was generally good, but the translation into the model is a little confusing. The fact that available time and procedures were rated positively seems inconsistent.

Summary of Operational Description Provided in the INL SPAR-H Analysis

According to the INL SPAR-H analysis, the most difficult aspect of this HFE is identifying that a steam generator tube rupture has occurred. This is primarily due to the misleading indications (loss of secondary radiation indications), but the added complexity of the main steam line break will also contribute. Crews will have to recognize the symptoms of the SGTR from indirect indications while they are dealing with the consequences of the main steam line break and automatic reactor scram. Once (and *if*) the SGTR is diagnosed, crews should be able to take the appropriate steps to identify and isolate the ruptured steam generator fairly easily, but it is probable that they will take more time than is permitted to meet success criteria.

As with HFE 1A, the SPAR-H analysis presumed nominal work processes because "we had no reason to predict otherwise...however, if work processes are poor for a crew, the probability of failure would be higher. For example, if a crew does not follow their procedures, the HEP quadruples. If the crew takes too long to take the appropriate steps (e.g., due to poor coordination, communication, or command and control), the probability of failure increases. Crews who have a culture of lax procedural adherence, crews who have a slow response time or who take too much time in crew meetings or discussing the plan of action, or crews with poor command and control would be likely to have more difficulty in succeeding on this HFE."

Assessment of Operational Description

The INL SPAR-H qualitative assessment is quite consistent with Halden's operational summary; they did not specifically discuss the difficulties with the procedures and the associated delays that the operating crews experienced, but this could be implied by the complexity factor and the impact of the missing radiation cues, which they identified as important. As discussed above, the INL SPAR-H analysis has noted that the procedures would not help with the diagnosis, though they treated them as nominal. Their assessment of the work-process-related factors that *could* lead to delays seemed reasonable, but they did not have enough information with which to make predictions. The work process requirements were high, and some of the noted crew factors may have influenced crew outcomes in the complex scenario.

3.2.3.5 Additional Comments on INL SPAR-H Analysis

Comments on INL SPAR-H Analysis of HFE 1A

The decision on how and which PSF to rate is based on the analysts' judgment in using SPAR-H, and it is not always obvious why the choices are made. Some additional guidance in SPAR-H as to how to consider the PSFs together and make such judgments would be very useful, and additional documentation on these decisions would improve traceability.

Decisions about multipliers can be based on a number of factors, and have large impacts on the HEPs. SPAR-H probably intends to be relatively flexible in this regard; that is, it is ultimately left to the analyst. If, however, analysts are expected to consider the relative weights across PSFs, which appears necessary, additional guidance and documentation would be helpful.

Insights For Error Reduction

Along with a good task analysis, several of the PSFs included in the SPAR-H method should allow insights into improving safety. That is, the method examines aspects that, when identified as problematic, could be improved to facilitate error reduction; however, this will depend heavily on the judgments made about the different potential PSFs and their levels. Since the crews generally did well in this scenario, the negative PSFs identified in the SPAR-H analysis were not really relevant to improving performance.

Impact on HEP (Sensitivity to Driving Factors)

Since there were no driving negative factors, the relatively low HEP seemed to reflect the analysis.

Guidance and Traceability

As with most HRA methods, there is room for improving method guidance (see discussion above, and below for HFE 1B). The SPAR-H method results are traceable as long as there are good discussions on the reasons for choosing the PSFs and their weights. For this analysis, particularly for HFE 1B, additional discussion would have been helpful.

<u>*Comments on INL SPAR-H Analysis for HFE 1B*</u>

The basis for the assignment of PSF levels was not clear in all cases, as the analysis indicated some inconsistency in the assignments. Determining which and how many PSFs to include as negative or positive influences, as well as how to assign the PSFs levels, can be a complicated process in SPAR-H, at least for these types of scenarios. There does not appear to be adequate guidance to help analysts make such judgments: for example, it is not clear why they rated available time, procedures, and training as nominal. In addition, misleading indications would seem to have aspects that would contribute to complexity. Thus, it is not clear why a complexity PSF was also needed, particularly when misleading indications had a multiplier of 50, while complexity had a multiplier of 5. These characteristics would seem to suggest analyst-to-analyst variability; nevertheless, the method results were consistent with the analysts' opinion that there would be a high probability of failure to respond within the given time frame.

Decisions about multipliers can be based on a number of factors, and have large impacts on the HEPS. SPAR-H probably intends to be relatively flexible in this regard: that is, it is ultimately left to the analyst. If, however, analysts are expected to consider the relative weights across PSFs, which appears necessary, additional guidance and documentation would be helpful.

Insights for Error Reduction

As noted above for HFE 1A, in conjunction with a good task analysis, several of the PSFs included in the SPAR-H method should allow insights into improving safety. That is, the method examines aspects that, when identified as problematic, could be improved to facilitate error reduction. However, this will depend heavily on the judgments made about the different potential PSFs and their levels. While this SPAR-H analysis identified some factors needing improvements, several important ones, particularly the problems with E-0, could have been covered more directly to support improvements.

Impact on HEP (Sensitivity to Driving Factors)

With a multiplier of 50, the HMI PSF (misleading indicators) was the main driving factor in the INL SPAR-H analysis. In the SPAR-H analysis, the resulting HEPs are generally sensitive to the multipliers assigned, unless many PSFs are included.

Guidance and Traceability

As with most HRA methods, there is room for improving method guidance. (See previous section and the same section for HFE 1A for discussion of needed guidance.) The SPAR-H method results are traceable as long as there are good discussions on the reasons for choosing the PSFs and their weights. For this analysis, particularly for HFE 1B, additional discussion would have been helpful.

3.2.3.6 INL SPAR-H Team Comments on the Original Comparison

No comments were provided.

3.2.4 SPAR-H (NRC)

See Section 3.2.3.1 for a short overview of SPAR-H and 3.2.3.2 for main reference [16].

3.2.4.1 NRC, SPAR-H Analysis of HFE 1A

Summary of Negative Influences Identified by NRC SPAR-H

For the Cognitive Part (Diagnosis):

No negative influences were identified. All but two of the PSFs were assumed to be nominal for this event.

For the Execution Part:

The execution part was treated almost the same as the cognitive part. No negative influences were identified. All but one of the PSFs were assumed to be nominal for this event.

Comparison to Empirical Data

Only one crew failed to complete the action in the designated time frame, and they were only about a minute and a half late. The main factor in the crews' completion time appeared to be the rate at which they progressed through the procedure, but it was also impacted by the time at which the crew tripped the reactor and entered E-0. The crew that failed performed slowly, but otherwise had many elements of good performance: they were cautious and held meetings, but lacked the sense of urgency that might have helped them to finish in time and avoid the danger of overfilling the steam generator. Thus, aspects of crew-to-crew variability, which could fall into the category of work processes (or some other crew-related category), would seem to be a potential contributor to crew performance. Regardless of their work processes or characteristics, however, all but one of the crews met the criterion for HFE 1A. Thus, work processes were not identified as a negative driving factor in the data for HFE 1A. In any case, the HRA methods were not expected to be able to address these effects in this study.

With respect to the execution, there was evidence that some crews had problems implementing E-3 due to "some simulator/home plant differences and to the complex build-up of this procedure step (eight steps mixing local and control room actions)." Execution complexity was identified as a "somewhat high" negative influence in the data, but this factor was not identified as a negative PSF in the NRC SPAR-H analysis.

Assessment of Negative Influences Identified by NRC SPAR-H

No negative factors were identified, and, with the exception of the minor problems identified with executing E-3, this result is generally consistent with the Halden results.

Summary of Positive and Neutral Influences as Identified by NRC SPAR-H

For the Cognitive Part:

Available time, stress, complexity, ergonomics, fitness for duty, and work processes were all assumed to be nominal. Experience and training were assumed to be high and procedures were categorized as diagnostic/symptom-oriented, so these two PSFs would be the most important in a generally good situation for success.

Note that in calculating the time available, the analysts apparently assumed that the crew would diagnose a problem and that the plant would be tripped within 5 minutes. This is because they added the 8 to 10 minutes estimated to get through E-0 and the 5 minutes to diagnose and perform the actions in E-3, leaving them with 13 to 15 minutes estimated for the time required for the actions. This would then leave 5 to 7 minutes for the original diagnosis to trip the plant, which they apparently assumed would be adequate (i.e., in the context of SPAR-H, the 5 to 7 minutes would on average be sufficient for diagnosing the need to trip the plant). Therefore, nominal time available was selected for both diagnosis and execution.

For the Execution Part:

Procedures were in this case assumed to be nominal, so training and experience was the only PSF assumed higher than nominal, giving a positive influence to performance.

Comparison to Empirical Data

Experience and training were assumed to be high, and procedures were categorized as diagnostic/symptom-oriented for the execution part of the task. With the exception of the effect of execution complexity in E-3, the latter appeared to be true. All other PSFs were considered nominal.

Assessment of Positive and Neutral Influences Identified by NRC SPAR-H

Treating available time, stress, complexity, ergonomics, fitness for duty, and work processes as nominal (which means generally good in SPAR-H) is not inconsistent with the results. Similarly, improving the HEP because of experience and training is not unreasonable.

Summary of Operational Description Provided in the NRC SPAR-H Analysis

It was assumed that there would be enough time for the diagnosis of the SGTR, for working through the procedures, and for isolating the faulted SG. It was decided that the diagnosis would be relatively easy due to the training and experience of the crews with this type of scenario, along with the diagnostic symptom-based procedures which would allow the SGTR to be identified with the relevant cues that would be available. Similarly, executing the needed actions was determined to be relatively easy, due to the nominal level of most of the PSFs and the high level of training and experience.

Assessment of Operational Description

The NRC SPAR-H qualitative assessment is generally consistent with Halden's operational summary.

3.2.4.2 NRC, SPAR-H Analysis of HFE 1B

Summary of Negative Influences Identified by NRC SPAR-H

For the Cognitive Part:

Two negative influences were identified, the complexity associated with diagnosis and the stress level. Moderate complexity was assumed instead of nominal complexity because "the progression of the accident contains many additional variables (beyond the base scenario) and requires concurrent diagnosis." High stress was selected due to the unexpected multiple annunciators creating a potentially disruptive atmosphere. All other PSFs were nominal for this event.

For the Execution Part:

The execution part was treated the same as the cognitive part, with moderate complexity and high stress associated with the response. All of the other PSFs were assumed to be nominal for this event.

Comparison to Empirical Data

The NRC SPAR-H analysis identified complexity, one of the more important negative influences named in the Halden analysis of operating crew performance. They noted that "in the complex scenario, the progression of the accident contains many additional variables and requires concurrent diagnoses." Moderate complexity was assumed, giving a multiplier of two. In the Halden summary of the operating crew results, stress was not identified as an important factor, although at least a few of the crews noted that, as is usual in these situations, they had felt some sense of time pressure. The NRC SPAR-H analysis listed stress as being an important factor because "the unexpected multiple annunciators create a potentially disruptive atmosphere." In this case, "high" stress was selected, giving a multiplier of two.

The Halden crew summaries also identified procedures, training, work processes/team dynamics, and execution complexity as factors bearing negatively on performance. The effects of factors like team dynamics and work processes varied among the crews, with some having characteristics that facilitated performance and others having effects.

In addition, the summary of the driving PSFs for the operating crews noted that the time allowed to complete the action in this study constrained the likelihood of success (rated as somewhat poor), that is, it was likely to decrease the success rate. In the SPAR-H analysis, time available was assumed to be nominal (sufficient).

In the NRC SPAR-H analysis, both stress and complexity were also assumed to be the most important PSFs for failing to execute the response, but there was no clear evidence that this was the case. The analysis did not seem to refer to the minor problems with the execution of E-3 step 3.

Assessment of Negative Influences Identified by NRC SPAR-H

In the Halden summary of driving PSFs for operating crews, the inadequacy of the procedures and training for the complex scenario, in conjunction with the missing radiation indications and several other factors, were thought to be very important. The NRC SPAR-H analysis treated training and procedures as nominal, so they did not directly influence the calculation of the predicted HEP (multipliers of 1.0). Similarly, the SPAR-H documentation did not explicitly address how the masking by the steam line break and loss of secondary radiation indications were considered. However, it is reasonable to expect that this aspect was covered under the complexity PSF. Overall, this SPAR-H analysis did not seem to anticipate a number of negative factors that would be influencing performance (significantly slowing it down relative to the base case), and indicated that the crews would frequently be successful (relatively low HEP); however, this scenario turned out to be fairly difficult. Most of the crews were confused to some degree, and eventually had to base the transfer to E-3 on a knowledge-based judgment using a different piece of information than was usual (SG level disparity). Seven out of fourteen crews failed to meet the time criterion of 25 minutes, even though all but one got it done within 32 minutes.

Summary of Positive and Neutral Influences Identified by NRC SPAR-H

For the cognitive part:

Available time, experience and training, procedures, ergonomics/MMI, fitness for duty, and work processes were all assumed to be nominal. Thus, a generally good situation for success is only impacted by the factors leading to less-than-nominal conditions.

For the execution part:

The execution part was treated the same as the cognitive part. No PSFs led to better-than-nominal conditions.

Comparison to Empirical Data

Although the empirical data suggested that training and indications had an overall negative influence on crew performance, they were also identified as having a secondary positive influence. The operating crews' general training apparently resulted in a good knowledge base, and, along with the availability of other information (e.g., SG level), they supported the eventual (if late) diagnosis. All crews eventually succeeded with a knowledge-based diagnosis, using SG-level disparity and the appropriate procedures after the diagnosis. Thus, while the lack of scenario-specific training appeared to hamper the crews and their normal training on SGTR may have made them overly reliant on radiation indications, their knowledge base eventually allowed them to solve the problem (but not always within the allotted time). It is this knowledge base, along with some strong "crew characteristics or work processes" for some crews, that allowed 7 out of 14 crews to succeed within the given time frame.

The NRC SPAR-H analysis identified available time, experience and training, procedures, HMI, fitness for duty, and work processes as nominal for this HFE. *Nominal* in SPAR-H generally indicates that the factors will support or enhance performance. The analysts' assumption that available time and procedures were nominal was generally inconsistent with the results; however, the assumption that the crews' experience/training was nominal, if referring to their resulting knowledge base, is consistent with the results.

Assessment of Positive and Neutral Influences Identified by NRC SPAR-H

It appeared that, in general, the assessment of positive factors was overly optimistic compared to the actual results, at least when the time delays are considered and the initial levels of confusion are experienced by the crews.

Summary of Operational Description Provided in the NRC SPAR-H Analysis

Based on the (relatively low) HEP, this event is not considered overly difficult. It is assumed to be somewhat complex due to the additional variables (beyond the base scenario) to be considered and the required concurrent diagnoses (SGTR and steamline break); additionally, high stress was selected due to the unexpected multiple annunciators creating a potentially disruptive atmosphere. However, it was thought that the crews receive enough relevant training on individual aspects of the scenario, and that, even though the isolation of the main steam isolation valves and the failure of a secondary radiation detector will reduce the effectiveness of the procedural guidance relative to the base scenario, training and procedures were at least nominal in supporting the correct diagnosis and response.

Assessment of Operational Description

The NRC SPAR-H qualitative assessment is not completely inconsistent with Halden's operational summary. The assessors noted the potential difficulties with the procedures ("the isolation of the main steam isolation valves and the failure of a secondary radiation detector will reduce the effectiveness of the procedural guidance"), but still rated them as nominal for the scenario. They also argued that the crews receive enough relevant training on individual aspects of the scenario, which was generally true from the standpoint that they all performed a knowledge-based diagnosis and eventually succeeded. Finally, they indicated that the complexity of the scenario would have an impact. However, they failed to predict the extent to which several crews would be slowed by the context.

3.2.4.3 Additional Comments on NRC SPAR-H Analysis

Comments on NRC SPAR-H Analysis of HFE 1A

Nothing specific about the method was indicated in the analysis of HFE 1A. Training and procedures were rated better-than-nominal. These types of judgments, which factors and why, seem relatively subtle in using SPAR-H, and it seems that additional guidance would be beneficial. Also see notes on the NRC SPAR-H analysis of HFE 1B below for further comment.

Insights for Error Reduction

Along with a good task analysis, several of the PSFs included in the SPAR-H method should allow insights into improving safety: that is, the method examines aspects that, when identified as problematic, could be improved to facilitate error reduction. However, this will depend heavily on the judgments made about the different potential PSFs and their levels. Since the crews did generally well in this scenario, no changes needed were identified.

Impact on HEP (Sensitivity to Driving Factors)

Since there were no driving negative factors, the low HEP seemed to reflect the analysis.

Guidance and Traceability

As with most HRA methods, there is room for improving method guidance. See the comments on HFE 1B for a discussion of necessary guidance. The SPAR-H method results are traceable as long as there are good discussions on the reasons for choosing the PSFs and their weights. For this analysis, additional discussion would have been helpful.

Comments on NRC SPAR-H Analysis of HFE 1B

In SPRA-H, how to rate each PSF is based on the analysts' judgment, and it is not always obvious why the choices are made. Determining which and how many PSFs to include as negative or positive influences, as well as how to assign the PSFs levels, can be a complicated process in SPAR-H, at least for these types of scenarios; for example, it is not clear from the analysts' discussion why they rated the procedures as nominal. Some additional guidance in SPAR-H as to how to consider the PSFs together and make such judgments would be very useful.

Decisions about multipliers can be based on a number of factors, and SPAR-H probably intends to be relatively flexible in this regard: that is, it is ultimately left to the analyst. If, however, analysts are expected to consider the relative weights across PSFs, which appears necessary, additional guidance and documentation would be helpful.

Insights for Error Reduction

In conjunction with a good task analysis, several of the PSFs included in the SPAR-H method should allow insights into improving safety: that is, the method examines aspects that, when identified as problematic, could be improved to facilitate error reduction. However, this will depend heavily on the judgments made about the different potential PSFs and their levels. While this SPAR-H analysis identified some factors needing improvements, several important ones, particularly the problems with E-0 (procedures) and training, could have been covered more directly to support improvements.

Impact on HEP (Sensitivity to Driving Factors)

In the SPAR-H analysis, the resulting HEPs are generally sensitive to the multipliers assigned, unless many PSFs are included. In this analysis, only two PSFs were identified as driving performance, and they were weighted equally. One, complexity, was relatively important in the Halden results, but the other, stress, was not identified as being very important.

Guidance and Traceability

As with most HRA methods, there is room for improving method guidance. See notes above for discussion of necessary guidance. The SPAR-H method results are traceable as long as there are good discussions on the reasons for choosing the PSFs and their weights. For this analysis, particularly for HFE 1B, additional discussion would have been helpful.

3.2.4.4 NRC SPAR-H Team Comments on the Original Comparison

No comments were provided.

3.2.5 CESA-Q (PSI)

3.2.5.1 Short Overview of the CESA-Q Quantification Method (by Luca Podofillini and Bernhard Reer)

CESA-Q is the quantification module of the Commission Errors Search and Assessment (CESA) method, which was developed by the HRA project at the Paul Scherrer Institut (PSI). The method is intended to guide the identification and prioritization of aggravating operator actions in post-initiator scenarios, such as EOCs. The PSI work on CESA started with development of an EOC identification module [17, 19]; continuous quality assurance effort led to the improved version presented in [18].

EOC quantification was addressed in a later PSI project, resulting in an outline of a method for EOC quantification, CESA-Q [20] (refer to [20] for a complete description of the method). In this method, the EOC is analyzed in terms of plant- and scenario-specific factors. Two groups of factors are introduced: situational factors, which identify EOC-motivating contexts, and adjustment factors, which refine the analysis of EOCs to estimate the strength of the motivating context. A reliability index is introduced, to represent the overall belief of the analyst regarding the positive or negative effects on the EOC probability (ranging from zero for strongly "error-forcing" contexts to nine for contexts with very low EOC probabilities). Quantification is performed by comparing the pattern of the factors' evaluations with patterns of catalogued reference EOCs (identified from 26 operational events, previously analyzed both qualitatively and quantitatively in [21, 22]).

It must be noted that the CESA-Q development and previous applications have focused on EOCs, while this HRA empirical study addresses errors of omissions. Therefore, the application of CESA-Q in the study was explorative. Modifications to the **method** are planned to account for the feedback from the first phase of the empirical study.

3.2.5.2 Main References to the CESA-Q Method

[17] V.N. Dang, B. Reer, S. Hirschberg: "Analyzing Errors of Commission: Identification and First Assessment for a Swiss Plant," Building the New HRA: Errors of Commission–from Research to Application, OECD NEA Workshop, Rockville, MD, USA, May 7-9, 2001 (published by OECD in 2002: NEA/CSNI/R(2002)3, 105-116).

[18] B. Reer, V.N. Dang: "The Commission Errors Search and Assessment (CESA) Method PSI Report Nr. 07-03," ISSN 1019-0643, May 2007, Paul Scherrer Institut, Switzerland.

[19] B. Reer, V.N. Dang, S. Hirschberg: "The CESA method and its application in a plant-specific pilot study on errors of commission," Reliability Engineering and System Safety 83 (2004), 187-205.

[20] B. Reer: "Outline of a Method for Quantifying Errors of Commission," LEA 09-302, Laboratory for Energy Systems Analysis, Paul Scherrer Institute, Villigen PSI, Switzerland, 2009.

[21] B. Reer, V.N. Dang: "Situational Features of Errors of Commission Identified from Operating Experience," LEA 09-303, Laboratory for Energy Systems Analysis, Paul Scherrer Institute, Villigen PSI, Switzerland, 2009.

[22] B. Reer: "An Approach for Ranking EOC Situations Based on Situational Factors," LEA 09-304, Laboratory for Energy Systems Analysis, Paul Scherrer Institute, Villigen PSI, Switzerland, 2009

3.2.5.3 PSI, CESA-Q Analysis of HFE 1A

<u>*Summary of Negative Influences Identified by PSI CESA-Q*</u>

For the Cognitive Part:

CESA-Q analysis focuses on decision errors (essentially errors at decision points in the procedures), and no negative influences or error-forcing conditions were identified for this event.

In the context of the CESA-Q model, only "random errors" were assumed plausible, and included:

- misperception of the behavior of parameters (SG pressure behavior and the status of the components to be manipulated in E-3 step 3)
- interpretation and communication of instructions or rules due to misinterpretation of the labels of the components to be manipulated in E-3 step 3
- time pressure associated with an incorrect response or with the task to be performed (SG isolation)

Time pressure (sense of urgency) was seen as a potential negative influence in the sense that it was thought to be plausible that the crews might feel some time pressure in performing the task, and that this could contribute to the potential for a random error.

Limited credit for recovery (e.g., from going down a wrong path in the procedure), which is part of the model, was given, owing to shortage of time and potential delays in perceiving feedback (e.g., SG level).

For the Execution Part:

The second bullet (interpretation and communication of instructions or rules due to misinterpretation of the labels of the components to be manipulated in E-3 step 3) has aspects associated with random errors in execution, but, again, no negative PSFs were identified.

<u>Comparison to Empirical Data</u>

Only one crew failed to complete the action within the designated time frame, and they were only about a minute and half late. The main influence on the crews' completion time appeared to be the rate at which they progressed through the procedure, but it was also impacted by the time at which the crew tripped the reactor and entered E-0. The crew that failed performed slowly, but otherwise had many elements of good performance: they were cautious and held meetings, but lacked the sense of urgency that might have helped them to finish in time and avoid the danger of overfilling the steam generator. Thus, aspects of crew-to-crew variability in what could fall into the category of work processes (or some other crew-related category) would seem to be a potential contributor to crew performance. However, regardless of their work processes or characteristics, all but one of the crews met the criterion for HFE 1A. Thus, work processes were not identified as a negative driving factor in the data for HFE 1A. In any case, the HRA methods were not expected to be able to address these effects in this study.

The CESA-Q analysis did not identify any negative PSFs, but evaluated the potential for "random error." The main negative aspects appeared to be the limited time available for recovery and the potential for time pressure to contribute to random error. As discussed above, only one crew failed to complete the action within the designated time, but limited time for recovery could have been relevant to some crews if a random error had occurred. In addition, the CESA-Q analysis listed "interpretation and communication of instructions or rules, due to misinterpretation of the labels of the components to be manipulated in step 3 of E-3" and "misperception of the behavior of parameters (SG pressure behavior and the status of the components to be manipulated in E-3 step 3)" as potential sources of random error. While the crews did not make any random errors, some minor problems in implementing E-3 step 3 slowed some crews. Whether it had to do with "misinterpretation of the labels of the components to be manipulated" or "perception of the behavior of parameters" is unclear.

Thus, CESA-Q's determination of no negative driving PSFs, except for potential time limitations, was consistent with the results.

Assessment of Negative Influences Identified by PSI CESA-Q

No driving negative factors were identified, which is consistent with the results.

Summary of Positive and Neutral Influences as Identified by PSI CESA-Q Analysis

For the Cognitive Part:

Based on this analysis, the conditions are essentially nominal for the correct responses. In CESA-Q terms, there were no exceptional conditions, no misleading conditions or instructions, no distractions, and no incentives for risky actions. Thus, the positive side of these characteristics was apparently assumed to be present. With no error-forcing conditions, conditions for success are apparently good (as indicated by the HEP).

Plus, even though the time to perceive cues relevant to recovery is short, there are cues available.

For the Execution Part:

Conditions were apparently sufficient.

Comparison to Empirical Data

No exceptional positive conditions were identified by CESA-Q; rather, everything was assumed to be nominal, which in CESA-Q produced a relatively low HEP even before recovery. The assumption of generally good conditions was consistent with the results (i.e., indications of conditions, training and experience, and procedural guidance were all identified as good for the crews).

Assessment of Positive and Neutral Influences Identified by CESA-Q

See comparison above.

Summary of Operational Description Provided in the CESA-Q Analysis

It was apparently assumed that there would be enough time for the correct decisions and responses. The events with highest likelihood of random error were associated with two procedure steps:

- Operators erroneously transfer from E-0 step 18 to E-2 ("Isolation of faulted SG").
- Operators fail to identify and isolate the ruptured SG (E-3 step 3).

However, there was no expectation that the operators would have any problem in 1A.

Assessment of Operational Description

CESA-Q's suggestions for the events with the highest likelihood of random error could not be verified in this study, but a few crews were slowed by aspects of E-3, step 3. Nevertheless, although cursory, the CESA-Q qualitative assessment is generally consistent with Halden's operational summary; their analysis indicated that there would not be any major problems for the crews.

3.2.5.4 PSI, CESA-Q Analysis of HFE 1B

Summary of Negative Influences Identified by PSI CESA-Q

For the Cognitive Part:

CESA-Q analysis focuses on decision errors (essentially errors at decision points in the procedures). The main negative influence is the "Adverse Exception," a negative PSF in the context of CESA-Q, created by the exceptional condition of the combined steamline break (SLB) and SGTR. This situation causes crews to miss important cues for the correct response, and some potential for misleading cues (e.g., the initial drop in all SG pressures has some potential to incorrectly lead the crew to E-2 (isolation of faulted SG)).

Negative adjustment factors to the HEP for this event are due to 1) the situation increasing the cognitive requirements, 2) the likelihood that the hint for the correct response in ES-1.1 (in the foldout page) will not be checked frequently (the analysis assumes that the crew will enter ES-1.1 from E-0), and 3) time pressure (limited time for making the correct decision).

In addition, limited credit for recovery (e.g., from going down a wrong path in the procedure), which is part of the model, was given, due to shortage of time, especially if focused on steps of ES-1.1. Also, the main cue for this event is masked (radiation level).

For the Execution Part:

Shortage of time seems to be the main concern, as it limits the potential for recovery.

Comparison to Empirical Data

Although the terminology differs somewhat, the CESA-Q analysis identified several of the main driving PSFs included in the summary of the operating crew results. The adverse exception PSF (situational feature) identified in CESA-Q is a negative PSF created by the exceptional condition of the combined SLB and SGTR. It is represented by "important cues missing for the correct response and some potential for misleading cues (e.g., the initial drop in all SG pressures has some potential to incorrectly lead the crew to E-2 (isolation of faulted SG))." This essentially refers to the masking effect in the scenario, which is covered under complexity in the data summary.

The analysts also argued that there were three "negative adjustment factors" to the HEP for this event. The first was "the situation increasing the cognitive requirements (which is consistent with complexity). The second was that the hint for the correct response in ES-1.1 (in the foldout page) may not be checked frequently. The CESA-Q analysis assumed that the crews "will enter ES-1.1 from E-0," and, in fact, many of the crews did enter ES-1.1. Only a couple of the teams actually used the foldout page to get to E-3, but it was not clear whether this was a checking issue. Many of the crews that entered ES-1.1 eventually left and managed to get to E-3 from a knowledge-based diagnosis and decision to jump to E-3. The third negative adjustment factor was time pressure, which they couched in terms of limited time for making the correct decision and for recovery.

In several respects, this discussion is consistent with what was identified as factors driving the crews' performance. Certainly, complexity made it harder to get to E-3 in a timely manner, and time limitations led to several crews not responding within the given time frame. In addition, the analysis appeared to recognize that E-0 (the main procedure) would not generally get the crews to E-3, and many crews did enter ES-1.1. However, only a couple of the crews actually got to E-3 directly from the foldout page. Thus, ES-1.1 also came up short as a procedure to solve this event. Nevertheless, the CESA-Q analysts based their assumption that the crews would enter ES-1.1 on the plots in the information package for the complex scenario, where the PRZ pressure starts to increase (from ~127 bar) at 12:50 (1 min after the IE). Based on this, one could infer that the criterion "RCS pressure stable or increasing" (E-0 step 21) for entering ES-1.1 was met; and, in fact, 5 out of 14 crews did enter ES-1.1. However, other crews apparently did not see the criterion as being met, so it was not obvious to all. In any case, the CESA-Q method identified several of the main "PSFs" driving the crews' behavior.

The analysis did not explicitly identify training as a negative factor (see discussion below [next page] in the section on the "Comparison to empirical data" for the positive influences in the CESA-Q analysis of HFE 1B), and the empirical analysis suggested that more specific training on this SGTR scenario would have improved performance. The crews appeared to rely heavily on the radiation indications, and it was not immediately clear what to do without them. Even when many of the crews recognized the diverging SG levels, the absence of radiation slowed them.

Assessment of Negative Influences Identified by PSI CESA-Q

The analysis of negative PSFs performed by the CESA-Q team was generally good and consistent with the results, except that the procedure set's inability to get the crews to E-3 (probably in conjunction with limited specific training on this scenario) may not have been "weighted" negatively enough. The relatively low HEP produced by CESA-Q suggests that more crews would have been successful than the 8 out 14 that succeeded.

Summary of Positive and Neutral Influences Identified by PSI CESA-Q

For the Cognitive Part:

Based on this analysis, the main positive factors were that 1) level indications were available and clearly visible, 2) negligible physical effort was required for verification of the correct cues, 3) there was no particular benefit in staying in ES-1.1 in this scenario, and 4) personal redundancy was available (could be considered teamwork/communication).

It was also thought that cues for recovery would be available (differences in SG level), and, even though time is short, some time for recovery is available.

For the Execution Part:

Conditions were apparently sufficient.

Comparison to Empirical Data

The availability of information (SG level) was identified as a positive influence by CESA-Q. The alternative indications (e.g., SG level) were identified as a secondary positive influence in the data analysis, the indications were more generally negative, but the CESA-Q prediction was consistent with these results. As CESA-Q also pointed out that there is no particular benefit in staying with ES-1.1, the crews looked for other ways to get to E-3, and finally found one, based on their knowledge and the SG level divergence. The CESA-Q analysis did not list the crews' general training on SGTR as clear support for their knowledge-based decision, which would have made it a positive influence; however, per comments from the CESA-Q team, situational features in CESA-Q are assessed in the context of operator training (e.g., to assess whether a particular instruction or indication can be classified as misleading (or not) or whether a required verification by backup signals is difficult (or not), training has to be taken into account). Thus, training may not be explicitly called out in the analysis, but could be inferred by those knowledgeable about the method. They also proposed personal redundancy as a positive situational feature, but it seemed that there was significant crew-to-crew variability along this dimension in this scenario. Thus, although this factor appeared to have an influence, it was not necessarily positive across crews. However, recognizing this type of influence was not expected for this exercise, so they had no clear basis for judgment.

Assessment of Positive and Neutral Influences Identified by PSI CESA-Q

The CESA-Q analysis recognized the strength of the SG level cue (availability of information), but did not explicitly note that the crews' training/knowledge base was an important positive influence (or situational factor). Good training apparently underlies the positive assessment of the strength of the SG level cue; as noted in a CESA-Q team comment, the role that training plays in the analysis needs to be more clearly outlined in the method description. The citation of "negligible physical effort being required for verification of the correct cues" may have been a positive situational factor, but it does not appear to drive performance. Nevertheless, it is hard to say that it was not a positive situational factor, and it is certainly preferable to the converse.

Summary of Operational Description Provided in the PSI CESA-Q Analysis

It is expected that the failure will most likely come from the fact that the operators focus on following ES-1.1 immediately after entering it, and consequently overlook the transfer to E-3 on the foldout page.

We expect that all crews will succeed in 1B; indeed, as suggested by the relatively low failure probability, we don't expect to see failures.

However, the decision to transfer to E-3 in this case is not straightforward, and has been modeled with a failure probability of 0.1 at the decision points. This value can be interpreted at different levels: at the "probabilistic level," we may expect that 1 out of 10 crews will take the "wrong" procedural path and enter into ES-1.1, start going through the procedure, eventually recover (i.e., realize that the increasing level in one SG is a cue for SGTR), and finally transfer to E-3.

This does not mean that we expect that 9 out of 10 crews will have no problem with 1B; on another level, the value of 0.1 can also be interpreted as a degree of difficulty with action 1B. In this sense, we expect that the crews, in general, will have some difficulties in fulfilling the task. We would therefore expect to see discussion among the operators about missing high radiation level indications, and whether or not to enter ES-1.1 when operators reach step 21, right after passing step 19.

Assessment of Operational Description

It is not clear that the first comment in the CESA-Q qualitative assessment is exactly correct. The crews didn't seem to get too caught up in following ES-1.1, but it was true that most did not use the foldout page. However, the CESA-Q analysis states that the crews "would enter into ES-1.1, start going through the procedure and then at some point recover (i.e., realize that the increasing level in one SG is a cue for SGTR), and finally transfer to E-3," which is generally an accurate statement of what happened for many crews.

They also note that the decision to transfer to E-3 in this case is not straightforward, but they give credit for the crews recovering from this problem. The main limitation in the CESA-Q qualitative assessment was a failure to recognize the degree to which the crews would be slowed by the situation and the number of crews that would follow a wrong path, at least for a while. They were correct that all crews would have some difficulty, and that they all would eventually succeed. Overall, the assessment was very good.

3.2.5.5 Additional Comments on PSI, CESA-Q Analysis

Comments on PSI, CESA-Q Analysis of HFE 1A and HFE 1B

In general, the method seemed to provide a reasonable set of situational factors with which to represent important factors in the scenario being analyzed. In this analysis, the selection process seemed relatively straightforward, but, without more experience with the method, it is difficult to determine whether the levels of the situational features can become hard to distinguish and whether the one set of factors is sufficient for most scenarios. It may have been the optimistic judgment about the time pressure situational feature that led to the optimistic HEP[8], but whether this was due to a lack of adequate guidance for discriminating between the situational factor levels is unclear. This would be an issue for the method developers to investigate and address as needed. Additional guidance might be useful. The method application did seem to benefit from a good task analysis, but it's currently unclear whether this was due to the method or to the analyst team. The credit for recovery, given by the analysis, seemed to result in a somewhat optimistic HEP, at least as is relative to the designated time frame; however, it may be more accurate for the HEP to reflect the overall performance (i.e., all crews eventually succeeded, and, of those that missed the deadline, all but one made it within 10 minutes of the criterion).

Alternatively, the CESA-Q system of comparing ratings on situational factors to event descriptions associated with a particular HEP may have produced the somewhat optimistic HEP, but this is not clear.

Insights for Error Reduction

In conjunction with a good task analysis, the PSFs and situational factors included in the CESA-Q method should allow insights into improving safety; that is, the method examines aspects that, when identified as problematic, could be improved to facilitate error reduction. However, this will depend heavily on the judgments made about the different potential situational factors. For this application, a number of areas that could stand improvement were identified.

Impact on HEP (Sensitivity to Driving Factors)

This aspect needs to be investigated further, but it appears that the resulting HEP did not reflect the discussed importance of the identified situational features[9].

Guidance and Traceability

It should be noted that the CESA-Q method was developed for EOCs, and was being adjusted for use in this application. Since the developers of the method were performing the application, it is hard to judge whether or not the guidance for selecting situational features will be adequate for most users. The guidance, as represented in the material sent for this application, seemed sparse. The derivation of the HEPs within the method and performance drivers is traceable, but the underlying basis for the final HEPs (underlying data) is not clear[10].

---

[8] Note that in reviewing the comparison, the CESA-Q team made the following observation: "We think our HEP was optimistic since it did not really reflect the impact of time availability, more than time pressure (see next comment). The CESA-Q time pressure factor evaluates whether there is an urgency to act (e.g., only a few seconds are available to decide whether the motivated action is wrong or not). It reflects a situation in which the crew would rush into a decision because they felt they didn't have much time. This was not observed in the simulator."

[9] In response to this comment, the CESA-Q team observed that "when we looked a posteriori at our analysis (informed by the experimental outcome), our impression was that our analysis, and thus our HEP, did not really reflect the impact of the operators being slow, but not committing any error. Therefore, one may say that our analysis was not enough sensitive to this specific aspect of the time availability. (Note that in our analysis, time availability impacts our HEP in the short time for recovery from an incorrect decision.)"

[10] Note that documentation of the underlying basis of the decision HEPs is in progress; see references 6 and 7 in the reference section above.

3.2.5.6 PSI CESA-Q Team Comments on the Original Comparison

On the whole, the comparison represents the CESA-Q analysis fairly. An important aspect of the comparison methodology has been the interaction of the assessment group (the comparison group) with the HRA teams. This has taken place at several points during the study; the feedback has ensured that our analyses and the resulting findings are expressed clearly.

We would like to emphasize that the application of CESA-Q in this study is exploratory: the method's development and previous applications have focused on errors of commission, while this study addresses errors of omissions.

The comparison of predictive analyses with a set of actual crew performances in the simulator has provided a number of insights into the CESA-Q method: for instance, in treating time, CESA-Q focuses on the effect of time pressure on the quality of decision-making. It seems that the method, in its current version, does not give proper credit to the effect of "running out of time" while making correct, procedure-guided decisions, which seems to be one of the drivers for HFE 1B.

We are looking forward to the comparison of the remaining HFEs in the SGTR. Generally, we are planning to use the insights to a) refine the guidance and b) evaluate the method to see whether additional factors or aspects of the factors need to be included. In this regard, empirical data provide invaluable input. Although an empirical model of performance needs to be based on more than one scenario (two variants in this case), this data contributes to such a model. This should lead to improvements in CESA-Q, as well as in other HRA methods.

3.2.6 Decision Trees + ASEP (NRI)

3.2.6.1 Short Overview of the NRI "Decision Trees + ASEP" Method (by Jaroslav Holy)

The method represents a combination of two well-known HRA principles, the decision tree approach and a modified ASEP approach. The basic idea for the method was originated in the first half of nineties by Gareth Parry during a collaboration between NRI Rez and NUS (later Scientech) in the first NPP Temelin PRA project. The basis for developing the decision tree type quantification method reflects the work of EPRI specialists, and some information about it can be found in [24] and [25]. The ASEP part of the method follows some of the principles presented in [26] and [27], which were significantly elaborated upon and modified in later NRI analyses [10]. Since the method was primarily used as an approximate tool for HRA with a significant lack of information typical of a plant under construction (as NPP Temelin was at that time), the method has to be updated later to be suitable for an HRA update for a plant under operation, taking into consideration specific operation features in WWER reactors.

When using this method, every human error is regarded as having three contributors. The first is from the failure to **d**etect, **d**iagnose, or **d**ecide on a plan of action (the DDD contribution). This is essentially the contribution addressing errors in work with informational and cognitive errors. The second contribution is caused by the delay in starting the manipulations based on previously performed information processing and diagnosis. The last contribution is from the failure to execute the planned action correctly. Different approaches are used to estimate the probabilities of all three phases of human intervention.

For estimation of information processing failure probability, the approach is based on the decomposition of each failure event into the contributions that represent different potential failure mechanisms. For each failure mechanism, a decision tree was constructed, which had as its branches factors that were thought to influence the likelihood of an error as a result of that mechanism. A probability is associated with each path through the decision tree to represent the analyst's assessment of the the factors' combined effect on that path. In those cases, in which more than one path through the decision tree must be included in the estimation, an average value derived from all the probabilities associated with all the paths under concern is used to represent DT contribution to the total HEP value. The probability of an error in the DDD phase for a particular scenario is then estimated by summing the contributions of the appropriate paths from the decision trees that are applicable to the scenario. Five basic failure mechanisms have been used for operators' actions, connected with information processing in recent versions of the method: 1) information not available or hardly available, 2) failure of attention, 3) information misread/miscommunicated, 4) procedure step skipped, and 5) procedure misinterpreted.

As soon as the control room staff have successfully processed the available non-standard status information, it is necessary to finalize the diagnosis act and to decide to perform manipulations with plant equipment. Time was defined as the basic factor influencing the success of this part of intervention. For the quantification of the probability of failure of this part of the action, approximate time reliability curves from THERP [11] are typically used. However, this part of crew interaction with plant equipment is, above all, connected with diagnosis. Since the symptom-based procedures provide the crew with all the necessary diagnosis support, this contribution to the total human failure probability is mostly expected to be negligible when crew actions are driven by this type of procedure.

The quantification model based on elements of ASEP [10] procedure was used as the basic methodical tool in analyzing the manipulative part of human action. Here, the appropriate HEP contributors are derived for the individual human manipulations using the following rules: 1) the probability of execution failure depends on the type of task and on the level of stress, 2) the type of task may be step-by-step (usual, standard), partly dynamic or dynamic, 3) the level of stress may be classified as moderately, increased, or extremely high, 4) a set of additional conditions is given to help in selecting an appropriate level of stress or dynamics attribute, and 5) for each possible combination of stress and task dynamics level, a general numerical value of HEP contributors was derived, which is used directly in quantification.

3.2.6.2 Main References to the "Decision Trees + ASEP" Method

[24] Beare, A.N., Gaddy, C., Singh, A., and Parry, G.W.: "An Approach for Assessment of the Reliability of Cognitive Response for Nuclear Power Plant Operating Crews," Proceedings of Probabilistic Safety Assessment and Management, Beverly Hills, CA., Elsevier, February 1991.

[25] Parry, G.W. et al.: "An Approach to the Analysis of Operating Crew Responses Using Simulator Exercises for Use in PSAs," presented at the OECD/BMU Workshop on special issues of level 1 PSA, Cologne, FRG, May 28, 1991.

[26] Parry, G.W., Holy, J., Kucera, L.: "Human Reliability analysis-Analysis file 6T47AF01," NPP Temelin PSA documentation, Revision 1, March 1996.

[10] Swain, A.D. (1987): "Accident Sequence Evaluation Program Human Reliability-Analysis Procedures," NUREG/CR-4772.

[27] Holy, J., Kucera, L.: "Human Reliability Analysis-Analysis file 6T47AF01," NPP Temelin update PSA documentation, Revision 1, December 2001.

3.2.6.3 NRI, DT + ASEP Analysis of HFE 1A

<u>Summary of Negative Influences Identified by NRI, DT + ASEP</u>

For the Cognitive Part:

Relative complexity of the procedure logic (DT5) covering these activities and length of procedure (DT4) were the most important negative influences. While workload was considered high due to the time constraints of the event, the rest of the outcomes in DT2 (insufficient attention of operator) were all positive, so this was not a contributor to the HEP. Note that outside of the workload (i.e., time stress), time available was not considered a limiting factor for this event because it was thought that good cues would be available in the time frame and that "since the symptom-based procedures provide the crew with all the necessary diagnosis support, this contribution (available time) to the total human failure probability is expected to be negligible when crew actions are driven by this type of procedure."

For the Execution Part:

The stress level was an important contributor for the first action, *closing of steam valve to turbine-driven AFW pump and closing main steamline isolation valve and adjacent bypass valve*, and a type of task (partially dynamic task) for the second action (or second part), *trip of AFW pump if the SG level is higher than 10% of nominal.* The dynamic changes in parameter values and the course of plant response to SGTR in the second part of the action make it partially dynamic.

<u>Comparison to Empirical Data</u>

Only one crew failed to complete the action within the designated time frame, and they were only about a minute and a half late. The main factor in the crews' completion time appeared to be the rate at which they progressed through the procedure, but it was also impacted by the time at which the crew tripped the reactor and entered E-0. The crew that failed performed slowly, but otherwise had many elements of good performance: they were cautious and held meetings, but lacked the sense of urgency that might have helped them to finish in time and avoid the danger of overfilling the steam generator. Thus, aspects of crew-to-crew variability in what could fall into the category of work processes (or some other crew-related category) would seem to be a potential contributor to crew performance. Regardless of their work processes or characteristics, all but one of the crews met the criterion for HFE 1A, so work processes were not identified as a negative driving factor in the data for HFE 1A. In any case, the HRA methods were not expected to be able to address these effects in this study.

The apparently limited effects of the time available in the crews' performance are consistent with the DT-ASEP analysis, which asserted that "time available was not considered a limiting factor for this event because it was thought that good cues would be available in the time frame" and that "since the symptom-based procedures provide the crew with all the necessary diagnosis support, this contribution (available time) to the total human failure probability is expected to be negligible when crew actions are driven by this type of procedure." This appeared to be true, with the exception of one relatively slow crew. Most crews completed the action well within the time available, so this assumption was generally consistent with the results.

The DT-ASEP analysis identified the relative complexity of the procedure logic (Decision Tree (DT) 5) covering these activities and the length of procedure (DT4) as the most important negative influences. These factors appeared to apply mainly to E-3 steps 3 and 4. There was some indication that aspects of step 3 slowed some crews, so execution complexity was assessed as somewhat high. Given that there were time limits (and the DT+ASEP analysis team was aware of the time constraints, as indicated in other parts of their documentation), the analysts focused on the relative complexity and length of the procedures as the only factors likely to cause problems (through delays), which was supported by the results. Despite this, their HEP value for the diagnosis was relatively low, which implies that these factors would not be expected to have strong effects.

However, the analysis also predicted higher-than-normal stress levels associated with "closing of steam valve to turbine-driven AFW pump and closing main steam line isolation valve and adjacent bypass valve." They also thought that the type of task (partially dynamic task) for the second action (or second part), trip of AFW pump if the SG level is higher than 10% of nominal, could contribute to the likelihood of failure during response execution. However, there was no evidence of high stress levels. The resulting HEP predicted for response execution was also relatively low, and it seems unlikely that anyone would argue that this HEP reflects a large effect of these factors.

Assessment of Negative Influences Identified by NRI DT+ASEP

In this type of scenario (base SGTR), it is difficult to evaluate predictions that are weighted such that they are expected to have limited effects on performance. The factors modeled in the DT+ASEP analysis may very well have such effects on some crews, and could contribute to some crews having some problems in execution. Nevertheless, the negative PSFs were not unreasonable and the analysis did not identify any major drivers for failure, which is consistent with the results.

Summary of Positive and Neutral Influences as Identified by NRI DT+ASEP

For the Cognitive Part:

Information availability (good cues and training in DT1) and the quality of the MMI (one-time evaluation of the main display in DT2) positively supported the crews. Time was not seen as a major limiting factor.

For the Execution Part:

The type of task was standard (step by step) for the first action, and nominal stress could be assumed for the second.

Comparison to Empirical Data

The identification of information availability as a positive influence is consistent with the results. It is difficult to assess the impact of the MMI, but in general it seemed to support performance. In following the decision trees, most factors were given the "better" choices, except for the negative factors noted above. Time was not seen as a limiting factor, which was generally consistent with the results, but the good procedural support was not explicitly identified. The selection of the generally positive choices in the DT's is consistent with the results.

Assessment of Positive and Neutral Influences Identified by NRI DT+ASEP

See comparison above.

Summary of Operational Description Provided in the NRI DT+ASEP Analysis

There was sufficient information at the crew's disposal in the control room. Since the success of the action is defined as "to identify and isolate failed SG within 20 minutes," not all alarm-type indications are fast enough to be relevant, but at least two important alarms would be available in a timely manner. At the base of detailed simulation, the secondary circuit radiation alarm is the first signal of SGTR potential, supported by a significant drop of pressurizer level alarm, which follows it very quickly (within one minute). The remaining alarms come later—damaged SG abnormal level in 18 minutes, pressure drop in pressurizer even later, so that they cannot be effectively used by control room crew. Still, two clear, strong alarms can be seen as sufficient alarm support for the action. In general, there was good separation of important signal information from the background; however, the list of substeps in E-3 step 3, which represents the base of the actions quantified with this HEP, is relatively long. In addition, E-3 step 4 is a kind of floating step; the logic of this type of step usually causes problems. Still other steps may require several actions, adding to the complexity. In general, the complexity of the procedure might hinder the crews.

Assessment of Operational Description

While it is difficult to verify all aspects of the operational description (particularly the negative factors), the generally positive description in terms of success is consistent with Halden's operational summary.

3.2.6.4 NRI, DT + ASEP Analysis of HFE 1B

Summary of Negative Influences Identified by NRI DT+ASEP

For the Cognitive Part:

The fact that necessary accurate information is not completely available (ambiguity of symptoms) and the probable lack of training on this combination of events (DT1) are negative influences. High workload, along with some missing alarms and the need to monitor parameters (DT2), also contributes. The other strongest negative influences are the relatively long and complex procedures (DTs 4 and 5).

For the Execution Part:

The stress level and partially dynamic aspects caused by the ambiguity of symptoms and external conditions were important contributors for the first action, *closing of steam valve to turbine driven AFW pump and closing main steam line isolation valve and adjacent bypass valve*, and types of tasks (partially dynamic task) for the second action (or second part), *trip of AFW pump if the SG level is higher than 10% of nominal.* The dynamic changes in parameter values and the course of plant response to SGTR in the second part of the action make it partially dynamic.

Comparison to Empirical Data

The NRI DT+ASEP analysis identified the ambiguity of the symptoms and the probable lack of training on this combination of events (DT1) as negative influences. This is consistent with the results of the Halden summary of crew performance. Both factors were thought to be important. Additionally, the DT+ASEP analysis assumed that the complexity and length of E-3 (particularly steps 3 and 4) could significantly affect the crews. The Halden summary also identified procedures as a problem, but it was the mismatch of E-0 with the available information that was thought to create the main problem. E-0 was not generally effective in providing a means to get the crews to E-3, and ES-1.1 did not work particularly well either; however, there was some indication that aspects of step 3 slowed some crews, as predicted by the NRI DT+ASEP analysis. The NRI analysis also suggested that there were some relatively complex steps in E-0 (there is a combination of AND and OR logic (step 21 in E-0 procedure)), and that the SI termination status checking represents a relatively complex step with non-negligible potential for error of commission occurrence.

The Halden crew summaries also identified scenario complexity, procedures, training, work processes/team dynamics, and execution complexity as factors bearing negatively on performance. As noted above, the NRI DT+ASEP touched on most of the factors. The effects of factors like team dynamics and work processes varied among the crews, with some having characteristics that facilitated performance and others having characteristics that hindered performance (though the HRA teams would have no basis for predicting this).

In addition, the summary of the driving PSFs for the operating crews noted that the time allowed to complete the action in this study constrained the likelihood of success, that is, it was likely to decrease the success rate. While the NRI DT+ASEP analysis acknowledged that the time frames for success of this HFE tended to be short (and not very realistic), they did not appear to directly use this factor in evaluating performance. This decision was inconsistent with the results.

The DT+ASEP analysis also identified high workload, along with some missing alarms and the need to monitor parameters (DT2), such as RCS pressure, as making contributions. While not directly identified in the Halden summary as having strong effects, these factors may have affected performance at least somewhat.

However, the analysis also predicted higher-than-normal stress levels associated with "closing of steam valve to turbine-driven AFW pump and closing main steamline isolation valve and adjacent bypass valve." The analysts thought that the type of task (partially dynamic task) for the second action (or second part), *trip of AFW pump if the SG level is higher than 10% of nominal*, could contribute to the likelihood of failure during response execution. However, there was no evidence of high stress levels. The resulting HEP predicted for response execution was also relatively low, and it seems unlikely that anyone would argue that this HEP reflects a large effect from stress or the problems with E-3, which was consistent with the results.

<u>*Assessment of Negative Influences Identified by NRI DT+ASEP*</u>

The analysis identified several of the most important negative drivers of performance. While the analysts did not explicitly discuss the problems the crews would have with E-0, this could be inferred from their noting the ambiguity of the symptoms (obviously relative to the procedures) and their probable lack of training on this combination of events. However, they did not appear to account for the impact of those factors' effects in terms of how long it would take the crews to diagnose the situation and complete the response. Thus, they may have underestimated the problems the crews would have with E-0 and ES1.1, in that seven crews failed to meet the time criterion. On the other hand, all crews were eventually successful.

<u>*Summary of Positive and Neutral Influences Identified by NRI DT+ASEP*</u>

For the Cognitive Part:

Good alternate sources of information (DT1) and the quality of the MMI (alternate information on main display with good separation (DTs 2 and 3)) positively supported the crews.

For the Execution Part:

Nominal stress could be assumed for the second action. In addition, the automatic closing of the main steam isolation valve makes the manipulations less complex than the base case.

Comparison to Empirical Data

The availability of alternate sources of information was identified as a positive influence by NRI DT+ASEP. The alternative indications (e.g., SG level) were identified as a secondary positive influence in the data analysis, since the indications were more generally negative, but the NRI DT+ASEP prediction was consistent with these results. However, the DT+ASEP analysis did not list the crews' general training in SGTR (secondary positive influence) as clear support for the knowledge-based decision that most crews ultimately relied on, making it a positive influence.

Assessment of Positive and Neutral Influences Identified by NRI DT+ASEP

The NRI DT+ASEP analysis recognized the strength of the SG-level cue (availability of information), but did not explicitly note the crews' training/knowledge base as an important positive influence (secondary positive influence in the data analysis). The crews' general training on SGTR seemed to eventually counter the insufficient training for this specific scenario, thus becoming an important factor in the crews' success. On the other hand, the analysts' use of path g in DT5 suggests they believed that the crews' training on the procedures would help with the outcome. This particular factor may have been somewhat difficult to represent within the DT approach, given that the lack of specific training was selected through path c of DT1.

Summary of Operational Description Provided in the NRI DT+ASEP Analysis

Due to the masking effect of the steam line break, the necessary accurate information is not completely at the crews' disposal in control room, but they have alternate information sources, which may lead them to the right conclusion and response. The missing direct alarms hinder the crews' ability to diagnose, and they will need to monitor the dynamic changes in RCS parameters in order to be successful. However, the information on the large screen display and the available alarms will help them. The list of sub-steps in E-3 step 3, which represents the base of the actions quantified with this HEP, is relatively long, and complexity is added by the combination of AND and OR logic in E-0 step 21.

Assessment of Operational Description

While the analysts did not explicitly discuss the problems the crews would have with E-0, it can be inferred to some extent from their discussion. Thus, much of the NRI DT+ASEP operational description is consistent with the results. Dynamic changes in RCS parameters probably contributed to the diagnosis, as they said, but the divergence of SG level appeared to be the main diagnostic parameter. There was indication that aspects of step 3 slowed some crews (execution complexity somewhat high), and, as noted above, the NRI analysis also suggested that there were some relatively complex steps in E-0 (there is a combination of AND and OR logic (E-0 step 21), and that the SI termination status checking represents a relatively complex step with non-negligible potential for error of commission occurrence. There was also evidence that the SI termination status checking represents a relatively complex step.

3.2.6.5 Additional Comments on NRI DT+ASEP Analysis

Comments on NRI DT+ASEP Analysis of HFE 1A and HFE 1B

For the most part, the method seemed to provide a reasonable set of factors (as represented by the decision trees and the paths through them) to represent the scenario being analyzed. However, it may not always be easy to determine how to capture the range of factors that might be relevant (e.g., more general vs. specific training, and, in this scenario, the effect of available time as a constraint on success). The decision tree approach has not traditionally been used to address time limited scenarios, but the time criterion, based on what was expected of the crews rather than on what may be the usual accepted time in such a scenario, may have made it difficult to use another approach, such as a simple TRC. Particularly complex scenarios, such as the one modeled for this analysis (HFE 1B), could make the decision tree approach somewhat cumbersome, and it may need improvements before it is used for the more complex scenarios.

Additionally, as with all HRA methods, decisions about whether particular factors will or could influence performance is based on the analysts' opinion. Whether a procedure is too long or too complex is often tied to other factors, such as training and the crews' knowledge base. Thus, such decisions are not always simple, and this creates the opportunity for analyst-to-analyst variability. Additional guidance for considering the relationships between factors may improve methods like the decision tree approach.

Insights for Error Reduction

Along with a good task analysis, the decision trees included in the DT+ASEP method should allow insights into improving safety; that is, the method examines aspects that, when identified as problematic, could be improved to facilitate error reduction. However, this will depend heavily on the judgments made about the different potential causes of failure. The analysis of this event suggested that training on this specific event might be improved, but did not explicitly identify the need to improve the procedures to support transfer to E-3 in HFE 1B. The analysis of HFE 1A suggested that improvements to procedure E-3 might be warranted.

Impact on HEP (Sensitivity to Driving Factors)

For HFE 1A, since there were no driving negative factors, the relatively low HEP seemed to reflect the analysis.

However, based on the analysis of HFE 1B, it appears that, with the use of the decision tree, identifying the relative weights of the contributors to the HEP is not always straightforward, since several factors are considered in arriving at the HEP for a tree. Despite this, the contribution of the overall failure mechanism can be evaluated based on the HEP obtained for each tree.

Guidance and Traceability

As with most HRA methods, there is room for improving method guidance. However, the guidance provided and the documentation of the method application and derivation of the HEPs was very traceable.

3.2.6.6 NRI DT+ASEP Team Comments on the Original Comparison

No comments were provided.

3.2.7 MERMOS (EDF)

3.2.7.1 Short Overview of the MERMOS Method (by Pierre Le Bot)

Principle of the Description of Failure with Methode d'Evaluation de la Realisacion des Missions Operateur la Sureté (MERMOS)

The reference to succeed is the requirements. Failure occurs when the requirements are not met, and is described through different possible explanatory "MERMOS scenarios" (« operational stories »).

The probability of failure of the mission is then (separated scenarios):

P (failure of a human reliability task) = $\Sigma_{\text{identifiable scenarios}}$ P (identifiable failure scenario) + Pr

Pr = Residual Probability: between 10-4 and 10-5 (reflects uncertainty).

Detection of the MERMOS scenarios (and learning of their exhaustiveness): Three main requisite functions for the "operating system"[11] (MERMOS systemic approach) that can fail: Strategy, Action, and Diagnosis. For example, with strategy, the analysts have to identify competing objectives that could become lesser priorities.

Hypothesis: We consider failure of one function at a time (to avoid counting a failure several times, we separate the scenarios).

Structuring of the scenarios: situation features (SF), important configurations of accident operation (CICA[12]) given the SF, non-reconfiguration during Tmission:

$$P(\text{failure scenario 1}) = P_{\text{non reconf/CICAs}} \times P_{\text{CICAs/SITU}} \times P_{\text{SITU}}$$

Where:

PSITU is the probability of the simultaneous presence of the situation features participating in the appearance of CICAs.

$P_{\text{CICAs/SITU}}$ is the probability of simultaneous existence of the CICAs, knowing that the characteristics of the corresponding situation[13] are present.

$P_{\text{non reconf/CICAs}}$ is the probability of the scenario appearing, knowing that the corresponding CICAs are met, or even the probability that the CICAs are maintained long enough to lead to failure of the task (non-reconfiguration probability).

---

[11] Operating system: crew + procedures + HMI (distributed cognition)

[12] CICA: Important Configurations of Accident Operation (describing a way of operating); failure occurs when a CICA is not adapted to a specific situation and when the operating system does not reconfigure itself on time.

[13] For a given task, the structural characteristics are fixed and therefore have a probability of existence equal to one, whatever the context. On the other hand, the contextual characteristics of the situation can vary. The possibility of CICAs appearing is consequently related only to the presence of the necessary contextual characteristics for the situation.

Note: the probabilities are conditional. They are to be determined according to discrete values if possible, so as to increase understanding of the analysis and to ensure a certain robustness and reproducibility of the use of these judgments. These values are defined with the following understanding:

- very improbable: 0.01
- improbable: 0.1
- probable: 0.3
- very probable: 0.9

3.2.7.2 Main References to the MERMOS Method

[28] C. Bieder, E. Desmares, F. Cara, P. Le Bot, J.L. Bonnet: "MERMOS: New Issues on Human Failure in Emergency Operations of Nuclear Power Plants," Probabilistic Safety Assessment and Management, PSAM4, New York, USA, September 13-18, 1998.

[29] C. Bieder, P. Le Bot, E. Desmares, F. Cara, J.L. Bonnet: "MERMOS: EDF's New Advanced HRA Method," Probabilistic Safety Assessment and Management, PSAM4, New York, USA, September 13-18, 1998.

[30] P. Le Bot, C.Bieder, F. Cara: "MERMOS, a second generation HRA method : what it does and doesn't do," International Topical Meeting on Probabilistic Safety Assessment, PSA'99, Washington, D.C., USA, August 22–26, 1999.

[31] C.Bieder, P. Le Bot, F. Cara: "What does a MERMOS analysis consist in?", International Topical Meeting on Probabilistic Safety Assessment, PSA'99, Washington, D.C., USA, August 22–26, 1999.

[32] F. Cara, C.Bieder, P. Le Bot: "What are the CICAs? Retrospective analysis of the Three Mile Island accident from the MERMOS viewpoint," International Topical Meeting on Probabilistic Safety Assessment, PSA'99, Washington, D.C., USA, August 22–26, 1999.

[33] P. Le Bot, C. Bieder, S. Vidal: "Feedback from the actual implementation of the MERMOS method," Probabilistic Safety Assessment and Management, PSAM5, Osaka, Japan, November 27-December 1, 2000.

[34] P. Le Bot, H. Pesme, F. Ruiz: "Methodological Validation of MERMOS through 160 analyses," Probabilistic Safety Assessment, PSA'02, Detroit, MI, USA, October 6–9, 2002.

[35] P. Le Bot: "Human reliability data, human error and accident models—illustration through the Three Mile Island accident analysis," *Reliability Engineering & System Safety*, Volume 83, Issue 2, February 2004, pages 153-167.

[36] P. Le Bot, A. Voicu: "Exploratory study of the impact of organisational factors on safety using Probabilistic Safety Assessments," presentation only, at Probabilistic Safety Assessment, PSA'05, San Fransisco, CA, USA, September 11-15, 2005.

[37] P. Le Bot: "Using expert judgement with MERMOS: from static assessment towards knowledge capitalisation," Seminar of the European Commission, The Use of Expert Judgment in Decision Making, Aix-en-Provence, France, 2005.

[38] P. Meyer, P. Le Bot, H. Pesme: "MERMOS: an extended second generation HRA method," Proc. 2007 IEEE 8th Human Factors and Power Plants (HFPP) and 13th Annual HPRCT Meeting, August 26-31, 2007, Monterey, CA, USA.

[39] H. Pesme, P. Le Bot, P. Meyer: "Little stories to explain Human Reliability Assessment: a practical approach of the MERMOS method," Proc. 2007 IEEE 8th Human Factors and Power Plants (HFPP) and 13th Annual HPRCT Meeting, August 26-31, 2007, Monterey, CA, USA.

[40] P. Le Bot, P. Meyer, H. Pesme: "The CICA concept for use in the MERMOS method redefined by a new organizational reliability model," Proc. 2007 IEEE 8th Human Factors and Power Plants (HFPP) and 13th Annual HPRCT Meeting, August 26-31, 2007, Monterey, CA, USA.

[41] Pierre Le Bot: "Can nuclear safety engineering deal with the human and organizational aspects of error?", Symposium INSS Human factors, Tsuruga/Tokyo, Japan, 2007.

3.2.7.3 EDF, MERMOS Analysis of HFE 1A

In an application of MERMOS, multiple scenarios are usually identified as contributors to the failure of the HFE. The assessors' summary of the influences, negative and positive, is based on selecting the dominant scenario (or scenarios) among these. By contrast, in reporting the influencing factors, the HRA team considered all scenarios that were identified in the MERMOS analysis, including scenarios with a small contribution to the HEP.

The dominant MERMOS scenario for HFE 1A is Scenario #1. This scenario contributes 88% of HEP, with eight other scenarios making up the rest[14]. It is described below, under "Summary of operational description in the Electricité de France (EDF) MERMOS analysis."

Scenario #1 takes place when "the system does not perform the procedural steps fast enough, and does not reach the step of the isolation of the ruptured SG within the allotted time (failure by no strategy)."

Summary of Negative Influences Identified by EDF, MERMOS Analysis

Factors linked to the existence of scenario #1: "Priorities prescriptions"–the priorities defined by the prescriptions (procedures and training) will require the operators to perform other actions before identifying and isolating the ruptured SG, at least during the HFE time window.

Factors related to occurrence of HFE, given the existence of scenario #1: "Training–safety culture," "Allotted time to act," "It is very probable that the operators follow the instructions step-

[14]Each MERMOS scenario has a conditional probability, given the PRA context. The sum of the conditional probabilities of the MERMOS scenarios yields the HEP.

by-step." This last factor, a step-by-step following of procedures, is identified as negative due to its impact on the time to get to the appropriate procedure steps.

Expressed in Form B terminology, the factor that is identified as a negative influence is "available time," referring to an inappropriate balance between the available and the required time. It should be emphasized that, at least in the dominant scenario, the analysis does not predict that performance is being challenged by a shortage of time or by stress arising from a perception of time pressure. On the contrary, the lack of time pressure or of a sense of urgency is explicitly mentioned as a negative factor, negative in the sense that the analysis predicts that the HFE failure will occur because there are no strong imperatives for responding within the time window defined for the HFE.

Comparison to Empirical Data

The main negative factors predicted in the submittal are consistent with the empirical data. The lack of time due to the time window being used for actions prescribed by procedures (but not required for HFE 1A success) and the lack of a sense of urgency or of imperatives to respond within the time window defined for the HFE are both present in the data.

This description corresponds well to the performance of the slowest crew, which completed the isolation after the time window assumed for HFE 1A. This performance was characterized as generally good and correct, but slow and marked by a lack of urgency. Their application of the procedures is described as thorough and controlled but slow. As a result of these factors being taken together, the crew completed the task late.

Assessment of Negative Influences Identified by EDF, MERMOS

Overall, the analysis predicted fairly good performance for HFE 1A (the HEP can be interpreted as an expectation of not observing any failures, with an upper bound of 1.3 crews) and no major negative factors. If HFE failure were to occur, it would primarily result from following the procedures step by step, without sufficient urgency.

The HRA team's analysis of the negative factors also matches the empirical data in the sense that, in the data, no significant negative factors were identified for HFE 1A. No notably specific difficulties were observed.

Summary of Positive Influences Identified by EDF, MERMOS Analysis

The positive influences were identified as:

- easy to diagnose transient, trained (crews would be expected to know the main actions to take without consulting the instructions)
- communication mode that supports error detection and correction
- shift supervisor redundancy

The first factor, ease of diagnosis, is predicted as a positive influence specifically for HFE 1A, whereas the second and third factors are identified as general positive factors that apply to both HFE 1A and 1B.

Comparison to Empirical Data

The positive factors identified, "easy to diagnose" and "trained" transient, are consistent with the Training and Experience–good to very good, HMI, and indication of conditions–very good.

Assessment of Positive Influences Identified by EDF, MERMOS

The main positive performance driver is indeed the match between the crews' training and the scenario.

Summary of Operational Description Provided in the EDF MERMOS Analysis (HFE 1A)

The assessment of predicted operational expressions focuses on the dominant MERMOS scenario (88% of HEP, with eight other scenarios making up the remaining 12%), which is scenario #1.

Description of HFE 1A MERMOS Scenario #1

The perceived difficulty faced by the crews is that, since the "required" time (success criterion) is defined "independently of functional operating objectives," actions taken in accordance with HAMMLAB or operating criteria may easily be delayed. The crews will not have an impression of urgency or of a need to catch up with respect to the defined time window.

Operational expression (from scenario 1): The "operating system" does not perform the procedural steps fast enough and does not reach the step of isolating the ruptured SG within the allotted time. It performs this way because it follows training and procedures and at the same time does not perceive any functional operating objectives that would motivate a faster response.

Operational expression (from scenario 1): Crew does not give (sufficient) priority to isolation of the ruptured SG. This occurs because there is an "absence of priority and acceleration of operation in the event of delay."

Operational expression: A late manual shutdown is assigned a probability of 0.1.

A separate operational expression (associated with HFE 1A Scenario #3, contributing about 8% of the HEP, and a distant second compared to Scenario #1 with 88%) relates to local actions. The isolation may be delayed while waiting for feedback from local actions. Local actions are not required for the execution of the isolation. However, other local actions may be optional or required (or prescribed) to confirm the SGTR diagnosis. If the crew waits for feedback from such actions to proceed, this increases their probability of not completing isolation within the time window (for instance, the action mentioned in E-3 step 3, "locally verify steam traps isolation valves from ruptured SG").

Assessment of Operational Description (HFE 1A)

The main operational expressions predicted for HFE 1A suggest a performance without specific difficulties in terms of situation assessment, decision making, or execution, and in which the operators follow the procedures routinely, without a strong sense of urgency with respect to the 20-minute time window that has been defined for HFE 1A success.

There is a good match between the predicted operational description of the failure of HFE 1A and the empirical data for HFE 1A. No major difficulties were observed, and there are no strongly negative factors identified for this HFE from the data. In addition, a lacking sense of urgency aptly describes the performance of the crew that took the most time to complete the isolation in this scenario.

It is worth noting that the analysis specifically identifies the local actions in E-3 step 3 as a potential cause of delays. This is reflected in the empirical data in the rating "somewhat high" for "execution complexity," which relates to the need to manage a mix of local and control room actions in this step.

3.2.7.4 EDF, MERMOS Analysis of HFE 1B

The dominant MERMOS scenario[15] for HFE 1B is scenario #1. This scenario contributes 89% of HEP, with seven other scenarios making up the remainder. Scenario #1 occurs when "the system does not perform the procedural steps fast enough and does not reach the step of the isolation of the ruptured SG within the allotted time (failure by no strategy)." It is described in more detail below, under "Summary of operational expressions in the EDF MERMOS analysis."

Summary of Negative Influences Identified by EDF, MERMOS Analysis

The negative influences identified for HFE 1B are the same factors as for HFE 1A, the same HFE in the base case scenario. Additional negative factors identified for HFE 1B are "training and experience" and "adequacy of time." The crews may take some time to check the criterion "uncontrollably rising SG level," due to their lack of experience with this task and to the judgment required to interpret this criterion, which negatively impacts the adequacy of the time window.

For reference, the negative influences described for HFE 1A that apply for HFE 1B are repeated here:

- Factors linked to existence of scenario #1: "Priorities prescriptions"–the priorities defined by the prescriptions (procedures and training) will require the operators to perform other actions before identifying and isolating the ruptured SG, at least during the HFE time window.
- Factors related to the occurrence of HFE, given existence of scenario #1: "Training–safety culture," "Allotted time to act," "It is very probable that the operators follow the

[15]Each MERMOS scenario has a conditional probability, given the PRA context. The sum of the conditional probabilities of the MERMOS scenarios yields the HEP.

instructions step-by-step." This last factor, step-by-step following of the procedures, is identified as negative due to its impact on the time to get to the appropriate procedure steps.

- Expressed in Form B terminology, the factor that is identified as a negative influence is "available time," referring to an inappropriate balance between the available and the required time. It should be emphasized that, at least in the dominant scenario, the analysis does not predict that performance is being challenged by a shortage of time or by stress arising from a perception of time pressure. On the contrary, the lack of time pressure or of a sense of urgency is explicitly mentioned as a negative factor, negative in the sense that the analysis predicts that the HFE failure will occur because there are no strong imperatives for responding within the time window defined for the HFE.

Comparison to Empirical Data

For HFE 1B, the empirical data in the complex SGTR scenario point to difficulties that reflect specific mismatches between the scenario and the procedures, and between the scenario and the training and experience of the operators. These difficulties are represented by identifying the factors scenario complexity and procedural guidance as negative drivers. In addition, training, the indications of plant conditions, execution complexity, and adequacy of time are rated negatively.

Assessment of Negative Influences Identified by EDF, MERMOS

There is a good match between the negative influences identified in the analysis and the negative drivers observed in the empirical data:

- Scenario complexity and indications of plant conditions, due to the lack of the radiation indications, combined with the reliance of these procedures on these indications.
- Training and experience, due to the lack of training on the use of alternative indications for diagnosis SGTR.
- Adequacy of time, which becomes problematic due to the combination of the above.

However, the PSFs associated in the empirical data with the inadequate procedural guidance for transferring to E-3 in the absence of radiation indications are missed (Procedural Guidance-poor). [See also the MERMOS team's point of view §3.2.7.6].

Summary of Positive Influences Identified by EDF, MERMOS Analysis

The analysis identified the following influences as positive. The first is specific to HFE 1B, whereas communication and shift supervisor redundancy were general positive influences that also applied to HFE 1A:

- strict adherence to instructions and the short intervals between SG levels helps the mission to be completed within the required time frame;
- communication mode that supports error detection, correction; and

- shift supervisor redundancy.

Comparison to Empirical Data

The empirical data shows a high level of adherence to the procedural guidance. However, because the procedural guidance relies on the radiation indications as the main indication of SGTR, this adherence does not tend to support completion of the isolation in time, as the crews had to use time to find procedural support for a transfer into E-3, after having identified the SGTR based on alternative indications.

In general, communications supported error detection and correction. In the crew performances for both HFE 1A and 1B, the few slips that did occur were subsequently recovered through communication, as well as through the redundancy provided by the shift supervisor.

Assessment of Positive Influences Identified by EDF, MERMOS

The general positive influences, the communication mode that supports error detection and correction and the redundancy provided by the shift supervisor, were observed in the empirical data. The successful crew performances show these factors are positive influences.

The positive influence of "strict adherence to instructions" was not supported by the data. Because the procedural guidance is poor for this scenario, following the procedures closely and strictly does not support a timely response.

Summary of Operational Description Provided in the EDF, MERMOS Analysis (HFE 1B)

The assessment of predicted operational expressions focuses on the dominant MERMOS scenario (89% of HEP, with seven other scenarios making up the rest), which is scenario 1. For HFE 1B, three other operational expressions were identified (each scenario contributing about 3% to the HEP). These are discussed further below.

Operational expression: The system does not perform the procedural steps fast enough and does not reach the isolation step within the allotted time (Scen. No.1, dominant)

Operational expression: "Identification of the SGTR by checking steam generator levels can cause problems or time wastage."

Operational expression: "The absence of radioactivity does not facilitate diagnosis or enable other hypotheses to be developed for the event in progress."

Operational expression: assisting reactor operator (ARO) takes time to check that the level in SG1 is rising uncontrollably. This is probable (assigned p=0.3). The ARO will not be fast because this check is not often included in SGTR training scenarios, which rely more strongly on other cues.

Three other operational expressions were identified (each scenario contributing about 3% to the HEP):

- Waiting for feedback from local actions leads to delays (isolation not completed in time window).
- The crew identifies a steamline break and fails to identify the SGTR as an additional fault. As a result, it proceeds to FR-H.5. In this scenario, the crew assumes that the steamline break is the only fault and does not refine the diagnosis or search for complications.
- In trying to control SG levels, the ARO encounters difficulties because it is not an easy operation during an SGTR. Having perceived that there are no radiation indications, the ARO does not communicate to his fellow crew concerning the problems he is having.

Assessment of Operational Description (HFE 1B)

The following table shows that the operational expression predicted in the MERMOS analysis for HFE 1B matches the empirical data quite well. It reflects the overall operational difficulty in the scenario and the mode of the failure: delay in some steps due to lack of training, delay while following the procedures, both of which lead to a delayed isolation of the SGTR.

One of the contributing causes in the empirical data that the prediction misses is the difficulty in finding a transfer step within the procedural framework, if the SG level mismatch (or uncontrollably rising level) is detected by the operators. (See also the MERMOS team point of view §3.2.7.6).

Table 3-1. MERMOS Operational Expressions (Dominant Scenario)

Operational expressions predicted in MERMOS submittal (HFE 1B dominant scenario)	Observed operational expressions
The system does not perform the procedural steps fast enough and does not reach the isolation step within the allotted time (Scen. No.1, dominant).	This overall performance summary matches the empirical data. The performance difficulties in this scenario are indeed related to proceeding quickly enough through the procedure (as opposed, for instance, to a failure to assess the scenario correctly). A major source of delay while proceeding through the procedures is due to the procedures not leading directly to the E-3 transfer, given the indications available in this scenario.
"The absence of radioactivity does not facilitate diagnosis or enable other hypotheses to be developed for the event in progress.	This matches the empirical data, and is, in fact, a dominant operational issue (when combined with the procedural guidance's reliance on this indication and an apparent lack of training on the alternative cues).
"Identification of the SGTR by checking steam generator levels can cause problems or time wastage."	This operational expression matches the empirical data, for instance, *"Crew cannot explain SG1 level without radiation and lack of level in PRZ. Long discussions and unstructured meetings."*
ARO takes time to check that the level in SG1 is rising uncontrollably. This is probable (assigned p=0.3). The ARO will not be fast because this check is not often included in SGTR training scenarios, which rely more strongly on other cues.	This relates to the previous operational expression (it provides a basis for the problems or time wastage). It also matches the empirical data, which has evidence supporting the lack of training on checking of alternative cues for SGTR.

Table 3-2. MERMOS Operational Expressions (Non-Dominant Scenarios)

Operational Expressions Predicted in MERMOS Submittal (Non-Dominant Scenarios)	Observed Operational Expressions
-Waiting for feedback from local actions leads to delays (isolation not completed in time window)	This refers to E-3 step 3. The evidence shows that the crews need a fair amount of time to complete this step, due to the local actions mixed in with the CR actions.
-Crew identifies steam line break and fails to identify the SGTR as an additional fault. As a result, it proceeds to FR-H.5. In this scenario, the crew assumes that the steamline break is the only fault and does not refine the diagnosis or search for complications.	This was not observed.
-In trying to control SG levels, the ARO encounters difficulties because it is not an easy operation during an SGTR. Having perceived that there are no radiation indications, the ARO does not communicate to his fellow crew concerning the problems he is having.	This was not observed. Note that this expression is not predicted to be likely.

3.2.7.5 Additional Comments on the EDF MERMOS Analysis

Comments on EDF MERMOS Analysis

The HRA team's characterization of the main driving factors reflects what they generally expect to see. The dominant scenario reflects the main way in which the team would expect the crews to fail, if these were to fail. Consequently, the positive factors may be seen in some or all of the MERMOS scenarios. The negative factors are most often associated with individual scenarios, but in some cases contribute to several scenarios.

There is a very good match between the qualitative analysis and the quantification in MERMOS, in part because the method allows each scenario to be modeled without requiring a translation into another model. (The analysts do not have to look for a way to express their findings in the terms of the method.)

Secondly, the qualitative analysis has a clear structure, with a clear search strategy. The qualitative analysis approach is MERMOS-specific; its structure is independent of the form B taxonomy of PSFs used in this study. By listing both the retained and non-significant scenarios, it is clear what elements of the search led to the retained scenarios as well as the scenario elements that made the excluded scenarios non-significant.

It is worth noting, especially in the HFE 1B scenarios that MERMOS scenarios seem to be very close to the "operational expression" referred to in Form A, in other words, the MERMOS scenarios explain in specific terms how the failure occurs. Whereas the operational expression must be deduced from the HEP components in many methods (e.g., no diagnosis, late diagnosis, execution), they are explicit in MERMOS.

With regard to the comparison between the analysis and empirical data in terms of driving factors, it should be stressed that the concept of PSFs is not inherent to the MERMOS method. The PSF concept is based on a general model of overall interactions among positive and negative factors. The MERMOS analysis is focused on the identification of specific failure scenarios. The driving factors identified in the submittal resulted from an analysis of the failure scenarios by the MERMOS HRA team, after these scenarios had been identified.

Insights for Error Reduction

The specificity of the MERMOS scenarios and the method's focus on operational expressions and their contributors tend to produce explicit insights for error reduction, that is, these insights can be used to identify specific error-reduction measures. In the case of HFE 1A, the analysis suggests that if the assumed 20 minute time window is critical to a successful response to SGTR, the operators would have to be made more explicitly aware of this time criterion. In the case of HFE 1B, these insights concern the procedural steps and tasks required to address SGTRs that lack radiation indications, which were confirmed by the evidence to be a dominant issue, and the checking of uncontrollably rising SG level, which is observed to be the case as well (although not a strong factor).

Impact on HEP (Sensitivity to Driving Factors)

The sensitivity of HEPs to driving factors in MERMOS analyses directly reflects the model of how the factors interact to result in the HFE (mission) failure scenario. In other words, the sensitivity to a given factor is not a constant characteristic of the method. The sensitivity depends instead on the scenario and on the interaction of a given factor with the other factors present, as reflected in the failure scenario (scenario for the failure of the HFE). As the HRA team notes, "those MERMOS operational stories are a way to express how different factors can combine to lead to failure (in opposition to considerations on PSFs without taking into account their combinations)"–in other words, as opposed to analyses of PSFs that do not account for their combined effects.

Guidance and Traceability

The analysis is structured systematically. The documentation of the submittal suggests that explicit guidance for the qualitative analysis is provided and combined with the analysts' expertise concerning operations. The qualitative analysis and quantitative analyses are traceable. A notable exception is the estimation of the factor probabilities (probabilities of certain performance characteristics, etc); this estimation is based on expert judgment. The guidance for this expert judgment consists of providing the experts with a set of simple values (see §3.2.7.1) and a process for expert judgment (that is, to have three experts first estimating independently, then sharing arguments and finally choosing a consensual value, or a majority value).

3.2.7.6 EDF MERMOS Team Comments on the Original Comparison

HFE 1B: We did not find a problem with transferring to E-3; we are not convinced yet.

Theoretically, as mentioned in the package, the procedures feature several paths to enter E-3. From our experience with French teams, we felt that the operators should have no difficulty entering E3 if they followed the procedure strictly, without using their knowledge to transfer directly to E3 from step 19. Because of this, we imagined that the main reason for failure was an overly strict following of the procedures, which would take up too much time. In actual fact, the opposite has been observed: Halden's teams transferred directly from step 19 to E3 by communicating and coordinating. Only a good knowledge of the Halden teams' habits could have alerted us and predicted that difference. We can explain the difference by the fact that current, state-based French procedures are designed to provide the operators with a solution in any case. Then the operators trust the procedure and adhere to it. We cannot agree with the statement that inadequate procedure guidance is an important PSF for HFE 1B; the operators' distrust in the procedures may be a more likely problem, though this has yet to be analyzed.

MERMOS Team Comments on the Comparison Methodology

It is not sufficient to compare the MERMOS operational expressions with the observed operational expressions of the failure teams: it's better to observe the MERMOS operational expressions in any of the teams' performances, successful or not, and then to check that the combination of the items in the MERMOS scenarios leads to failure, even if the whole MERMOS failure scenario has not been observed.

MERMOS is not focused on general micro individual performance (including success) prediction, but on macro collective and systemic failure at a safety mission level. However, systemic failures occur only in very specific contexts, which include some "operational expressions" that we can observe on simulator, as this comparison demonstrates.

MERMOS Team Comments on the Draft Report as a Whole

Despite the fact that the study is centered on PSFs and that a comparison with simulator results could be irrelevant, we appreciate the philosophy of the pilot study:

- Priority to comprehending the method and qualitative aspects.
- Respect for methods and analysts.
- Caution for benchmark.

3.2.8 HEART (Ringhals)

3.2.8.1 Short Overview of the HEART Method (by Steve Collier)

The HEART (Human Error Assessment and Reduction Technique) assessment team (HAT) consisted of two plant experts and an HRA advisor. The assessments were initially made by the plant experts. The results were reviewed separately by each team member, then discussed and altered to achieve consensus in a series of three-way telephone conversations and email exchanges. This was not an ideal process, but the team expects that it will meet the needs of the pilot study.

The HAT used the method described by Williams in [42]. The team was not aware of any publicly available or "official" HEART manual, so the assessment does not use the various other papers, comments, guidance, or derivative versions of HEART and its use (e.g., NARA, a further development of HEART), or company-confidential documents.

The steps used for the assessment were:

Step 1: Select a generic task-type

Step 2: Select EPCs ("Error-Producing Conditions") and assess proportions of effect

Step 3: Calculate HEP

Step 4: Consider dependencies

Step 5: Document HEP

These steps were completed for each HFE to be assessed. They are described in turn in the subsections below.

Step 1: Select a generic task-type

A generic task-type (GTT) was selected, based upon the task information supplied. Although the choice of GTT is an important matter of judgment, it was not subjected to peer review outside the HAT. In an HRA assessment, it would normally be a matter of judgment as to how far a task is decomposed. The HAT had a strong preference to retain the HFEs as described in the supplied information, and there was no occasion where the HAT felt it necessary to decompose the tasks further.

Step 2: Select EPCs and assess proportions of effect

"Error-producing conditions" (EPCs) (EPC is the HEART name for what other techniques might call a negative PSF) were assigned based upon the information supplied. Based on its understanding of the tasks, the HAT selected EPCs that could be defended; EPCs that did not achieve a consensus, or that were speculative, were dropped. The HAT was aware that HEART tends to be overly pessimistic if many speculative EPCs are selected.

The HAT tried their best to check for the known problem of "double-counting," either between GTTs and EPCs, or between multiple EPCs. This was difficult and judgmental, and there was no rigorous guidance in the method's documentation to help avoid possible double-counting.

The assessment of the proportion of effect for each EPC was also judgmental, and the team recognized that this can be a source of variation or inconsistency between assessors. Consensus weights were reported.

Step 3: Calculate HEP

The final HEP for each HFE (pHFE) was calculated in the normal HEART way:

pHFE = pGTT * (((EPC1size – 1) * EPCweight) + 1) * ((EPC2size – 1) * EPCweight) + 1) ...)

where: pHFE means probability of Human Failure Event

pGTT means probability of failure due to a Generic Task Type taken from HEART tables

EPCsize is the multiplying effect of an EPC taken from HEART tables

EPCweight is the assessed proportion of an EPC

It should be noted that HEART EPCs always make the generic pHFE more probable (pessimistic). If the above equation resulted in a number greater than one, then pHFE was set to one, as is usual for HEART.

Step 4: Consider dependencies

The HAT noted HEART's suggested method for assessing human error dependencies is the THERP model. Since the pilot study wants to assess issues with HEART itself, the team felt it would cloud the results to modify the HEART HEPs with THERP-dependency-adjusted HEPs. The team also felt that dependencies should normally be reviewed first, after viewing a list of minimum cut-sets, which was not available. If desired, the HAT can return to this issue, or it can be discussed as one of the outcomes of the pilot study. (The HAT preferred not to model implausibly low HEPs because of what one might call residual dependencies or CCFs, but the occasion did not arise, and HPLVs (human performance limiting values) are not part of HEART).

Step 5: Document HEP

Results of the HEART assessments were recorded in typical HEART tables, as well as in the formats required for the pilot experiment. A blank HEART table devised by the HAT is shown below.

Table 3-3. HEART Assessment Record

HEART Assessment Record			
HFE Code: HFE 1B	**Functional HFE Description:**		
Generic Task Code: F	**Generic Task Description:** Restore or shift a system ... following procedures with some checking ...		**Nominal HEP:**
EPC Code & Weight	**EPC Descriptions**	**Assessed Prop**	**Effect**
Assessment Notes			
Assessed Failure Probability:			
Uncertainty Bounds From Manual:			

3.2.8.2 Main References to the HEART Method

[42] Williams JC, (1986): "HEART–A proposed method for assessing and reducing human error," Proceedings of the 9th Advances in Reliability Technology Symposium, University of Bradford, UK, 2-4 April, 1986, pp B3/R/1-B/3/R/13.

3.2.8.3 Ringhals, HEART Analysis of HFE 1A

Summary of Negative Influences Identified by Ringhals, HEART Analysis

Time: EPC no. 2, "*A shortage of time available for error detection and correction,*" has been assessed at 0.8 of maximum effect (x11), resulting in a multiplier of nine (applied to the nominal value).

If no manual scram is performed, the time to perform is estimated at 27 to 29 minutes (including 11 minutes until automatic scram), while the success criterion is 20 minutes. If a manual scram is performed, the time to perform would be 16 to 18 minutes, plus time to decide to scram manually. The latter is not estimated.

The EPC "*stress*" is also noted, but with 0.4 assessed proportion of maximum effect (x1.3), this only increases the HEP by 12%.

Comparison to Empirical Data

The main negative driving factor is partly supported by the empirical data. All but one crew completed the task within the time window defined for the HFE; however, the average performance time took over 15 minutes of the 20 minute time window, which partially supports the prediction that if the crews had committed an error while performing the required task, they could have trouble completing the required tasks within the defined time window. There was no evidence for the secondary negative factor "stress."

Assessment of Negative Influences Identified by Ringhals, HEART

The analysis of the negative influence matches the empirical data in that there are no major negative driving factors. In the data, the "adequacy of time" is rated *good*, but, as the analysis indicates, there would not be much time to detect and correct any errors that might occur.

Summary of Positive Influences Identified by Ringhals, HEART Analysis

No positive factors are noted in the qualitative analysis.

The selection of Generic Task Type F, "Restore or shift a system to original or new state following procedures, with some checking," gives this HEP a nominal value of 0.003. This is one of the lowest HEART GTT nominal values. The positive factor is that the task is guided by procedures (as opposed to not having a procedure).

Comparison to Empirical Data

The selection of the HEART GTT (nominal value) accounts for the fact that the action is guided by the procedures; however, it does not explicitly credit the positive effect of the "good" rating for the Procedural Guidance. In other words, it acknowledges only that the procedural guidance exists and is followed, and does not consider the quality of the procedural guidance.

Assessment of Positive Influences Identified by Ringhals, HEART

(Not applicable. HEART has no explicitly positive PSFs (EPCs are by definition negative). As a result, the method should not be evaluated on the basis of how well it identifies positive influences.)

Summary of Operational Description Provided in the Ringhals HEART Analysis (HFE 1A)

Operational expression: Omission due to lack of time. In other words, the crew does not manage to isolate within the success window.

Assessment of Operational Description (HFE 1B)

The analysts note that the performance should be fairly routine, due to thorough practice ("operators should be able to handle the situation"); "the only difficulty is...the short time available." This suggests that the task analysis, while not detailed, was adequate to characterize the performance in operational terms. On the other hand, HEART's one-sided nature does not allow it to explicitly reflect "good performance."

3.2.8.4 Ringhals, HEART Analysis of HFE 1B

Summary of Negative Influences Identified by Ringhals, HEART Analysis

The most negative factor is EPC no. 1, "*Unfamiliarity with a situation which is potentially important but which occurs infrequently, or which is novel*," assessed at 0.4 of max. effect (x17), resulting in a multiplier of 7.4.

Time: EPC no. 2, "*A shortage of time available for error detection and correction*," has been assessed at 0.4 of maximum effect (x11), resulting in a multiplier of five (applied to the nominal value).

The quality of feedback "*Poor, ambiguous or ill-matched system feedback*" and "*stress*" are additional negative factors. Their assessed proportion of effect 0.2, 0.8, and maximum effects (x4, x1.3), respectively, mean that these values result in multipliers of "only" 1.6 and 1.24.

The Generic Task Category is an important driver of the HEP. Category "C" has been selected ("Complex task requiring high level of comprehension and skill"), and results in an HEP value before adjustment of 0.16[16]. (Note: In contrast, if Task F, "Restore or shift a system to original or new state following procedures, with some checking" had been selected, the starting value would have been 0.003. This was the category selected for HFE 1A.)

Comparison to Empirical Data

Two of the negative factors predicted as most important (high impact on HEP) are supported by the evidence (refer to the table below). The third, "shortage of time for error detection and correction," is not supported by the empirical data. Time was identified in the empirical data as "Adequacy of time–Somewhat poor."

Assessment of Negative Influences Identified by Ringhals, HEART

Overall, the prediction of driving PSFs includes some of the most important factors observed in the data. However, the prediction fails to address the issues with Procedural Guidance that are quite central to the performance of HFE 1B, which is rated "poor." Secondly, the way in which the PSFs interact is not reflected at all in the HEART analysis.

[16]It is worth noting that the HEART method makes no provision for adjustments that reduce the HEP. This results in the "conservative tendencies" mentioned in the interpretation of the HEP distribution.

Table 3-4. Predicted and Observed Negative Factors

Negative Factors Predicted in HEART Submission	Negative Driving PSFs Observed in the Empirical Data (HFE 1B)
Selection of Generic Task Category for "complex task requiring high level of comprehension and skill"	Corresponds to driving factors "complexity (scenario complexity)–somewhat high to high and "execution complexity"–somewhat high
EPC "unfamiliar...infrequent, novel situation"–poor, high impact on HEP	Corresponds to driving factor "training"–somewhat poor
EPC "shortage of time for error detection and correction"–poor, high impact on HEP	Not supported by the empirical data
EPC "poor system feedback"–somewhat negative, low impact on HEP	Corresponds to driving factor "indications of plant conditions"–somewhat poor to poor
EPC "stress"–very negative, relatively low impact on HEP	Not supported by the empirical data

Summary of Positive Influences Identified by Ringhals, HEART Analysis

No positive factors are noted in the qualitative analysis.

Comparison to Empirical Data

(Not applicable.)

Assessment of Positive Influences Identified by Ringhals, HEART

(Not applicable. HEART has no explicit positive PSFs (EPCs are by definition negative). As a result, the method should not be evaluated by how well it identifies positive influences.)

Summary of Operational Description Provided in the Ringhals, HEART Analysis (HFE 1B)

Operational expression: Due to the masking effect, the crew may have difficulties identifying the SGTR–this relates to perceiving the relevant indications. In addition, they may address the steamline break and not continue to verify for additional faults (at least not immediately, which may cause problems in meeting the time criterion defined for the HFE).

In this case, the application of the method requires the analyst to consider whether the situation may be unfamiliar to the operators. In considering this EPC, they identify why it may be unfamiliar.

Assessment of Operational Description (HFE 1B)

The operational expressions predicted in the HEART submission do not correspond to the observed performances at all. In the observed performances, the difficulty of using the procedural guidance was central, and interacted with a number of PSFs. The HEART submission does not address the procedural guidance, which tended to impede or delay success. A number of the crews appeared to have diagnosed the SGTR, and struggled to find an appropriate transfer to the E-3 procedure. The predicted possible (secondary) scenario with crew fixation on the steamline break was not observed.

3.2.8.5 Additional Comments on the Ringhals, HEART Analysis

Comments on Ringhals, HEART Analysis

The HRA team's analysis and the HEART application match exactly. The HRA team has justified or explained its HEART application, providing no other information. In this submission, there are no negative factors identified that are not represented in the method.

The HEART method does not include a means to explicitly account for positive PSFs.

Insights for Error Reduction

In terms of error reduction, the prediction is that performance will be good (relatively high rates of success), and that only an increase in the time available could potentially improve it.
For HFE 1B, with the exception of procedural guidance issues, the derived insights for error reduction point to the appropriate factors. On the other hand, these insights are not expressed in sufficiently operational terms that would point to specific error reduction measures.

Impact on HEP (Sensitivity to Driving Factors)

HFE 1A is practically characterized by the lack of any "strong" to "very strong" negative driving factor. The method predicts that performance should be sensitive to the time window, which is correct.

The HEP result for HFE 1B is strongly sensitive to two driving factors identified as most important (the negative assessments of these factors do contribute to a relatively high HEP).

Guidance and Traceability

The HEART method guidance is generally lacking, as it does not refer to a defined task analysis method, describe the EPCs, or provide support for assessing the "proportion of effect." A specific task analysis method is not mentioned in the submittal, and the EPCs have been analyzed and identified based on expert judgment. Although the documentation of the qualitative analysis for HFEs 1A and 1B is clearly structured and the HRA team has justified or explained its HEART application, the traceability of the qualitative analyses for HFEs 1A and 1B remains relatively poor due to expert judgment.

The quantitative analysis is thoroughly traceable due to the structure of the method.

3.2.8.6 Ringhals, HEART Team Comments on the Original Comparison

No comments were received.

3.2.9 PANAME (IRSN)

3.2.9.1 Short Overview of the PANAME Method (by Véronique Fauchille)

PANAME for Level 1 PRA

Starting in the nineties, French nuclear power plants have been gradually replacing their post-accident procedures, which are based on an event-oriented approach, with procedures based on a symptom-oriented approach. Symptom-oriented procedures are characterized by a looped structure, a succession of action/supervisory/reorientation to another sequence or another procedure, and a high level of human redundancy.

To model the specificities of the Symptom-Oriented Approach, the French Institut de Radioprotection et de Sûreté Nucléaire (IRSN) has launched a project called PANAME, which stands for "new action plan for the improvement of the human reliability analysis model." One of the main features of PANAME is that it supports the analysts by providing tools of assistance, such as decision trees, to quantify the parameters of the model. These are based on performance shaping factors, such as influence, training, situation complexity, workload, communication, procedural quality, and environmental quality. Data used in PANAME were provided by EDF, which performed specific simulator studies where operators used the new set of procedures.

PANAME is still a first generation HRA model. The structure of an HEP can be defined as follows:

- One part is dedicated to the modeling of the operating team (diagnosis (Pd), action and recovery by the team itself (Pe)).
- The second part is dedicated to the modeling of a recovery outside the operating team, the safety engineer, then the national crisis organization (Pr2).

$$P_{EH} = (P_d + P_e) \times P_{r2}$$

The probability of error during the diagnosis is obtained from curves, which gives the probability that the operator will fail, depending on the time allowed. Several curves are available, and a decision tree helps the analyst to choose the most appropriate curve, depending on the context. The probability of error in performing the action is a combination of three factors (a basic probability adjusted by a context factor and a probability of recovery by the team itself, depending on the time allowed to recover (Pr1)).

A second level of recovery is brought by the safety engineer and by the national crisis organization after a delay of four hours from the beginning of the abnormal situation, which corresponds to the necessary delay for the organization to be operational (Pr2).

In order to support the analysts in their choice and to facilitate the traceability of the issues of their evaluation, a software application was developed. Decision trees for the main parameters of the HRA model can be found there.

Applications

IRSN used PANAME's HRA model once in 2004 for the updating of the IRSN level 1 PRA model for 900 MWe reactors series, in order to take into account symptom-oriented procedures. At present, a similar update is being performed for the 1300 MWe reactors series. No international publication is currently available on the subject.

3.2.9.2 Main References to the PANAME Method

None currently available.

3.2.9.3 IRSN, PANAME Analysis of HFE 1A

Summary of Negative Influences Identified by IRSN PANAME

Available time, both for diagnosis and execution (including any recovery action), is the single most important factor influencing the estimated failure rate. This time could be tight, based on the defined 20 minutes, where there will be an estimated 13 to 15 minutes to diagnose the event. There also needs to be about five minutes to transition to E-3 and execute its first steps, including the isolations. A recovery factor given supervisory oversight is provided in the HEP estimate, but not much is credited because of the limited time.

The fact that there has been a safety injection also impacts the HEP estimate. (It is assumed that this has to do with the perceived seriousness of the situation, and, hence, some stress, but this is not clear from the analysis documentation).

Comparison to Empirical Data

The IRSN PANAME analysis identifies two negative factors in the base case, available time and the occurrence of safety injection. Halden's crew analysis identified one negative driving factor, execution complexity, due to the long list of actions that require both local and control room actions for performing E-3 step 3. Procedural guidance was identified in the Halden analysis as mostly good, but with a negative secondary influence, due to the observed difficulties in executing E-3 step 3, as noted above.

Assessment of Negative Influences Identified by IRSN PANAME

The IRSN PANAME analysis differs from the Halden analysis in its identification of negative factors; however, the IRSN PANAME analysis states that the available time is "the single most important factor influencing the estimated failure rate," which is also recognized in the crew simulations. All the teams managed to handle the scenario, although one crew required more time than was allowed. The IRSN PANAME analysis also mentions that the time to complete the first steps of E-3 (where this HFE ends) would be about five minutes, and those are steps that were identified as complex in the Halden analysis.

Summary of Positive and Neutral Influences Identified by IRSN PANAME

Training/familiarity with this type of event is provided, and matches the event well. It is judged that the crews should be well qualified to respond to the base case SGTR and its symptoms.

Of equal importance is the fact that there are clear symptoms/indications of the SGTR, particularly on the basis of the available alarms.

The environment is cited as nominal, allowing the operators to focus strictly on the correct response for the event. In response to subsequent questions about the analysis, other factors, such as complexity, communication, and quality of procedures are all cited as nominal/normal.

Comparison to Empirical Data

The IRSN PANAME analysis identifies training, indications, and environment as positive factors within the method, and subsequently states (in response to questions from Halden) that other factors, such as complexity, communication, and quality of procedures are all at the nominal level.

The Halden analysis of the operating crews mostly agrees with these findings, with differences in the adequacy of time and execution complexity. Adequacy of time is identified as a positive driving PSF in the Halden analysis, while it is assessed as a negative one in the IRSN PANAME analysis. Execution complexity is identified as somewhat high in the Halden analysis, while it is assessed at nominal level in the IRSN PANAME analysis; however, there is only a minor difference between "somewhat high" and "nominal."

Assessment of Positive and Neutral Influences Identified by IRSN PANAME

The analysis of positive factors in the IRSN PANAME method is straightforward for the SGTR scenario. Analysis of the simulator crews supports most of the IRSN PANAME method's assessment of positive factors. Classifying time availability as a negative or a positive factor is subjective in the IRSN PANAME method, and affects the base probability for failure in the diagnosis phase through a time correlation failure probability curve; thus, it is not a performance shaping factor, and there is no "nominal" baseline value for time. In all cases, a greater time availability leads to a lower failure probability.

Summary of Operational Description in the IRSN PANAME Analysis

The training, and, hence, familiarity level of the crews with such an event, the clear indications of the event, and a lack of environmental concerns and other strong negative influences should make this a relatively easy and straightforward task (i.e., identify and subsequently isolate the SGTR). However, the assumed 20 minute time for success is short. This is the main driver for the estimated HEP, based on the allowed time for diagnosis.

Given 14 crew simulations of this event, the estimated failure rate suggests that it is likely that no more than one or two crews will fail to meet the 20 minute assumed time for success. Seeing no failures at all is also within expectations, given the uncertainties. If such a failure does occur, the analysis suggests that it would not be due to any particular weakness in those influences that could affect performance. Instead, it would simply be due to the insufficient time in which to carry out the desired diagnosis and execution based on the success definition.

Assessment of Operational Description

The qualitative assessment based on the IRSN PANAME HRA accurately predicted the results of the simulator runs. One crew failed to handle the scenario within the time frame considered for success, while the IRSN PANAME method predicted that one or two teams would fail. The IRSN PANAME analysis also specifically predicts that any failure will be due to missing the deadline rather than to any other negative influence, which is supported by the simulator runs.

It should be noted that the qualitative assessment was not submitted as part of the IRSN PANAME analysis, but was constructed from other information submitted by the IRSN PANAME analysis team.

3.2.9.4 IRSN PANAME Analysis of HFE 1B

(Noted to be dominated by diagnosis failure; execution failure is negligible)

Summary of Negative Influences Identified by IRSN PANAME

Available time for both the diagnosis and the execution (including any recovery action) is an important factor influencing the estimated failure rate. This time could be tight, based on the defined 25 minutes, where it is estimated that there will be 15 to 20 minutes to diagnose the event. There needs to be about five minutes to transition to E-3 and execute its first steps, including the isolations. A recovery factor with supervisory oversight is provided in the HEP estimate, but not much is credited because of the short time.

The lack of clear and adequate indications and the lack of secondary radiation cues are cited as equally important and contribute to the HEP, since this makes the diagnosis particularly difficult.

The dual initiating event is also cited as a contributing factor.

Comparison to Empirical Data

The IRSN PANAME method analysis directly identifies two of the driving negative PSFs also identified in the Halden analysis of crew performance, the lack of clear and adequate indications and the relatively short time window considered for success. Dual initiating events (main steamline break and steam generator tube rupture) are also identified in the IRSN PANAME analysis as negative factors. Since this is the main reason behind the complexity of the scenario, this finding corresponds to a driving PSF scenario complexity, identified by the Halden crew analysis.

In addition to these three PSFs, Halden analysis identified procedural guidance, training, work processes, and execution complexity as driving negative PSFs. Of these, procedures were evaluated as a poor influence, while training, execution complexity, adequacy of time, and task complexity were evaluated as somewhat poor/high. While training was evaluated by Halden as a negative PSF, it has a secondary positive influence. Scenario complexity and work processes were evaluated as high.

Assessment of Negative Influences Identified by IRSN PANAME

In the Halden analysis, procedures and work processes, as well as the complexity of the scenario, were identified as the most important negative PSFs. The IRSN PANAME analysis treated training and procedures as nominal/normal, meaning that those factors did not affect calculation of the HEP, at the same time correctly identifying the scenario complexity. The IRSN PANAME analysis includes a specific PSF for the occurrence of a dual initiating event, correlated to the PSF scenario complexity in the Halden analysis. The analysis teams were not expected to predict work processes.

The IRSN PANAME analysis recognizes that even though training is sufficient for the scenario, other negative factors, such as misleading indicators and the dual initiating event, may lead to diagnosis confusion. Analysis of the simulator runs supports this conclusion; all teams were eventually able to make the correct diagnosis and transition to E-3, though not within the time frame required for success. In short, the IRSN PANAME analysis captures some of the driving negative PSFs, but not all.

Summary of Positive and Neutral Influences Identified by IRSN PANAME

Training is viewed as sufficient for this situation, particularly for performing the actual isolations, though it is recognized that the dual initiating event and lack of secondary radiation cues adds to the diagnosis confusion.

Equally importantly, the environment is nominal, allowing the operators to focus strictly on the correct response for the event.

In response to subsequent questions about the analysis, other factors, such as complexity, communication, and quality of procedures are all cited as nominal/normal.

Comparison to Empirical Data

In IRSN PANAME analysis training, complexity and procedures are identified as positive or neutral influences. Of those, only training and the working environment were included in the original analysis while the other positive or neutral influences were provided in response to subsequent questions from Halden, and were identified by the HRA team as nominal/normal. Additionally, the IRSN PANAME analysis team notes that while the training is at nominal level, there is likely to be some diagnosis confusion. The complexity factor in IRSN PANAME analysis corresponds to the Halden analysis PSF "execution complexity," not "scenario complexity."

The Halden analysis of the operating crews identified the PSF indications of conditions and training as having a secondary positive influence despite being negative driving PSFs. While the radiation indications—the single most important indicators for SGTR—were unusable due to the main steamline rupture and isolation, there are other indications for noticing the SGTR and making the correct diagnosis. Training was also listed as a negative driver because the crews had not been trained in this kind of complex scenario; however, the generally high level of training and experience among the crews allowed them to make the correct diagnosis, although it took a while for some crews, and make a knowledge-based transition into E-3.

In short, the IRSN PANAME analysis identifies one of the two positive secondary influences identified by the Halden crew analysis. The basis for their decision reflects similar findings by the Halden analysis; while they have not trained for the scenario in question, the generally high level of training is sufficient, despite the complications.

Assessment of Positive and Neutral Influences Identified by IRSN PANAME

The most important differences between the findings of the IRSN PANAME analysis and Halden crew analysis are the procedure and the execution complexity PSFs.

Procedures were cited as nominal/normal by the IRSN PANAME analysis, while the Halden crew analysis assessed them as poor, noting that most of the crews could not transfer to E-3 by just following the procedures. The IRSN PANAME guidelines for assessing procedures include criteria which could rate procedures with foldout pages as unfavorable (negative). This is supported by the Halden analysis, as six crews should have transitioned to E-3 from ES-1.1 because of the foldout page, though only two teams did so. The IRSN PANAME method guideline for assessing procedures states that "the quality of the procedures is unfavorable if the request for action is carried out at the end of the procedures and if there is a shift (theoretical or practical) between the request for action in the procedure of the reactor operator and that in the procedure of supervisor."

It seems that a procedure with foldout pages (at least) would fill the requirement for assessing the PSF as unfavorable. It is not clear why the IRSN PANAME analysis overlooked this. While some teams were identified as having problems reading the foldout page, the IRSN PANAME team also did not notice the main fault in the procedures: shifting to E-3 is based mainly on radiation alarms, which were absent in the scenario (unless the crew took a really long time to complete it), although this could also be explained by the scenario complexity PSF.

Task complexity was assessed as nominal/normal by the IRSN PANAME analysis and moderately high by the Halden analysis, as explained by the IRSN PANAME guidelines for assessing task complexity. The guidelines do take into account false indications, and whether changes in strategy are needed; however, the false indications are considered only for the operations part, when the diagnosis has already been performed. While misleading indicators and dual initiating events are very important in the scenario, they mainly affect arrival at the correct diagnosis. According to IRSN PANAME guidelines, one change in strategy (from dealing with main steamline break to SGTR in this scenario) is not enough to rate the task complexity as anything other than "nominal."

In both cases above (complexity, documents/procedures), there are very specific guidelines on assessing a value for the PSFs.. This may have led the analysis team to overlook things outside the guidelines, which might in reality have made the PSFs negative.

Summary of Operational Description in the IRSN PANAME Analysis

The assumed 25 minute time for success is short, given that the crew has to get through E-0 and ultimately enter E-3, and perform its first few steps to accomplish the desired isolations. This and the added difficulty of the lack of the secondary radiation alarms (which is typically a classic cue for an SGTR) make it difficult to diagnose the event within the time allowed (per the assumed success criteria). This is what primarily drives the estimated HEP. The crews are considered well trained on the implementation of the actual isolations, which are supposed to be easy, so failure to properly perform the isolations, assuming eventual diagnosis, is not a dominant contributor to the overall HEP.

Given 14 crew simulations of this event, the estimated failure rate suggests the likelihood that a few crews will fail to meet the 25 minute assumed time for success. Based on the given upper and lower values for the HEP, it is judged that we would see as few as one or two or as many as about a third of the crews failing to meet the assumed time frame. If such failures do occur, the analysis suggests that it would not be due to any particular weakness in those influences that could affect performance, although the lack of secondary radiation indications and dealing with the dual initiating event are expected to impede the diagnosis. Given this difficulty, and considering the workload required to get through E-0, eventually enter E-3, and perform its first few steps, there may not be enough time to carry out the desired diagnosis and execution based on the given success definition.

Assessment of Operational Description

The IRSN PANAME analysis's qualitative assessment cites some of the main difficulties that could cause a team to fail in the scenario, that is, misleading indicators combined with a relatively short time required for success. The operational story from the simulator crews notes the same difficulties, but there is the additional confusion created by the insufficient procedures, which seems to be the factor that sends many teams over the time limit. This is evident from the simulator runs; the majority of the teams make a knowledge-based transfer to E-3, some completely bypassing the procedures, others using them for confirmation of their previously made knowledge-based decision.

In conclusion, the largest deviation between the IRSN PANAME analysis predictions and the actual simulator runs is caused by the assessment of the procedures. Reasons for this are discussed in the assessment of positive factors, seen above.

It should be noted that the qualitative assessment was not submitted as part of the IRSN PANAME analysis, but was constructed from other information submitted by the IRSN PANAME analysis team.

3.2.9.5 Additional Comments on the IRSN PANAME Analysis

The IRSN PANAME HRA method captured the driving factors effectively in the base scenario, with mixed results in the complex scenario. It divides the human failure event into diagnosis and operations phases, where the diagnosis phase adjusts the base failure probability with different PSFs than in the operations phase. In the complex scenario, the analysis of the diagnosis part fit quite well, while the operations phase analysis missed two of the driving factors.

The IRSN PANAME method assigns values to the PSFs in a very structured way, using decision trees and extensive guidelines on what makes a factor positive or negative (and how positive or negative). This approach has advantages and disadvantages: for instance, clear and structured guidance reduces the amount of subjective decisions, which leads to consistency between different applications of the method and also makes the analysis work traceable. However, it also means that when analyzing complex situations, some interactions that are not covered by the method might be overlooked by the analysis team. This was evident in the mischaracterization of the procedural quality in the complex scenario–some parts of the procedures fit the guidelines for assigning it a negative value, but the main deficiency (reliance on radiation indicators for transfer to E-3) did not.

Insights for Error Reduction

No specific error correction guidelines were provided in the method; however, it is possible to identify areas for improvement within the method's PSF framework with sensitivity analysis.

Impact on HEP (Sensitivity to Driving Factors)

The PSF's impact on the HEPs is well defined in the method guidelines, with each factor's effect on the HEP ranging from 1/3 to 3 (or 9 in special cases). The values used in this application were not easily discerned from the submitted response. Direct comparison of the HEP values for the base case and the complex case yields a difference of factor three; as most of the error is assessed to result from the diagnosis step, most of the difference results from the use of different diagnosis failure probability time correlation curves, easy for the base case and average for the complex case. This means that the results are not sensitive to the PSFs used.

Guidance and Traceability

The guidance for assessing PSFs or context factors is structured with decision trees, and the criteria for assigning a specific value to a factor are well detailed. This increases the consistency and the traceability of the results.

3.2.9.6 IRSN PANAME Team Comments on the Original Comparison

With the PANAME model, HFE 1A and HFE 1B are dominated by the diagnosis failure. Execution failure is negligible, a little more so for HFE 1B than for HFE 1A.

The main difference between HFE 1A and HFE 1B is the experimental curve giving the probability of the team's failure in performing the chosen diagnosis.

In the base case scenario, the detection of the steam generator rupture is easy:

- "The SGTR is sufficient to cause nearly immediate alarms of secondary radiations and other abnormal indications, such as SG1 abnormal level and lowering pressurizer pressure."

Consequently, the PANAME curve corresponding to an easy diagnosis (without safety injection, both cases being available in the model) was chosen.
In the complex scenario:

- "The detection of the SGTR is more difficult because secondary radiation indications are quite normal due to the lack of steam flow...After a while, due to main steam isolation valve leakage, radiation indications could begin to rise and be noticeable, but this will likely take considerable time and not be evident during the initial response through E-0."

Consequently, the PANAME curve corresponding to an average context diagnosis was chosen.

In the exercise, the mission time is 20 minutes in the case of HFE 1A, 25 minutes in the case of HFE 1B, which is a little more favorable for the success of the action and which partially compensates for the difficulty of the context. However, due to the weight of the context, the diagnosis failure probability is 7,7 10-2 for HFE 1A and 2,0 10-1 for HFE 1B.

3.2.10 Enhanced Bayesian THERP (VTT)

3.2.10.1 Short Overview of the VTT-Enhanced Bayesian THERP Method (by Jan-Erik Holmberg)

The Enhanced Bayesian THERP is based on the use of a slightly modified version of the time-reliability curve introduced in the Swain's HRA handbook [11] and on the adjustment of the time-dependent human error probabilities with five PSFs: (1) support from procedures, (2) support from training, (3) feedback from process, (4) need for coordination and communication, and (5) mental load, decision burden [43]. The HEP is given by

$$p(t) = \min\{1 , p_0(t) \cdot \prod_{i=1,\ldots,5} K_i\}, \qquad (1)$$

where $p_0(t)$ is the basic human error probability taken from [11], t is the time available for identification and decision making, and $K_1, \ldots, K_5$ are the performance shaping factors.

The method is divided into a qualitative and a quantitative analysis. The qualitative analysis consists of a description of the scenario, with a block diagram and a description of the basic information in each operator action. The purpose of the block diagram is to define the operator actions in relation to relevant process events; its representation is close to a PRA event tree, but is usually somewhat more detailed. The block diagram is also used to present the dependencies between operator actions belonging to the same scenario. The purpose of describing the basic information of each operator action is to uniformly characterize the main aspects of an operator action, such as the initiating event, scenario, time windows, support from procedures and MMI, practical maneuvers needed in the action, and other noteworthy information.

The block diagram and the basic description of the operator action form the reference material when a number of experts independently make a quantitative rating of the PSFs, K1, ..., K5. The expert judgments are then aggregated in a Bayesian manner [44], and experts are asked to give qualitative comments for the justification of the judgment levels. In the real PRA projects, where the method has been used, the experts have acted as the control room operators.

3.2.10.2 Main References to the Enhanced Bayesian THERP Method

[11] A.D. Swain and H.E. Guttmann: Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications. NUREG/CR-1278, Sandia National Laboratories, Albuquerque, USA, 1983, p. 554.

[43] P. Pyy and R. Himanen: "A Praxis Oriented Approach for Plant Specific Human Reliability Analysis-Finnish Experience from Olkiluoto NPP," Cacciabue, P.C., and Papazoglou, I.A. (eds.), Proc. of the Probabilistic Safety Assessment and Management '96 ESREL'96—PSAM III Conference, Crete, June 24–26, 1996. Springer Verlag, London, 1996, pp. 882–887.

[44] J. Holmberg and P. Pyy: "An expert judgement based method for human reliability analysis of Forsmark 1 and 2 probabilistic safety assessment," Kondo, S. and Furuta, K. (eds.), Proc. of the 5th International Conference on Probabilistic Safety Assessment and Management (PSAM 5), Osaka, JP, 27 Nov.–1 Dec. 2000. Vol. 2/4. Universal Academy Press, Tokyo, 2000, pp. 797–802.

3.2.10.3 VTT, Bayesian THERP Analysis of HFE 1A

<u>Summary of Negative Influences Identified by VTT THERP</u>

Some mental load/stress is expected as a result of the need to scram, and also as long as the situation is unclear to the operators. The analysis assumes that there are 12 minutes to diagnose the event, based on the full 20 minutes from the time of the SGTR minus the three to five minutes to transition to E-3 and begin its initial steps, and minus the three minutes to implement the isolations.

Some degree to communicate and coordinate activities is the only other negative influence cited, though this is offset by good communication conditions, as stated in the analysis documentation.

<u>Comparison to Empirical Data</u>

The VTT THERP analysis identifies two minor negative factors in the base scenario, mental load/stress and communication and coordination activities. The Halden analysis of the crews identified one driving negative factor, execution complexity, and one positive driving factor with a negative secondary influence, procedural guidance. Procedural guidance was identified in the Halden analysis as mostly good, but some difficulties were observed in executing E-3 step 3. Similarly, this could be viewed as increased execution complexity.

Assessment of Negative Influences Identified by VTT THERP

The VTT THERP analysis differs in its identification of negative factors from the Halden analysis of simulator crews. Mental load/stress was not identified as a negative influence in Halden. This difference could be explained by the guidelines used in the VTT THERP method for assessing values of performance shaping factors, though the guidelines are somewhat vague; for example, a moderately high mental load/stress PSF is defined as "considerable, situation is serious, a serious decision needs to be made." Even a trivial decision can have serious consequences in the control room, but is that going to increase the operators' stress level? Is SGTR always a "serious situation"? An SGTR scenario in a simulator, in which the crews have been trained twice a year, and an SGTR in a real power plant might have different effects on the crews' stress level or mental load.

A large influence on the HEP in VTT THERP analysis is the availability of time for diagnosis. VTT THERP analysis assessed this time as 10 to 12 minutes, while in the simulator most teams entered E-3 in 8 to 10 minutes (average 9.5 minutes). The VTT THERP assessment is based on the maximum time for success, so it is accurate enough, since only one team required more time than the 20 minutes needed for success.

Summary of Positive and Neutral Influences Identified by VTT THERP

Lack of complexity is cited as the single most positive influence. It is manifested largely by the simplicity of the event itself, coupled with clear symptoms of the SGTR that can be easily observed and identified, especially those involving secondary radiation indication. No unusual decisions or actions are required.

There are good instructions that are applicable for the situation.

Also training for this type of event has been provided.

Comparison to Empirical Data

Complexity (including clear indications), instructions, and training are cited as the most important positive influences in the VTT THERP analysis. Halden analysis of the simulator crews also generally identifies the same factors as positive driving factors.

In addition to the positive factors mentioned in the VTT THERP analysis, the Halden analysis of its crews also identified adequacy of time, procedural guidance, and experience as positive driving factors.

Assessment of Positive and Neutral Influences Identified by VTT THERP

VTT THERP HRA and Halden crew analyses generally agree on the most important positive PSFs. Adequacy of time is not mentioned in the VTT THERP analysis as a performance shaping factor, but it does have an explicit effect on the HEP. The correlation between diagnosis time and base HEP is based on a time and failure probability curve, but there is no "baseline" time for normal time level. The VTT THERP method does not include a factor for work processes/team dynamics.

Summary of Operational Description in the VTT THERP Analysis

The simplicity of the event, the applicable procedures, the positive training and familiarity level of the crews with such an event, and the clear indications of the event should largely make this a relatively easy and straightforward task (i.e., identify and isolate the SGTR). However, the assumed 20 minute time for success is short. This, along with the need to scram and the operators' potential confusion over the situation, is likely to provide a level of stress/mental load that could contribute to the HEP.

Given 14 crew simulations of this event, the estimated failure rate suggests that the crews would not be expected to overshoot the 20 minute assumed time for success; neither would one failure be a total surprise. If such a failure does occur, the analysis suggests that it would not be due to any particular weakness in those influences that could affect performance. The method identifies the stress of the situation and the coordination required to accomplish everything within the 20 minute time period as contributing factors to the probability of failure.

Assessment of Operational Description

The qualitative assessment based on the VTT THERP HRA predicted, to a certain extent, the results from the simulator runs. In the simulator runs, one crew failed to handle the scenario within the time frame considered for success, while the VTT THERP method predicted that zero to one teams would fail. The VTT THERP analysis also predicted that the stress level experienced by the operators would contribute to the probability of the failure. This is not fully supported by the simulator runs, as stress was not identified as an important contribution in this relatively simple and well-taught scenario.

It should be noted that the qualitative assessment was not submitted as part of the VTT THERP analysis, but was constructed from other information submitted by the VTT THERP analysis team.

3.2.10.4 VTT, Bayesian THERP Analysis of HFE 1B

Summary of Negative Influences Identified by VTT THERP

Some mental load/stress is expected as a result of the immediate scram and the seriousness of the event (starting with a secondary breach). The analysis assumes that there are 15 minutes to make the correct diagnosis, by approximating five minutes for the first indication and five minutes for completing the actions, and subtracting these from the available 25 minutes.

Deficient feedback from processes associated with this event is considered nearly as important as the stress. This complexity is manifested by the multiple faults associated with the event (secondary breach and SGTR), and the lack of clear feedback on the event with the failure to point directly to the SGTR as masked by the steam line break. This makes it hard to notice the SGTR indications, leading to detection/interpretation problems.

The insufficient training for this event (compared to the base case SGTR) is a contributing factor to the HEP. The training of the operators on SGTR scenarios, while applicable to some extent, probably does not address this scenario directly.

Comparison to Empirical Data

The VTT THERP method analysis identified mental load/stress and deficient feedback, combined with multiple failures, as the main contributors to the HEP. Lack of training was cited as a minor contributing factor to the HEP.

Of the negative factors identified in the VTT THERP analysis, deficient feedback due to multiple failures and lack of training were also identified in the Halden analysis of operating crews as driving PSFs. Stress or mental load was not identified as a driving PSF, even though some crews did report increased stress levels.

Other negative driving PSFs identified by the Halden crew analysis, including procedural guidance and task complexity, were not identified by the VTT THERP analysis. While adequacy of time is not specifically mentioned in the VTT THERP analysis, it does affect the HEP calculation due to the used time correlation curve for base failure probability.

Assessment of Negative Influences Identified by VTT THERP

The VTT THERP analysis used a list of five performance shaping factors—quality and importance of procedures, quality and importance of training, feedback from process, mental load, and communication and coordination—to adjust a base failure probability assessed from a time-probability correlation curve. The PSF values were assessed by expert judgment, then combined within a Bayesian framework to produce the final PSF weights used to calculate the HEP values.

With the available PSFs in the VTT THERP method, it would be possible to assess the main driving negative influences identified in Halden's crew analysis, though the analysis did not identify one of the most important negative PSFs, quality and importance of procedures. As discussed in the assessment of VTT THERP's base case analysis, the guideline for assigning a value to the mental load PSF was vague. For assigning a moderately negative value to the quality and importance of procedures PSF, assessors were told that "instructions are important but they are imperfect." That applies in this scenario, but in the VTT THERP method the factor was assessed as a positive or a neutral influence; however, the procedural deficiency becomes apparent only in situations with deficient feedback (missing radiation cues), and in the VTT THERP analysis the experts judged this to be a problem with the MMI, not with the procedures.

Summary of Positive and Neutral Influences Identified by VTT THERP

Though the event is complex, it is judged that it is within the capabilities of the existing procedures, and that, if they are followed, they can lead the crew to success. Communication and coordination activities were also cited as minor positive factors.

Comparison to Empirical Data

The VTT THERP analysis correctly concludes that it is possible to handle the complex scenario within the procedures. The Halden analysis of the simulations showed that most crews had problems following the procedures, specifically in making the switch to E-3 (i.e., identifying the SGTR), because of the weight given to radiation indicators, while the VTT THERP analysis assessed this primarily as a problem with the deficient feedback rather than with the procedures.

The Halden analysis of the simulator crews identified two negative driving PSFs with secondary positive influences, training and indicators. The VTT THERP analysis cited training as a moderately poor PSF, so it was not seen as completely negative.

Assessment of Positive and Neutral Influences Identified by VTT THERP

Reasons for overlooking the procedural quality in this scenario is already discussed in the assessment of negative factors. VTT THERP's performance shaping factors use a single scale, which is the main information the method produces. The main neutral/positive driving PSFs identified by Halden for this scenario were certain elements of otherwise negative PSFs, which are not covered by the single scale used in VTT THERP for each PSF. As supplemental information, the VTT THERP method provided some comments given by the expert panel that assigned values to the different PSFs. These comments were, however, sometimes contradictory, and did not go into enough detail to assess each PSF's positive and negative element.

Summary of Operational Description in the VTT THERP Analysis

This dual event and its subsequent masking effects, including the lack of secondary radiation indications, should make this a somewhat difficult event for the crews. This difficulty can be characterized in the VTT THERP method as a ratio of the posterior/prior HEP rates–in the complex case the ratio was 10 times the ratio of the simple case. Thus, in the VTT THERP analysis, the complex case is assessed to be 10 times more difficult than the base case.

The stress level will be somewhat high, based on the scram and the initial secondary breach. Coincidentally, the deficiencies in the process feedback add to this difficulty because of the dual faults (steamline breach and SGTR) with the initial masking, making it hard to detect and properly interpret the SGTR. The minimal training for this type of SGTR event, especially with the lack of training in secondary radiation indications as an important cue for an SGTR, adds to the overall difficulty.

Given 14 crew simulations of this event, the estimated failure rate suggests that an average of two to three crews are expected to fail to meet the 25 minute assumed time for success, although all crews would be expected to fail under the 95th percentile estimate. The analysis concludes that the stress level and overall complexity associated with the event as well as the lack of familiarity with such a dual event will be the main difficulties in the scenario.

Assessment of Operational Description

The qualitative assessment constructed from VTT THERP analysis captures, to a degree, the operational story observed in the simulator runs. The confusion created by a fairly complex and untrained scenario is predicted by VTT THERP analysis, but the major observed confusion resulting from the procedures poor guidance is not. The analysis also correctly noted that handling the scenario is within the crews' ability–all the crews managed the situation, though some exceeded the time limit.

It should be noted that the qualitative assessment was not submitted as part of the VTT THERP analysis, but was constructed from other information submitted by the VTT THERP analysis team.

3.2.10.5 Additional Comments on the VTT THERP Analysis

The VTT THERP method captured the most important drivers in both scenarios, except for the PSF "quality of the procedures" in the complex scenario. As a method that uses expert judgment, the guidelines for assigning values for the different performance shaping factors are minimal, which leaves the results to the analysis team's discretion. On the other hand, it allows experts greater freedom in analyzing the PSFs.

The impact of the PSFs on HEP is explicitly stated in the method. PSFs are evaluated by a group of experts acting individually; the expert judgments are combined in a Bayesian framework to rate each PSF, and this rating is then used to increase or decrease the base error probability. The method uses five performance shaping factors, and, while the completeness of that set could be argued, they do include most of the main drivers in both scenarios. Scenario complexity is not explicitly assessed, but is covered by a combination of other factors (for example, incorrect indications and mental load).

Insights for Error Reduction

The main results of the VTT THERP method are the PSF values and the actual HEP value. The method does not provide specific insights for reducing errors, but these results can be used to identify difficult areas within the PSF framework. Sensibility analysis could also be used to evaluate HEP's response to the altered time window or the PSFs. In this manner, the method can be used to identify areas for improvement, like any other mathematical model.

Impact on HEP (Sensitivity to Driving Factors)

Between the base and the complex scenarios, the difference in prior/posterior HEP ratio caused by the identified PSFs was a factor of 10. Thus, when considering just the effect of the PSFs, the VTT THERP method is sensitive to the values assigned to the PSFs.

Guidance and Traceability

Guidance for applying the method is clear in the quantitative part of the method, that is, for calculating the HEP from the PSF evaluations and available time. Less guidance is available for the qualitative part, for example, how to generate the block diagram for the scenario. Instead of extensive guidelines for assigning values for the PSFs, the VTT THERP method relies more on the experts' insight and preferences. The PSF rating guidelines are somewhat vague, as was discussed earlier in analysis of base case. This allows more freedom to account for specific scenarios, but might create inconsistent results when used with several HRA teams.

3.2.10.6 VTT THERP Team Comments on the Original Comparison

The VTT analysis team commented that the qualitative assessment (prepared by the assessment group) may not be completely justified, due to the low number of comments generated in the expert judgment phase of the VTT analysis. More generally, VTT commented that the failure to anticipate some negative PSFs in their analysis, like the procedures PSF in the complex scenario, may be attributed to the application of the method (in their case, the expert panel part of the analysis) instead of to the method itself. As noted under the "Guidance and traceability" subsection, above, there is always some ambiguity in tracing the basis for PSF evaluations in a method that uses generic guidelines for PSF strength evaluation, if no extensive comments by the expert panel are available.

3.2.11 ATHEANA (NRC)

3.2.11.1 Short Overview of the ATHEANA Method

"A Technique for Human Event Analysis," or ATHEANA (NUREG-1624, Rev. 1 [45], NUREG-1880 [46]), is a human reliability analysis (HRA) methodology developed by the U.S. Nuclear Regulatory Commission to support the understanding and quantification of human failure events (HFEs) in nuclear power plants. Based on reviews of operating experience in technically challenging domains, such as nuclear power plants, a key observation that drives the ATHEANA approach is that HFEs that contribute to equipment damage or other severe consequences, and that involve highly trained staff using considerable procedure guidance, do not usually occur randomly or as a result of simple inadvertent behavior, such as missing a procedure step or failing to notice certain out-of-the-way indications. Instead, such HFEs occur when the operators are placed in an unfamiliar situation, where their training and procedures are inadequate or do not apply, or when some other unusual set of circumstances exists. In such situations, even highly trained staff can often make incorrect assessments regarding the status of the system being monitored or controlled, and subsequent human actions may not be beneficial, or may even be detrimental.

ATHEANA provides guidance for searching for and estimating the probability of making an error in such situations, for use in a PRA. Such situations are said to have an error-forcing context (EFC) in ATHEANA terminology. In addition, because situations with a strong EFC may not always be likely, ATHEANA also addresses evaluating and quantifying behavior in the more nominal case that is typically modeled in a PRA. More specifically, ATHEANA provides 1) guidance for identifying and modeling HFEs to be included in PRA models (including errors of omission and commission), 2) search processes for understanding a range of conditions that could drive operating crew performance in accident scenarios with respect to those HFEs, 3) guidance for representing those conditions in a PRA, and 4) a facilitator-led, expert opinion elicitation process to quantify HFEs based on those conditions.

3.2.11.2 Main References to the ATHEANA Method

[45] "Technical Basis and Implementation Guidelines for A Technique for Human Event Analysis (ATHEANA)," NUREG-1624, Rev. 1, U.S. NRC, Washington, D.C., May 2000.

[46] "ATHEANA User's Guide," NUREG-1880, U.S. NRC, Washington, D.C., June 2007.

3.2.11.3 NRC, ATHEANA Analysis of HFE 1A

Comparison–Identification of Driving PSFs

ATHEANA produced a set of six scenarios contributing to this HEP. The dominant ATHEANA scenario (contributing 99% of HEP) is Case 3A (described below). Note: the summary of factors by the HRA team considered each scenario, including the dominant scenario and all other scenarios, as a whole.

Summary of Negative Influences Identified by NRC ATHEANA

By far, the most dominant influence is the lack of available time, with the twenty minutes viewed as arbitrary and not particularly relevant to the operators since nothing in the procedures or training appears to drive them toward meeting this particular timing for this action. This time appears too tight based on the defined 20 minutes, since (a) some time will elapse between the first signs of the event and when the plant may be tripped (it appears that this could be as late as ~11 min for auto trip), and (b) it takes time to get through E-0 to at least step 19 before E-3 is even entered. The time it takes to properly get to E-3 and actually perform the isolations is therefore most affected by the methodical performance of these other serial activities, so, even though the crews will not necessarily do anything wrong, they will need more than the 20 minute criterion. Interestingly enough, as a part of the analysis to come up with the HEP, the HRA team made estimates of the time it might take them to successfully complete the action; the analysts estimated that the majority of the crew times would fall between 18 and 26 minutes. While it was noted that there are opportunities to branch off to other procedures (e.g., E-2) both before or after reaching of E-0 step 19 (to include the crews' potential decision not to go to E-3 at step 19), and possibly delay identifying and isolating the SGTR (this possibility was examined in the analysis as a form of deviation contexts), the ATHEANA team does not believe that these other possibilities will significantly contribute to the failure probability.

Some level of unfamiliarity with the HAMMLAB man-machine interface (MMI) due to limited training or exposure could be a negative factor and cause the crews to isolate the wrong SG during the isolation execution. Undetected communication error might also lead to isolation of the wrong SG, particularly in light of the slightly different numbers and roles of the operators in HAMMLAB, as opposed to the crews' normal configurations, and also since not all crews are comfortable with using two-way communication. Communication before procedure transitions may also slow down the crews.

Procedural and display deficiencies are also cited as contributing to the failure, including, for instance, lack of clear cautions in EOPs and lack of Critical Safety Function displays, as in the home plant.

Comparison to Empirical Data

The strong negative factors that were predicted are the lack of time in which to accomplish the procedural actions (but not required for HFE 1A success) and the lack of a driver to respond within the time window defined as the success criteria for this HFE are only weakly present in the base case.

Additionally, ATHEANA predicted three additional negative factors:

- Unfamiliarity with the man-machine interface due to limited training could lead to a slip during execution.
- An undetected communication error might also lead to isolation of the wrong SG.
- Procedural and display deficiencies could impact performance.

For the base case of HFE 1A, only procedural guidance was observed as a negative driving PSF (due to step 3 being a long list of local and control room actions). For the one crew that failed to meet the assumed time window in HFE 1A, it was not noted whether the operators a) followed procedures closely and/or b) did not act with a sense of urgency.

Assessment of Negative Influences Identified by NRC ATHEANA

The ATHEANA assessment of HFE 1A was driven by the assumed lack of time, which was modeled as being affected by the decision to trip the reactor, though this was only weakly borne out in the empirical data collected in the simulator.

Summary of Positive and Neutral Influences Identified by NRC ATHEANA

Clear symptoms of the SGTR (secondary radiation indications, SG-level indications, steam flow mismatch indications) are judged to be unambiguous cues. Collectively, these available indications and the excellent matching of the procedure to the event are the most positive influencing factors for this action.

The EOPs are viewed as fitting precisely to the event, especially with the above available cues for this scenario.

Experience/training/familiarity with this type of event is nearly as important a positive influence on the estimated failure rate as the previous two influences. It is judged that the crews should be quite familiar with the base case type of SGTR and its symptoms, based on their frequent training in such an event.

The task is viewed as straightforward.

Comparison to Empirical Data

Out of the five observed positive driving PSFs, three match directly to those identified in the submittal. The factors: HMI and indications (identified as "clear symptoms"), procedural guidance ("EOPs are viewed as fitting precisely"), and training are consistent with: HMI and Indication of conditions–very good; Procedural guidance–good; and Training and Experience–good to very good observed in the simulator.

Assessment of Positive and Neutral Influences Identified by NRC ATHEANA

The A Technique for Human Event Analysis (ATHEANA) assessment of positive factors directly matched the important positive factors that were observed.

Summary of Operational Description Provided in the NRC ATHEANA Analysis

The assessment of predicted operational expressions here focused on the dominant ATHEANA scenario, since it is 99% of the predicted HEP.

Description of HFE 1A ATHEANA Case 3A:

This case model failed to detect and isolate the affected SG within the time limit prescribed by the success criteria. The modeled situation in this case is that even if the operators successfully isolate SG1, they may not complete this action within the allotted 20 minutes. There is nothing driving them to this specific time. The time it takes is affected by the number of procedural steps as well as the fact that some of the actions are in the control room and some are local, manual actions.

Operational expression (from Case 3A): The operating team follows training and procedures but does not feel any drive or imperative that would motivate a faster response. As a result, it does not perform the procedural steps fast enough and does not reach the step of isolating the ruptured SG within the allotted time.

Result: A late manual shutdown is assigned a probability of 0.83.

A separate operational expression (associated with HFE 1A Case 3B) contributes about 1% of the HEP, and the remaining four are negligible. Case 3B models operator selection errors or undetected communication errors, resulting in a 0.01 contribution to the total HEP.

Assessor comment: The HRA team's characterization of the main factors reflects its experience in observing simulator crews under a range of situations.

There is a very good match between the qualitative analysis and the quantification in ATHEANA, partly because the method allows each scenario to be modeled without having to translate it into another model (i.e., the analysts do not have to express their findings in terms of the method).

Assessment of Operational Description

In HFE 1A, the operational expression of predicted performance developed a relatively high HEP (the highest predicted of all the teams) in which the expected actions take nearly all of the time available, which one would expect that it would be within the observable range. While the observed performance was not as error-prone as predicted, it would be worthwhile to compare the ATHEANA Case 3A scenario with the performance of the crews that completed the PRA-required tasks close to (or beyond) the 20 minute time window.

3.2.11.4 NRC, ATHEANA Analysis and Comparison, HFE 1B

ATHEANA produced a set of six scenarios contributing to this HEP, with the three dominant ATHEANA cases (Case 4, Case 3A, and Case 3B; listed in order of priority) contributing to nearly 97% of the total HEP. Note: the summary of factors by the HRA team considered each scenario, including the dominant scenario and all the scenarios as a whole.

Summary of Negative Influences Identified by NRC ATHEANA

Negative factors identified in HFE 1B are similar to the ones identified for HFE 1A, the base case, with additional emphasis as provided below.

The most dominant influence is the complexity involved in diagnosing the event based on the dual event scenario (steamline breach and then SGTR) and the masking effects caused by the steamline breach that drives the plant's the initial thermal-hydraulic response. The initial symptoms/cues/indications of the event that masks the SGTR, coupled with the isolation of most of the secondary radiation indications (making them appear normal) and the failure of the remaining possible secondary radiation indication, makes it very difficult for the crew to diagnose the SGTR event until well into the scenario evolution, when other SGTR cues begin to emerge (especially SG level control issues and the need for continued or repeated SI). The initial cues (or lack thereof in the case of the secondary radiation indications) will seemingly cause the crews to avoid branching off to E-3 at E-0 step 19, assuming they are diligently following procedures (because there are no high secondary radiation indications). Instead, the ATHEANA team estimates that the most likely paths to E-3 and isolating the SGTR will be (a) via E-0 step 24, when the SG levels are to be checked for any uncontrollable rise, especially when the crews will probably have already tried to control feed flow to the SGs and found it difficult for the SG with the failure, or (b) branching off to procedure ES 1.1 at step 21 of E-0, when the crew has to check whether SI can be terminated. In the latter case, conditions may be such that foldout condition #4, involving at least one SGTR symptom (level control issue) which would lead to E-3, will be met. Another case that occurs after branching off to ES 1.1 involves actually attempting to terminate SI (i.e., configure back to normal charging) and finding that SI restarts (because of the continued loss of primary coolant), thereby being a strong hint of the SGTR and/or suggesting the crew go back to E-0 and eventually confirming the stronger indications of the SGTR later in the scenario. Other possibilities of going to E-3 from E-0 step 18 or 25 are also analyzed but are not expected to lead to E-3, so their failure contributions to the overall HEP are not estimated to be significant. Interestingly enough, as part of the analysis to come up with the HEP, the HRA team estimated the amount of time that might elapse before this action was successful, postulating that the majority of the crew times would fall between 18 and 26 minutes.

The crews may simply run out of time to meet the "arbitrary" 25 minute time frame, especially with the possibility that the plant symptoms may cause them to branch to other procedures, such as E-2 or ES 1.1, before they finally reach E-3. For instance, the initial plant conditions could look like a case of a faulted SG(s), so the operators could branch to E-2 first, delaying arrival at E-3. Also, without explicit training/procedural requirements to perform this task within 25 minutes, the designated success criterion is not necessarily apparent to the crew.

The procedures are not especially helpful with dealing with this dual event scenario, especially with its masking effects. Additionally, the convoluted negative logic in E-0 step 18 could cause misinterpretation of this step, leading the crews to branch to E-2, preventing a timely arrival at E-3.

There is apparently less familiarity with SGTR events, where event diagnosis has to be based on SG level and other indications, without the benefit of secondary radiation indications.

The extreme conditions of a steamline breach and SGTR, coupled with some stress for having to deal with a steamline break, could also be a contributing factor.

Comparison to Empirical Data

In HFE 1B, the simultaneous steamline break that masks the steam generator tube produces complexity problems, which matches the dominant scenario modeled by ATHEANA (Case 4, 65% of the total HEP). Cases 3B (21%) and 3A (11%) are the other significant contributors, modeling failure to isolate the affected steam generator within the time limits when the crews terminate a safety injection (SI) in response to a signal (Case 3B) or use the ES-1.1 foldout page to transition to E-3 (Case 3A).

Assessment of Negative Influences Identified by NRC ATHEANA

ATHEANA did well in predicting the dominant scenario, and also predicted several other scenarios that might account for some of the variability of between-crew responses. The most significant negative factors in the ATHEANA submittal are complexity, time limitations, and procedures. These three are interrelated, and together provide a relatively good match with the negative PSFs from the empirical data (Procedural Guidance-poor). The submittal recognizes the need to use alternate cues for SGTR diagnosis, namely SG levels.

Summary of Positive and Neutral Influences Identified by NRC ATHEANA

The procedures are written in a way that is likely to lead the crew to E-3 (e.g., from E-2 or ES 1.1); however, it is noted this may not occur within the allotted 25 minutes, since going through these other procedure paths will take some time.

Once E-3 is entered, the task of isolating the SGTR is viewed as straightforward.

The procedures and training are well suited to perform the isolation once E-3 is actually entered.

Comparison to Empirical Data

Only one predicted positive factor ("Training") matched with the observed positive factors, although the "Training and Experience" factor for this case was only rated as somewhat poor. The other predicted positive factor was "Procedures," which did not match "Indications of plant conditions."

Assessment of Positive and Neutral Influences Identified by NRC ATHEANA

For the positive and neutral factors contributing to HFE 1B, the ATHEANA prediction only weakly matched the empirical data. The ATHEANA team noted that a lack of information typically used during their approach to the HRA (e.g., operator interviews) was not typically available, which may have affected the predictions.

Summary of Operational Description Provided in the NRC ATHEANA Analysis

The assessment of predicted operational expressions here focused on the top three ATHEANA cases, since they contribute to 97% of the predicted HEP.

Description of HFE 1A ATHEANA Case 4:

The complexity and dual nature (steamline break and SGTR) of the event and the masking effects that make up the event, along with the resulting ambiguous or potentially misleading indications, especially during the initial phase of the event (e.g., it might first look like one or more faulted SGs), will make this a difficult diagnosis, particularly within the designated 25 minutes. The possibility of branching to other procedures first (which could result to the use of many of the 25 minutes), as well as the relatively infrequent training in diagnosing SGTRs without secondary radiation indications, exacerbates the difficulty in correctly diagnosing the SGTR within the allowed time. Additionally, neither the training nor the procedures appear to focus on meeting this particular time requirement. Nevertheless, the HEP for this scenario is estimated to be less than that for the corresponding HEP for the base scenario because (a) the plant trip will occur immediately, so there is no delay time for tripping the plant and entering E-0, and (b) 25 minutes are allowed, rather than the base scenario's 20.

Operational Expression (Case 4)–Failure to accomplish the required actions within the prescribed time due to the masking and associated complexity of the indications.

Result: A late isolation for Case 4 has an evaluated probability of 0.16.

[Note: This final observation was not made in the analysis documentation, but is surmised on the basis of the results]. Given 14 crew simulations of this event, the estimated failure rate suggests that a few crews will fail to meet the assumed 25 minute success time. The analysis suggests that such failures would stem from the combination of negative influences already discussed, as well as from loss of time if/when branching off to other procedures is performed first, especially when the operators have no explicit drive to meet this time period.

Description of HFE 1A ATHEANA Case 3B:

Case 3B models failure to isolate the affected steam generator within the prescribed time limits when the crews terminate an SI in response to a signal, adding about 10 minutes to the response time. In this case, the modeled situation considers the possibility that, even if the operators successfully isolate SG1, they may not complete this action within the allotted 20 minutes.

Operational Expression (Case 3B): Failure to accomplish the required actions within the prescribed time due to distraction by other critical activities.

Operational expression: A late isolation for Case 3B has an evaluated probability of 0.05.

Description of HFE 1A ATHEANA Case 3A:

Case 3A models failure to isolate the affected steam generator within the prescribed time limits when the crews use the ES-1.1 foldout page to transition to E-3. The foldout page is checked every 10 to 20 minutes, and, when SG level increases above 10% (about 13 minutes), the operators attempt to control SG level, so the foldout page is not likely to provide the needed cue until after 16 minutes.

Operational Expression (Case 3B): Failure to accomplish the required actions within the prescribed time when the crews wait for the alternate cue (SG level) to provide sufficient indication of an SGTR.

Operational Expression–A late isolation for Case 3A is assigned a probability of 0.05.

The remaining three cases contribute to 3% of the total HEP. These primarily modeled erroneous branch points in the procedure, such as E-2.

<u>Assessment of Operational Description</u>

In HFE 1B, the operational expression of predicted performance developed a relatively high HEP (greater than 0.2), though this was lower than the HFE 1A prediction. The overall operational expression for the performance matches relatively well with the empirical data.

Table 3-5. ATHEANA Operational Expressions (Dominant Scenario)

Operational Expressions Predicted in ATHEANA Submittal (HFE 1B Dominant Scenario)	**Observed Operational Expressions**
Failure to accomplish the required actions within the prescribed time due to the masking and associated complexity of the indications.	Supported by the evidence. In general, the performance difficulties appear to be associated with the fact that the procedures do not lead to the E-3 transfer. Most of the delay is caused by this difficulty in the transfer.
Failure to accomplish the required actions within the prescribed time due to distraction by other critical activities.	Supported by the evidence. In general, the performance difficulties appear to be associated with the fact that the procedures do not lead directly to the E-3 transfer. Most of the delay is caused by this difficulty in the transfer.
Failure to accomplish the required actions within the prescribed time due to the crews waiting for the alternate cue (SG level) to provide sufficient indication of an SGTR.	Supported by the evidence. In general, the performance difficulties appear to be associated with the fact that the procedures do not lead directly to the E-3 transfer. Most of the delay is caused by this difficulty in the transfer.

Table 3-6. ATHEANA Operational Expressions (Non-Dominant Scenarios)

Operational Expressions Predicted ATHEANA Submittal (Non-Dominant Scenarios)	Observed Operational Expressions
Failure to accomplish the required actions within the prescribed time due to branching to E-2 or taking too long for field verification.	Supported by the evidence. In general, the performance difficulties appear to be associated with the fact that the procedures do not lead directly to the E-3 transfer. Most of the delay is caused by this difficulty in the transfer.
<Training was not one of the dominant contributors to the operational expression.>	This is similar to the second operational expression above. Note that the lack of training on checking alternative cues for SGTR is strongly supported by the empirical data.

3.2.11.5 Additional Comments on NRC ATHEANA Analysis

Insights for Error Reduction

The ATHEANA approach is capable of providing sufficient specificity of the error-producing conditions, such that insights into error reduction may be gained. However, in this particular application, the failure modes and error-producing conditions were documented in a more general manner that did not lend itself to providing insights into error reduction, that is, "some level of unfamiliarity with the HAMMLAB man-machine interface."

Impact on HEP (Sensitivity to Driving Factors)

This quantification explicitly modeled a base case and five sensitivity cases, which is a general characteristic of this method.

Guidance and Traceability (of HFE 1B Analysis)

The qualitative and quantitative analyses are traceable; a notable exception is the estimation of the factor probabilities (probabilities of certain performance characteristics, etc), which is based on expert judgment.

3.2.11.6 NRC ATHEANA Team Comments on the Original Comparison

The ATHEANA team was surprised at how short the assessment section was, relative to the SGTR complex case. A great deal more should have been said about the ATHEANA analysis; although the section presents the table comparing the operational stories developed by ATHEANA to those observed in the simulator, there is no discussion. Timing information that the ATHEANA team developed should also be discussed a great deal more, especially with respect to our interpretation of provided information, associated assumptions, and resulting quantification for the SGTR base case. For example, the analysis summary states that the operators "simply ran out of time," while the ATHEANA analysis describes various reasons why they may have run out of time.

The assessment team stated incorrectly that SG level mismatch was not identified as a predicted positive or neutral factor, when in fact it was predicted by the ATHEANA analysis team. The ATHEANA discussion then went on to consider at what point level mismatch would be such that an SG might be considered to be increasing in an uncontrolled manner (note that procedure requires this consideration). The issue is more complex than simply observing mismatch.

The analysis was conducted under limited information conditions, including:

- Lack of access to usual sources of information, such as:
  - access to plant engineers, operators and trainers, and procedure writers
  - access to oral history of crews' performance in many simulator drills
  - ability to observe several crews in the simulator, in a variety of accidents and transients
  - access to all procedures, including administrative procedures that define policies for the use of EOPs and AOPs
  - access to a variety of T/H analyses for the specific plant and the ability to have particular T/H cases run
- Lack of access to operator interviews. Since we have not interviewed or observed the operators in the Halden simulator, we acknowledge that they may use the procedures quite differently than our operators do. We used the EOPs in the way intended by their developers, and made our estimates accordingly. We didn't want any shortcuts in E-0 because it is the diagnostic procedure, with the order of diagnostic checks based on the priorities set by the critical safety functions analyses.

3.2.12 K-HRA (KAERI)

3.2.12.1 Short Overview of the K-HRA Method (by Wondea Jung)

The K-HRA [47] is a first-generation HRA method, which is a kind of modified method developed by Korea Atomic Energy Research Institute (KAERI) based on the ASEP HRA and the THERP. In the K-HRA, HEP for an HFE can be quantified by assessing two parts separately, a diagnosis part and an action part. The HEP of a diagnosis part is primarily determined by available time for diagnosing a relevant event, and is modified based on other PSFs.

Basic Framework of the K-HRA Method

Fig. 3-1 shows the framework of the K-HRA method. One of the arguing points in HRA is the selection of performance shaping factors (PSFs). The K-HRA selected PSFs based on the systematic review on conventional HRA methods and the ASME PRA standard. A set of comprehensive PSFs, as shown in the box on the right side of Fig. 3-1, is used in the K-HRA method's qualitative and quantitative analyses.

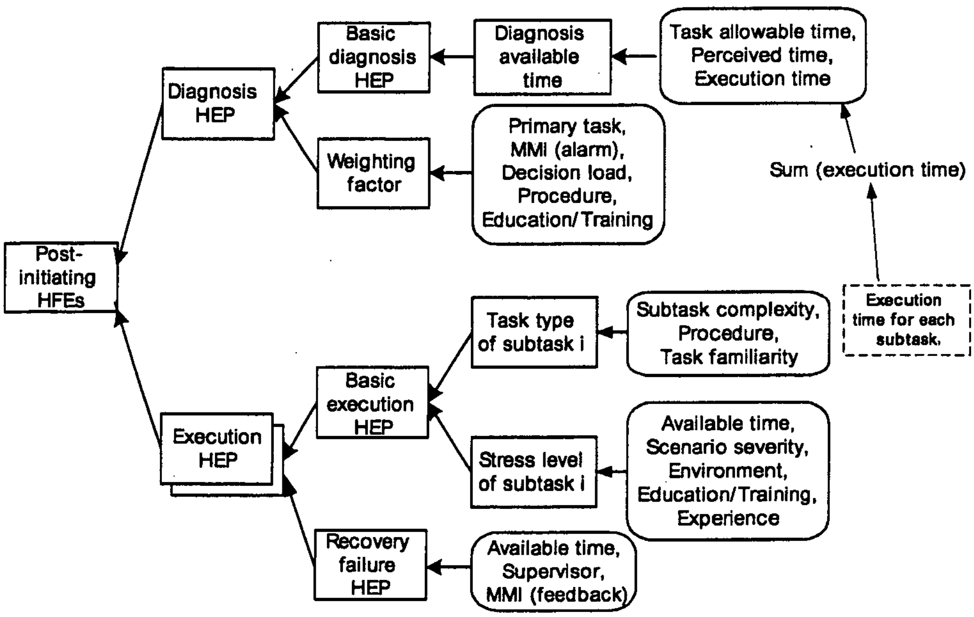

Figure 3-1. Analysis Framework for a Post-HFE of the K-HRA Method

In the K-HRA, it is assumed that human error probability can be assessed by analyzing the diagnosis and the execution separately, and the method separates nuclear power plant (NPP) human tasks into pre- and post-initiating HFEs. Pre-initiating HFEs are the human errors which occur in daily routine tasks, such as tests, maintenances, and calibrations during normal operation. This kind of routine task is performed based on procedures and pre-defined task plans, so the role of diagnosis in human behaviour is almost negligible. Thus, diagnosis error does not need to be assessed for the pre-initiating HFEs. On the other hand, human tasks related to post-initiating HFEs need both parts of human behaviour, diagnosis and execution. According to the human behaviour model, the standard method has two separate analysis processes for pre-initiating and post-initiating HFEs.

HEP for a diagnosis part, HEP(D), can be estimated as follows.

HEP(D) = Basic HEP(D) x M (weighting factor)

Where: Basic HEP(D) = *f* (diagnosis available time)

M = *f* (MMI, education/training, procedure, decision load, task priority)

Basic HEP of a diagnosis error can be represented as a function of an available time for the diagnosis of a certain task. M is a weighting factor that is selected by considering a set of PSF-related diagnosis processes, such as level of man-machine interface (MMI), quality of education/training, level of procedure, and so on. The K-HRA method provides a structured decision tree for determining M.

To assess the execution HEP, the analyst breaks down the execution part of a task into a sequence of sub-tasks. A set of technical rules on how to split the execution part is supplied by the K-HRA method. Execution HEP can be estimated as follows.

$$HEP(Ei) = \sum [\text{Basic } HEP(Ei) * HEP(Ri)]$$

where Basic HEP(Ei) = *f* (task type(i), stress level(i)),

HEP(Ri) = *f* (recovery (i) by supervisor or worker himself),

task type(i) = *f* (sub-action complexity, procedure, action familiarity),

stress level(i) = *f* (available time, situation severity, education/training, experience, environment),

recovery(i) = *f* (available time, supervisor, MMI)

Basic execution HEP for a subtask can be represented as a function of two factors, task type and stress level. Task type and stress level are determined by other related PSFs, as shown in the above equation. The basic HEP can be modified by considering recovery possibilities.

3.2.12.2 Main References to the K-HRA Method

[47] Wondea, Jung et al.: "A Standard HRA Method for PSA in Nuclear Power Plant; K-HRA Method," KAERI/TR-2961/2005, 2005.

3.2.12.3 KAERI, K-HRA of HFE 1A

Summary of Negative Influences Identified by KAERI K-HRA

Available time, both for diagnosis and execution (including any recovery action), is the single most important factor influencing the estimated failure rate. This time could be very tight, given the defined 20 minutes, since (a) some time will elapse between the first signs of the event and the time that the plant is tripped (assumed in this analysis to be five minutes), and (b) it takes time to get to at least E-0 step 19 before E-3 is even entered (assumed in this analysis to take at least seven minutes). Thus, the diagnosis time and the time to actually perform the isolations, including any time to allow for any recovery by the supervisor or another operator, are affected by these other serial activities.

Complexity associated with the execution process (multiple isolations to be executed) is the only other negative influence identified, though this is offset by the fact that the base case SGTR appears to be a familiar scenario for these crews.

It is acknowledged that some level of unfamiliarity with the HAMMLAB MMI could be a negative factor, but this is not strong enough to affect the overall neutrality to the positive influence assigned for MMI. It is also noted that the procedures do not always name specific components to be used or manipulated, but this was not sufficient to affect the overall neutral influence assigned for procedural quality.

Stress level is judged to be very high, a factor in the execution HEP value (note: based on the available analysis documentation, this is presumably because of time pressure, though the specific reason for assigning the very high stress level is not provided in the available documentation covering the analysis).

Comparison to Empirical Data

While K-HRA predicted "Available Time" as the strongest of three negative factors affecting the HEP, only one crew failed to complete the action in the designated time frame, and they were only about a minute and a half late. The main influence on the crews' completion time appeared to be the rate at which they progressed through the procedure, but it was also impacted by the time at which the crew tripped the reactor and entered E-0. The crew that failed performed well, but slowly.

K-HRA predicted that "Complexity" associated with the execution process was a negative PRA, matching one of the two observed negative PSFs in HFE 1A, "Procedural Guidance" (noted as a secondary negative influence since the overall effect of procedures was identified as positive) and "Execution Complexity." The rating of the observed Execution Complexity PSF was somewhat high.

K-HRA's judgment that stress level would be high and be a negative factor was not supported by the data, but some crews did mention some mild time pressure, which is usual for these situations.

Assessment of Negative Influences Identified by KAERI K-HRA

Other than the emphasis on stress as a negative influence, the K-HRA analysts did a good job of identifying the negative influences. The complexity of the isolation, and, by inference, the associated procedure steps in E-3, were noted. While adequacy of time was identified as positive rather than negative in the data, several crews came close to not meeting the time criterion; thus, although only one crew failed (barely) to meet the criterion, there was certainly not a lot of extra time.

Summary of Positive and Neutral Influences as Identified by KAERI K-HRA

Experience/training/familiarity with this type of event is the single most positive influence affecting the estimated failure rate, especially for the diagnosis. It is judged that the crews should be quite familiar with the base case type of SGTR and its symptoms, and the isolation tasks are not judged to be difficult.

Clear symptoms of SGTR (MMI influence) are also judged quite positively.

Availability of supervisor oversight to affect a recovery if needed, and good MMI feedback allowed for recovery in assessing the failure rate.

Other influences (procedures, work environment, decision burden, etc) are judged to be neutral influences.

Comparison to Empirical Data

The prediction of "Experience/Training" and "Clear Symptoms" strongly matched two of the four observed positive PSFs. The observed "Training and Experience" was rated as good to very good, while "HMI and indications of conditions" was rated as very good. The other observed positive PSFs were "Adequacy of Time" and "Procedural Guidance." K-HRA identified the procedures as neutral, but emphasized that the crews' familiarity with the scenario, which comes from training with the procedures, would be a positive factor. The K-HRA analysts thought, not unreasonably, that time available would be tighter than the data showed.

Assessment of Positive and Neutral Influences Identified By:

See comparison to empirical data above.

Summary of Operational Description Provided in the KAERI K-HRA

The clear symptoms of the event, and the positive experience/training/familiarity level of the crews with such an event, should generally make this a relatively easy and straightforward task (i.e., identify and subsequently isolate the SGTR). However, the assumed 20 minute time for success, given the delay before the plant may even be tripped and the necessity of reaching at least E-0 step 19 before even entering the SGTR procedure, provides time traps that give the overall success time little margin. The estimated failure rate is significantly affected by this, and, as implemented in this analysis, the diagnosis time does not credit any diagnosis for the event before the plant trip occurs, as evinced by the assigned diagnosis time of 11 minutes. (Note that it could be argued that the diagnosis of the SGTR event could occur before the trip, and may be the reason for the operators to trip the plant. Other analysts, if they were to credit the time before the trip as available for the SGTR diagnosis, would come up with at least 16 minutes of diagnosis time, and, hence, a lower diagnosis error rate).

(Note: this final observation was not made in the analysis documentation, but is surmised from the results.) Given 14 crew simulations of this event, the estimated failure rate suggests that crews would not be expected to miss the 20 minute assumed time for success; however, one failure might not be a total surprise, either. If such a failure does occur, the analysis suggests that it would not be due to any particular weakness in those influences that could affect performance, and would instead be due to the difficulty of performing the necessary serial activities within the assigned 20 minute time period.

Assessment of Operational Description

The operational description provided in the KAERI K-HRA showed a good understanding of scenario issues that could affect crew performance.

3.2.12.4 KAERI, K-HRA of HFE 1B

Summary of Negative Influences Identified by KAERI K-HRA

Available time, both for diagnosis and execution (including any recovery action), is the most important factor influencing the estimated failure rate. The 25 minute time frame would be very tight, since (a) considerable time will elapse between the initial steamline break with the resulting plant trip and the available signs of the SGTR event (assumed in this analysis to be 15 min), (b) it takes time to get through E-0 before E-3 is even entered, and (c) the masking effect of the initial steamline break includes the negation of the secondary radiation indications, due to the steamline isolation and to the delay in being able to observe parameters indicating that an SGTR has also occurred. Thus, the diagnosis time and the time to actually perform the isolations, including any time to allow for any recovery by the supervisor or another operator, are affected by these other serial activities and by the masking effect of the scenario.

The crew's experience/training relative to diagnosing the SGTR in this dual event scenario (especially because of its masking effects) is also a highly negative influence that affects the diagnosis error rate; in fact, the analysis assumes a low level of experience/training familiarity to diagnose this event, and it is judged that the crews receive no (zero) training for it. It is stated that this unfamiliarity with such a dual event is exacerbated by the need to diagnose the SGTR with no high radiation information, and it is possible that the crew would go first to EOP E-2 (based on the secondary breach), further delaying a response to the SGTR itself. The experience/training level assessed for executing the isolations (once the SGTR is diagnosed) was identified as a medium positive influence.

Complexity associated with the execution process (multiple isolations to be executed) is the only other negative influence identified. (It is noted that the stated complexity associated with the dual event diagnosis appears to be handled by assuming a low experience/training level for the crew, rather than by explicitly treating this complexity as a factor in assessing the diagnosis error rate.)

It is acknowledged that some level of unfamiliarity with the HAMMLAB MMI could be a negative factor, but this is not strong enough to affect the overall neutrality to the positive influence assigned for MMI. It is also noted that the procedures do not always name specific components to be used or manipulated, but this was not enough to affect the overall neutral influence assigned for procedural quality.

Stress level is judged to be very high, a factor in the execution HEP value (note: based on the available analysis documentation, this is presumably caused by time pressure, though the specific reason for assigning the very high stress level is not provided in the available documentation covering the analysis).

Comparison to Empirical Data

The Halden summaries that were based on empirical observations identified scenario complexity, indications, procedures, training, adequacy of time, execution complexity, and work processes as factors bearing negatively on performance. The K-HRA predicted "Available

Time" as the strongest factor, and the observed data called this "Adequacy of Time," which was rated as somewhat poor. Another predicted negative PSF was "Training," which mapped to the observed PSF of "Training" (rated as somewhat poor). Additionally, the KAERI K-HRA noted that "the masking effect of the initial steamline break includes the negation of the secondary radiation indications, due to the steam line isolation and to the delay in being able to observe parameters indicating that an SGTR has also occurred." This is consistent with the observed PSF of "Indications" being somewhat poor to poor. The K-HRA also noted that "it takes time to get through E-0 before E-3 is even entered," and that "this unfamiliarity with such a dual event is exacerbated by the need to diagnose the SGTR with no high radiation information, and it is possible that the crew would go first to EOP E-2 (based on the secondary breach), further delaying a response to the SGTR itself." Thus, the K-HRA was aware of potential problem PSFs, such as poor indications, but treated these through their impact on available time. Stress was also predicted, but was not noted explicitly in the observations; however, stress can result from time pressure, complexity, workload, and lack of training. Since these latter PSFs were actually observed, as well as complexity with a somewhat high to high rating, it is not unreasonable to think that stress could be a negative PSF in this scenario, and it was seen to some extent in the observations (but did not seem to play a major role).

Assessment of Negative Influences Identified by KAERI K-HRA

The assessment team correctly identified most of the dominant negative factors. The impact of some factors (e.g., indications and procedures) appeared to be modeled via the simultaneous experience/training for conditions of both SGTR and SLB, with neither radiation alarms nor their impact on time requirements. Additionally, the K-HRA predicted that the crews would generally fail 68% of the time; in actuality, 7 out of 14 (50%) crews failed to meet the time criterion of 25 minutes, even though all but one got it done within 32 minutes, so the prediction was relatively close to the actual data.

Summary of Positive and Neutral Influences Identified by KAERI K-HRA

Availability and layout of the indications (MMI influence) is judged to be quite positive. (Note that the stated lack of secondary radiation indications does not appear to be treated as a negative MMI influence; instead, this lack of available cues appears to be captured under the low experience/training familiarity assumed in the analysis for responding to such a masking-type scenario, with a corresponding lack of radiation indications. The problems with the indications also seemed to be considered in assessing the time requirements for the actions.)

Availability of supervisor oversight (to effect a recovery if necessary) and good MMI feedback allowed for recovery in assessing the failure rate.

Other influences (procedures, work environment, decision burden, etc) are judged to be neutral influences.

Comparison to Empirical Data

The two positive PSFs from the Halden data are training and indications of plant conditions; however, they are noted as secondary positive influences in the data, since their overall effect for HFE 1B was identified as negative. The secondary positive effects noted that general knowledge and training helped with the decision to go to E-3, and that alternative indications (e.g., SG level) eventually supported the decision. The K-HRA team identified "Availability and Layout of Indications (MMI)" as a positive influence as well.

Assessment of Positive and Neutral Influences Identified by KAERI K-HRA

While the K-HRA team identified "Availability and Layout of Indications (MMI)" as a positive influence, this does not appear to have afforded much credit in terms of recovery. In general, the PSFs were negative for this HFE in the time criterion assigned; thus, the K-HRA decision not to credit many positive factors for HFE 1B was generally consistent with the results. However, their decision to treat procedures as nominal/neutral seemed to conflict somewhat with their assumption of strong negative influences from training and the impact of the time required to get through the procedures. However, this may be a characteristic of the method.

Summary of Operational Description Provided in the KAERI K-HRA

This dual event and its subsequent masking effects, including the lack of secondary radiation indications, should make it difficult for the crews to properly respond within the assumed 25 minute success time, this difficulty exacerbated by the lack of experience/training familiarity with such a dual event. Additionally, the limited time and the built-in delays of (a) the time required for the masking effects of the secondary breach to stop hiding the SGTR event, (b) having to get through E-0 before even entering the SGTR procedure, and (c) the possibility of going into EOP E-2, further delaying explicit response to the SGTR, all provide time sinks that make the overall success time very tight. The estimated failure rate is significantly affected by these two factors (little time, lack of experience/training); further, as implemented in this analysis, the diagnosis time does not credit any diagnosis for the event for the first 15 minutes, due in part to the masking effects of the scenario. Thus, this is a critical aspect (an assigned six minutes for available diagnosis time in the analysis), leading to the resulting diagnosis HEP.

(Note: this final observation was not made in the analysis documentation, but is surmised on the basis of the results.) Given 14 crew simulations of this event, the estimated failure rate suggests that crews are expected to miss the assumed 25 minute success time; in fact, the estimated HEP suggests that we should see over half the crews fail to meet this time frame. The analysis concludes that such failures will be caused by unfamiliarity with a dual event that masks, at least initially, the signs of the SGTR, including loss of abnormal secondary radiation indications. This difficulty, along with the possibility of first entering E-2 from E-0, adds to the high likelihood of failure. The resulting time delays mean that there is insufficient time to meet the assumed success time.

Assessment of Operational Description T

The K-HRA qualitative assessment is consistent with Halden's operational summary. Although some crews did have problems with deciding whether or not to enter ES-1.1 due to the RCS pressure trend, all crews had trouble using the procedures (E-0 or ES-1.1 to get to E-3). Most crews diagnosed the SGTR based on the diverging SG levels and made a knowledge-based judgment to go to E-3, which was difficult in the time available.

3.2.12.5 Additional Comments on KAERI K-HRA

Comments on KAERI K-HRA of HFE 1A and HFE 1B

In Form B, the K-HRA identified the correct PSFs; however, the analysis identifies contributors to the HEP Weighting Factors, such as Low, Moderate, or High. The selection of these Weighting Factors is based on the analysts' judgment, and their bases are not always obvious. Additional guidance in K-HRA as to how to evaluate or rate the PSFs and make such judgments would be very useful, this being a common sentiment in regard to most methods[17].

Insights for Error Reduction

It is unclear whether the K-HRA for this event could produce meaningful insights into error reduction, though this is possible. Time is modeled explicitly via curves, so the impact can be varied. K-HRA employs weighting factors for the diagnosis; these factors are primary tasks, MMI level, Procedures, and Education/Training, and provide information regarding the scenario context, though they appear to be modeled at such a high level that it may be difficult to obtain insights that could be acted upon.

Impact on HEP (Sensitivity to Driving Factors)

For these HFEs, the use of the diagnosis curve and the choices for PSFs expected to affect the response execution HEP appeared to be reflected in the HEP, and are consistent with the results.

Guidance and Traceability

Derivation of the HEP results in K-HRA is traceable if the method is used as described. However, it is not clear that sufficient guidance is provided to adequately address all of the types of scenarios and conditions that will need to be addressed in PRA.

---

[17] With respect to this point, the KAERI K-HRA team commented that "we have a User's document of K-HRA written in Korean which includes analysis process, decision trees and criteria for PSFs evaluating. I know that the summary document submitted for the empirical study was insufficient to fully understand the contents of K-HRA. But the K-HRA provides the basis of the selection for PSF rating (such as Low, Medium, or High). One of the most important goals in developing the K-HRA was to reduce the uncertainty caused by the analyst him/herself. Therefore, the method was developed to clarify the decision process and to provide strict criteria at every decision step so that analysts could derive same or at least similar analysis result. It has several decision trees for quantifying weighting factor, determining task type or stress level, which is based on assessment on the level of PSFs. The K-HRA gives strict criteria to an analyst for selection of the level of PSFs. We understand that such kind of standardized method like the K-HRA might limit the freedom of expression for expertise; however I believe that we need such a standardized method for risk-informed activities since the freedom often leads to high uncertainty in a result of PRA" [*sic*].

3.2.12.6 KAERI K-HRA Team Comments on the Original Comparison

KAERI's HRA team agreed with the comparison summarized in the draft report (3.2.12 K-HRA), which pointed out the exact agreements and disagreements between the analyses of K-HRA and Halden simulator data.

One comment on the comparison was that we would need a common viewpoint on the PSF drivers. PSF driving factors need to be identified from the viewpoint of absolute assessment, not relative assessment, between HFE 1A and 1B. In the case of HFE 1A, the assessment group identified "adequacy of time" as one positive driving PSF. The HRA team could not agree that the "time" factor positively influences the crew's performance in the SGTR base scenario; as we understand it, the crew must feel high pressure to diagnose the event and to try to isolate the leakage as soon as possible. They must be under high stress despite their training since SGTR is a very rare event, so we would like to say that the factor of "time adequacy" would negatively influence the crew's performance, or at least be neutral.

3.2.13 CREAM (NRI)

3.2.13.1 Short Overview of the CREAM Method (by Jaroslav Holy, NRI)

Before starting a direct application of the Cognitive Reliability and Error Analysis Method (CREAM) [48], task analysis is needed to specify the basic elements of crew activities (modeled together with the given single HE-oriented basic event of the plant PRA model), which are going to be the individual separate subjects of CREAM analysis. CREAM is then applied to these basic elements, and the final HEP is enumerated as a sum of partial HEPs (provided by CREAM) corresponding to the basic elements specified.

At the beginning of a CREAM analysis, nine common performance conditions (CPC) for the given task are evaluated: 1) Adequacy of organization, 2) Working conditions, 3) Adequacy of MMI and operational support, 4) Availability of procedures/plans, 5) Number of simultaneous goals, 6) Available time, 7) Time of day (circadian rhythm), 8) Adequacy of training and experience, and 9) Crew collaboration quality, regarding the expected effect on crew performance within the given task. Three or four levels are typically used to evaluate CPC status, and are named differently for the individual CPCs (for example, the possibilities are 1) very efficient, 2) efficient, 3) inefficient, and 4) deficient for "adequacy of organization"). They are then transformed into universal levels of "expected effect on basic conditions," 1) improved, 2) not significant, or 3) reduced. The first basic step of the analysis is supplemented with an estimate of potential interactions among the individual CPCs and additional corrections of the CPCs levels assigned before.

In parallel, the given step/action of the task is linked with activity type. Fifteen activity types are available for this step in CREAM: 1) diagnose, 2) identify, 3) communicate, 4) coordinate, 5) monitor, 6) verify, 7) plant, 8) observe, 9) execute, 10) regulate, 11) scan, 12) compare, 13) evaluate, 14) maintain, and 15) record. Each activity belongs to one or more general cognitive functions, 1) observation, 2) interpretation, 3) planning, or 4) execution.

For each general cognitive function identified, the most relevant cognitive function failure mode is selected in the next step (the number of failure modes at disposal varies from two to five for the individual cognitive functions). For example, for the general cognitive function "observation," the possible failure modes are 1) observation of wrong object, 2) wrong identification made, and 3) observation not made (signal overlooking).

Each identified failure mode is firmly connected with nominal HEP value in CREAM, and with lower and upper HEP boundary values as well. This nominal value is then adjusted based on common performance conditions levels identified at the beginning of the analysis. The correction coefficients are given by CREAM and vary from 0.5 to 0.8 for "improved performance conditions," and up to two to five for "reduced performance conditions." The updated HEP value is the final quantitative result of the analysis.

Since 2003, NRI has been using CREAM in a number of cases of post-accident human errors analysis in PRAs of all six NPP units operated in the Czech Republic. In practice, CREAM has proven itself a very efficient tool, providing consistent and "reasonable" results for those crew actions where a relatively high level of complexity and cognitively causes problems when using first generation HRA methods.

3.2.13.2 Main References to the CREAM Method

[48] Hollnagel, E.: "Cognitive reliability and error analysis method CREAM," Elsevier Science Ltd., 1998.

3.2.13.3 NRI, CREAM Analysis of HFE 1A

Summary of Negative Influences Identified by NRI CREAM

The CREAM analysis views failure as a sum of individual failures of contextual control model (COCOM) elements, observation, interpretation, planning, and execution, which are considered in terms of common performance conditions (CPCs). For this HFE in the base case scenario, there were no negative or deficient CPCs. A failure at the HFE level would be the result of potential failures in observation, interpretation, and execution; the planning aspect of COCOM is considered negligible for this HFE. The analysis noted that delayed interpretation is possible because the time window for crew reaction is not excessively long, but its impact was assumed to be negligible.

Comparison to Empirical Data

The CREAM analysis did not highlight any negative influencing factors for the base case, whereas the Halden study suggested that there were some performance issues with E-3 step 3, associated with prioritization and coordination of tasks, potentially increasing the execution complexity.

Assessment of Negative Influences Identified by NRI CREAM

The CREAM method emphasizes the type of failure mode in terms of COCOM (e.g., observation, interpretation, or execution). This approach is successful in guiding the analyst to consider the full range of activities inherent in any given task. In practice, because a CPC may be present as a driving factor but not manifest in the COCOM as a failure, there may be difficulty in linking all considered CPCs to failures. Some CPCs may be undercounted in terms of their contribution to the final HEP. Thus, there may be a slight disconnection between the CPCs and the PSFs as driving factors, in terms of the way they are used throughout this report.

For example, the CREAM analysis identified potential failures resulting from observation, interpretation, or execution, which are common problem areas when entering into emergency operating procedures (EOPs). In the context of these COCOM factors, the CREAM analysis correctly highlighted the potential for the large number of simultaneous goals to affect crew performance, noting a large number of subtasks to be completed as part of the EOPs. Ultimately, however, the large number of simultaneous goals is not considered a likely contributor to failure in the CREAM analysis. Although a driving factor observed in the control room study (execution complexity in E-3) is correctly identified in the CREAM analysis, its effect is not elevated as a potential source of failure.

Summary of Positive and Neutral Influences as Identified by NRI CREAM

Several CPCs were considered to be neutral or insignificant for this scenario, and were not considered in the analysis:

- Adequacy of organization.
- Working conditions.
- Time of day (circadian rhythm).
- Crew collaboration quality.

The remaining CREAM CPCs were considered, and the following CPCs were assumed to have a positive influence:

- Above standard quality of the human machine interface.
- Completion of content and quality of ergonomics of the symptom-based procedures (the analysis team includes an entire appendix devoted to the ergonomic discussions of the symptom-based EOPs).
- Training and experience regarding the given action.

Comparison to Empirical Data

The CREAM analysis accurately considers three of the positive driving PSFs observed for the base case: training and experience, procedural guidance, and the HMI. Specifically, the CREAM analysts observed that the operators were frequently trained on SGTR, and that the E-0 and E-3 symptom-based procedures fully covered the required crew response to the SGTR. The analysts also noted that the HAMMLAB large screen display, coupled with the availability of alarms, would help operators identify and resolve the problem appropriately. These factors are encompassed by the "supportive man-machine interface" CPC level.

In contrast to the observed results from the simulator crews, the CREAM analysis does not credit adequacy of time and team dynamics. Regarding adequacy of time, the analysis assumes adequate time with only a small margin. While the time was adequate for most crews and therefore assumed to be "good" in the data analysis, one crew did respond late relative to the criterion. The CREAM analysts excluded consideration of the team dynamics (modeled under the crew collaboration quality CPC), out of the belief that it was not relevant to the analysis; however, the HRA teams were not expected to address this factor, since it was not possible to provide adequate information with the current experimental design.

Assessment of Positive and Neutral Influences Identified by NRI CREAM

The CREAM analysis was quite accurate in modeling several of the positive influences. Its failure to credit time reflects a reasonable conservatism in the analysis. The analysts' decision to exclude the crew collaboration quality CPC is not fully explained in the comparison, though the assumption that this is not a significant factor may reflect a difference in operational culture between the crews studied in the HAMMLAB and the crews located on-site with the CREAM analysis team. The analysts note elsewhere in the analysis that there may a stricter adherence to procedural steps at their home plant than in the HAMMLAB simulator crews, which may regularly hold meetings to discuss the best way to proceed in the face of a plant upset; this difference in crew response strategies would logically be reflected in the analysis, as a control room culture that emphasizes a strict adherence to procedures may not depend strongly on team dynamics to diagnose and correct plant upsets. Alternatively, it is possible that the CREAM analysis team viewed team dynamics as a factor that would be difficult to model prospectively in an HRA, particularly given that adequate information to assess this factor was not provided due to the design of the study. In the absence of preliminary information to suggest team dynamics at play in the scenario, it may be reasonable to assume that team dynamics are not a risk-significant contributor to the overall success or failure of the crews, and, as noted above, the HRA teams were not expected to address this factor.

Summary of Operational Description Provided in the NRI CREAM Analysis

The most probable failure types identified in the analysis are:

- wrong identification,
- delayed interpretation, and
- missed action.

The majority of the causes of failure for this HFE 1A action stem from cognitive and perceptual factors; the execution seems to be a reflection of the cognitive activity. The HMI, EOPs, and previous training on recognizing and responding to this event are dominant considerations when analyzing the activity. Generally, crews are expected to be successful with this activity, and the failure types seem more reflective of performance variability due to cognitive/perceptual factors than to systematic (e.g., context-induced) factors.

Assessment of Operational Description

The CREAM analysis did not directly predict the procedural difficulty with E-3 observed in the crews, but it did account for failure types consistent with the HAMMLAB findings. The CREAM analysis identified the many steps in the procedures as a potential factor but excluded these as dominant driving factors in the analysis, due to the crews' extended experience with and training in SGTR scenarios. Had the analysts considered planning as a COCOM feature, it would have aligned perfectly with the operational observations.

3.2.13.4 NRI, CREAM Analysis of HFE 1B

Summary of Negative Influences Identified by NRI CREAM

For this HFE in the complex case scenario, several potentially negative CPCs are specifically identified but found to be insignificant influences. The CREAM analysis suggests that the number of simultaneous goals might be at the upper limit of crew performance because of the large number of steps in the EOP, and that the crew may not be as well trained for the complex case as for the base case. The analysis further notes the absence of the radiation alarm as a potential CPC, but assumes that sufficient other indicators will point the crew to the correct problem diagnosis. Thus, none of these three CPCs are viewed as negatively affecting performance.

Comparison to Empirical Data

The Halden data identified a number of negative influencing factors, roughly categorized as high scenario complexity, somewhat poor to poor indications of plant conditions, poor procedural guidance, poor experience and training, somewhat poor adequacy of time, and somewhat high execution complexity. The CREAM analysis correctly identified the potential for the majority of these factors, but incorrectly assumed that the effect of these factors would be negligible; for example, while the analysis identified differences in appropriate training between the base case and complex case SGTR scenarios, it underestimated the magnitude of these differences. Crews had, in fact, little training on this variant of the SGTR, specifically on the determination that an SGTR had occurred in the absence of radiation alarms. Similarly, the analysis identified the deficiency caused by the missing radiation alarm, but assumed that the crew would be able to compensate through other available indicators. In reality, crews exhibited an overreliance on the alarm as the primary indicator of an SGTR, thereby slowing the correct diagnosis of the SGTR; however, they did eventually compensate through the other indicators, though not in time for many of the crews to meet the time criterion for the HFE.

Assessment of Negative Influences Identified by NRI CREAM

As in the base case, the CREAM analysis team successfully identified the main potential negative influences but failed to assign them at a significant level of impact that matched crew performance. Thus, the negative influences are underrepresented in the final analysis, compared to the findings from the empirical study.

Summary of Positive and Neutral Influences Identified by NRI CREAM

A single CPC is identified that has a positive influence in the CREAM analysis, availability of procedures. As in the base case analysis, the general completeness of the EOPs is noted.

Comparison to Empirical Data

General training and knowledge (not specific to the scenario) and the presence of alternative indications were identified as secondary positive influences in the crew data, since the training and indications were generally negative for this scenario. They were not identified as positive influences in the CREAM analysis.

The CREAM analysis identifies procedures as a positive influencing factor, while the observational data suggest that procedures may have had a negative influence due to lack of guidance in identifying the SGTR in the absence of the radiation alarm and due to the complexity in transitioning to the E-3 procedure.

Assessment of Positive and Neutral Influences Identified By

The CREAM analysis noted the complexity of the EOPs under the number of simultaneous goals CPC, but did not weight this as a consideration for the procedures CPC. The complexity of the task and the quality of the procedures do not appear to be orthogonal considerations, but are treated independently in the CREAM analysis.

Summary of Operational Description Provided in the NRI CREAM

The most probable failure types identified in the analysis are:

- observation not made (meaning a failure to attend to the most relevant aspects of information),
- delayed interpretation (about the cause of the fault which would impact transitioning to the correct procedure), and
- wrong action type.

In this HFE, a combination of perceptual and intentional factors (i.e., errors in developing the correct intention or plan) dominates the failure types and resulting probability. Although the procedure ergonomics were evaluated as positive, the complexity of the scenario and resultant cognitive aspect of procedure execution dominate the failure description. The failures are still cognitive in nature, and the actions simply reflect the cognitive state of the operators. The factors that contribute to the difficulty of the action include ergonomic aspects of HMI, technical content of the procedures, a lack of direct feedback or indication (i.e., vice the base case scenario), information noise, time pressure, and probably a lack of experience with this type of scenario. However, all of these appear to be on such a level that they neither increase nor significantly decrease the failure potential given by nominal HEP values provided by the CREAM method.

Assessment of Operational Description

The CREAM analysis successfully identified the fault types that were observed in the crews as failure to observe or promptly interpret the SGTR. The analysis also suggested the potential for taking the wrong action, which was not observed in the crews, but did not include one CREAM COCOM function that was observed in the crews (an inadequate plan). Inadequate planning in the CREAM sense is manifested by the difficulties some crews had in proceeding to E-3. It was an unfamiliar situation to the crews, who were neither experienced with nor trained on this variant of the SGTR, and the procedures did not quickly guide them to a diagnosis. The crews also undertook time-consuming discussions on procedure at several points.

3.2.13.5 Additional Comments on NRI CREAM Analysis

Comments on NRI CREAM Analysis of HFE 1A

Insights for Error Reduction

The consideration of COCOM factors in the analysis provides a systematic means for considering error factors; in fact, the CREAM analysis identified several opportunities for error that were not observed in the crews but which are plausible extrapolations of the actual performance.

Impact on HEP (Sensitivity to Driving Factors)

The driving CPCs—adequacy of MMI, availability of procedures, number of simultaneous goals, and adequacy of training and experience—serve to lower the basic HEPs for the three COCOM features that were considered. Given the analysts' identification of the number of steps in the procedures, it seems problematic to exclude this as a driving factor in the HEP calculation. Even if the factor is not a strong driver, it is a reasonable precaution to consider its potential impact in increasing the HEP.

Guidance and Traceability

The analysis was thorough, with separate appendices of analyses conducted on the procedures and the factors influencing control room reliability. This level of analysis, especially the analysis of the procedures, is not a strict requirement of the CREAM. In addition to supplemental analyses, the provided analysis thoroughly documents each CPC-level assumption. The analysts provide complete tables of CPCs across all scenarios, in which each level of assignment is shown and a corresponding footnoted comment explaining the decision process is provided.

It is possible, of course, to perform a CREAM analysis without such thorough documentation, instead employing a checklist approach. The level assignments approach, like many HRA methods, relies heavily on expert judgment. To ensure that a comprehensive review of the scenario is performed and that all decisions are made on the basis of available plant information and plant operations expertise, it is necessary to document all assumptions and decisions thoroughly, as the analysts have done.

Comments on NRI CREAM Analysis of HFE 1B

Insights for Error Reduction

Although the CREAM analysis captured the main fault types, it did so without full consideration of the negative influences observed in the crews. The COCOM functions serve as a strong characterization of fault activities in plant operations, and correctly point to observed performance. The COCOM functions therefore appear, when properly considered, to hold considerable prescriptive power for driving error mitigation.

Impact on HEP (Sensitivity to Driving Factors)

The CREAM analysis considers only one non-nominal COCOM function, planning, which is credited positively, presumably due to the assumed completeness of procedures for the scenario. The relationship between this COCOM and the positive CPC is not clearly articulated in the analysis. The overall effect of crediting procedures is to decrease the HEP over the nominal state; while this decrease is consistent with the weighted CPCs considered in the CREAM analysis, it does not fully consider several negative factors that were documented but not considered as negative factors.

Guidance and Traceability

See also comments under the HFE 1A base case comparison for CREAM.

Using the current analysis as a representative case, the CREAM method seems to lack guidance on a suitable level of conservatism. Although negative influences were correctly identified by the analysis team, they were not given suitable weighting to make them significant factors in the overall analysis; on the other hand, the analysis was quick to credit positive influences, and therefore fails to strike a proper balance between the negative and positive influences. As evinced in this analysis, the CREAM method might benefit from more explicit guidance on including negative influences in the analysis.

3.2.13.6 NRI CREAM Team Comments on the Original Comparison

No comments were provided.

3.3 Comparison Summary

3.3.1 The Nature of PSFs-Insights from the Empirical Data

The approach to the empirical data analysis was an attempt to identify "drivers" in terms of the PSFs. The results of the analysis of the crew responses were characterized in terms of the PSFs, and could generally be categorized into two classes:

- PSFs that are the same for all crews. These are essentially the PSFs that are external to the crews, and are determined by the scenario and the nature of the plant response, procedures, and plant interface, which are the same for all crews. These PSFs contribute to the overall/average level of performance (on an absolute scale), from which one would expect the faster crews to be above this level and the slowest crews below.

- PSFs that relate to crew characteristics, such as leadership style, crew experience, etc. These contribute to variability of crew performance.

If the observations show that a PSF appears to have affected all of the crews in the same direction, this PSF (whether generally positive or negative in terms of the effect on performance) is a driver of average performance. The crew-specific PSFs may, however, affect the degree to which the external PSFs affect the individual crews; for instance, a crew's collective knowledge will affect the degree to which the signature of the scenario affects the interpretation of the procedures. Therefore, these two sets of PSFs are not independent.

3.3.2 Implications for the Comparison with HRA Predictions

Most HRA models have been developed to support PRA models that, in order to make the solution tractable, represent a discretization of the possible range of scenarios. In a PRA model, a specific scenario is typically chosen as a representative of the spectrum of scenarios that can result from the particular initiating event type and the identified system or functional failures. In such a model, the crew-to-crew variability is an aleatory factor in the PRA context; the crew on duty at the time of the initiating event is a random variable, since the time of the initiating event itself is random. The mathematically correct way to deal with this variability in a PRA model is to perform a probabilistically weighted average over the HEP results for different crews (see discussion in Section 5.3.3.5 of NUREG-1792). The appropriate HEP to be used in a PRA is therefore the average of all crews. Consequently, many HRA models are based on the identification of those characteristics that reflect the general plant requirements for training, interpretation of procedures, etc, and that thus represent the characteristics of an "average" or representative crew. It should therefore be expected that many of the predictive analyses will not address all the factors that were found to be significant in determining individual crew performance.

When comparing the predictions of the various methods with the conclusions from the empirical data analysis, it is important to understand how the various methods incorporate the PSFs. Some of the more recent modeling approaches, such as ATHEANA and MERMOS, may formally consider crew-to-crew variability in the qualitative assessment on which they based their HEP estimate. Methods based on time reliability curves (TRCs) that are based on simulator observations (rather than those that are generic in nature, such as in ASEP), implicitly incorporate crew to crew variability since it is this variability that determines the dispersion of the TRC. The systematic PSFs, that is, those that are the same for all crews, would primarily impact the central tendency of performance, such as the median response time. Such TRC models are not predictive in the sense of identifying the PSFs responsible for the variability. By contrast, crew-to-crew variability is not explicitly called out in any of the analyses, such as SPAR-H, HEART, or CBDT, that depend on an assessment of PSFs or scenario-specific factors to directly estimate the HEP. It is therefore unlikely that the effect of this variability on the HEP has been accounted for in these analyses. The HEP is typically evaluated on the assumption of some undefined, average crew characteristic.

While there is no guidance for most HRA methods on how to account for crew-to-crew variability in determining the appropriate level of a particular PSF, it also has to be recognized that it would be difficult to assess whether variability exists without observing the crews on a number of exercises. Those HRA methods that rely on scenario-specific simulator observations to derive the HEPs (e.g., HCR/ORE) would capture this variability explicitly, but it would be very difficult to assess it a priori.

Before using an HRA model, the analyst needs to develop an understanding of the scenario, such as how the plant behaves, what information is available and when, etc, and assembles an explanation of what he or she thinks will happen, then interprets that understanding in the context of the model he or she is using to derive the HEPs. The HRA method he or she is using will largely dictate how he or she will construct his or her explanation. This qualitative assessment of the scenario should be an important step in the analysis, regardless of what HRA model is used.

3.3.3 Implications for the Conclusions to be Drawn from this Comparison

This comparison of method predictions to empirical data has been made for two HFEs only, and both those HFEs are, for different reasons, challenging to a large number of the methods—the first because it involves a decision (when to trip the reactor), for which the crews may use their judgment up to the point where the defined criteria are met, the second because it involves a very low probability scenario where the primary cues are not as expected from training or procedures, which makes it diagnostically complex. Therefore, it should not be concluded that a method is necessarily poor if the predictions did not match the empirical evidence particularly well. The various methods have been developed with a specific region of applicability in mind, as discussed above, and are generally considered to be valid within that region. This exercise highlights the areas in which those methods may need to be improved, as well as those circumstances (scenarios) where application of the HRA method may be limited.

3.3.4 Implications for Model Improvement and the Next Phase of this Research

The current phase of the comparison has not focused on the values of the HEPs. The next phase of the study will go more in-depth on both the qualitative insights on performance generated through the HRA methods and the quantitative findings of the different HRA methods. In this context, it becomes even more important to address crew-to-crew variability in the estimates. While specific HRA models may not address crew-to-crew variability in the estimation of the HEPs, they likely have the facility to explore the potential impact of the crew-to-crew variability with sensitivity studies—for example, by varying the level of a PSF (e.g., team dynamics), or by choosing a different path through the procedures. Since the crew-to-crew variability is an aleatory factor in PRA space, as discussed earlier, the appropriate HEP to be used in a PRA is the average over all crews. However, the sensitivity approach discussed above provides a way to assess the potential significance of the crew-to-crew variability. Introducing this concept into the next phase of this project could provide additional insights into the regions of validity of the methods and identify potential improvements in the use of the HRA models. For example, for situations where there may be a large variability among the crew performances, and where this variability can be demonstrated to have a potentially significant impact on future decisions, this feedback would be useful to decision makers, who may then implement multiple contingency actions or provide some form of procedure revision that may reduce this variability.

provided for the application of the HRA methods. That is, if the method applied did not provide guidance on the importance of developing and documenting the analyst's understanding of the scenario being analyzed and of the likely factors, HRA teams did not provide such documentation. On the basis of the pilot, it becomes clear that an appropriate examination of what is involved in the actions under analysis should be a critical factor for any HRA method; it also became clear that, when supported by a careful examination of the crews' required tasks and given information on the scenario, most HRAs can do a good job of identifying potential drivers of crew performance. Therefore, it is important for HRA methods to include guidance on analyzing the scenario characteristics, so that the analysts understand the cognitive and execution demands on operating crews. Furthermore, it appears that many of the methods could benefit from additional guidance on incorporating these qualitative insights into the quantification of the human error probabilities (HEPs).

- In the HRAs, the evaluation of the degree of influence of the different PSFs considered by the methods was also an important factor. In many cases, the judgments for evaluating the strength of a PSF can be difficult, and some of the results were very sensitive to the subtler ones. In most HRA methods, the guidance provided to support these judgments is limited. In addition, there is evidence that PSFs do interact to produce effects on performance. Most HRA methods note that these PSF interactions can be important, but do not provide guidance on addressing them.
- One factor contributing to the differences in the HEPs obtained by the different methods is the way the teams evaluated the extent to which the response time criterion would influence the crews' likelihood of success. Based on the information provided to the HRA teams about timing aspects of the scenarios, there were different (but not unreasonable) assumptions about when the crews would be expected to perform different aspects of the task, and some teams thought that the time criteria were somewhat arbitrary with respect to time criteria used in PRAs. The different assumptions made by the teams could to some extent be considered an artifact of the way the HFEs were defined, which is something to be considered for future experiments. However, this was certainly not the only factor that influenced the differences in the results.
- It appeared that crew factors, such as team dynamics, work processes, communication strategies, sense of urgency, and willingness to take knowledge-based actions, can have significant effects on crew performance. These effects can be positive for some crews and negative for others within the same accident scenario, as they are moderated or reinforced by other crew characteristics and/or situational features. The variability of these factors is not normally evaluated by most HRA methods. While such factors are certainly worth investigating in the context of an HRA, the effects may often have to be evaluated with sensitivity analyses of the HRA results.

4.2 Observations on the Methodology Used in the Pilot Study

The study developed a methodology for collecting crew performance observations suitable for comparison to HRA results, and demonstrated that it is possible to benchmark HRA methods using simulator experiments. Observations on the methodology and related issues include the following.

- Many issues related to the use of simulator data can be addressed through thoughtful experiment design, relative to the needs of HRA methods. For example, the perceived differences of a human action analyzed in a PRA versus one in a simulated scenario can be addressed. Constraints related to the use of the simulator can be identified a priori and documented in the "information package" provided to the analysts, allowing them to understand the intricacies of the simulated scenario and to focus their analysis on the simulated actions. The study examined the application of HRA methods within a well-defined context that enabled a comparison between method results and data.
- The study provided a detailed understanding of how HRA methods are applied. Since all analysis teams were given the same information package, this experiment showed how and to what extent analysts use the information provided in applying their method. This is an important aspect of the study, since, based on the comparison, method capability-related insights can be developed to consider relevant information and its potential impact on the HRA results.
- With respect to the analysis of the crew performance data, the derivation of the driving PSFs, and the method-to-data comparisons, this work identified some issues that need to be addressed in the follow-up studies. During this pilot phase, such methodological issues were addressed with expert judgment and through close interaction between the assessors and the experimenters. Thus, based on the internal review process and the review of the results with the HRA teams, the assessment group and the experimenters believe that the overall characterizations of the performance influences from the crew data, as well as the comparison of the empirical data with the predictions of the HRA methods, are sound.
- One area for improvement of the Empirical HRA Study methodology concerns the "variable" factors, which relate to crew attributes and functioning. Although these are in some cases important determinants of performance variability, as observed in the simulator study, most HRA methods have not been developed to address the variability of these factors, focusing instead on an "average" characterization. In the next phase of the study, efforts will be made to better define the meaning and use of the "variable" PSFs.
- Even though the challenging complex scenario simulated in the study may be of very low likelihood, the pilot study has demonstrated that the comparisons of the data from such scenarios with HRA results provide useful insights into the methods. Care must be taken to maintain the realism of the scenarios, to the extent that scenarios with highly likely HFEs should be included in future studies.

- Since only one HRA team used each method (with one exception), the pilot study was unable to clearly separate the effects of analyst or team characteristics from method-specific effects on results. Nevertheless, the current pilot methodology has already provided many useful insights. After the first two phases of this study are finished, future experiments could include multiple HRA teams per method, and, at least initially, efforts should be made to control experience level to the extent possible. Insights from such comparisons will help to reduce analyst-to-analyst variability in the application of methods.
- The HRA method requirements in terms of the resources required to perform a thorough analysis are of interest in HRA method selection, but the trade-off evaluations relevant to this issue are not addressed in this study. The goal is simply to examine the validity and reliability of the methods, regardless of their resource demands.
- As noted above, timing issues can have significant effects on the results of HRA method applications. The success for HFEs 1A and 1B was based on an expectation of how long it would take a crew to negotiate their way through the procedures, and on the time available to avoid steam generator overfill. The main focus in the crews' SGTR training was to avoid overfilling the steam generators, and to be efficient in this task. Thus, an estimate of the time needed to get through the procedures properly was used as input to the success criterion. PRA criteria, however, are typically based on the time at which the plant is considered to be in an irreversible state, which in this case would be when the steam generator power operated relief valves (PORVs) are forced open. This is some time later than the 100% level indications in the steam generators. Thus, some of the HRA teams felt that the success criteria for HFEs 1A and 1B were to some extent unrealistic. While the HFE definitions used were not necessarily unreasonable (and they did allow useful results to be obtained), this aspect is an undesirable source of variance to be avoided in the future, especially if emphasis is to be given to comparing the HEPs.
- The pilot study did not examine all of the capabilities of some HRA methods, such as the identification of HFEs and the treatment of errors of commission. The pilot study's focus on quantification as the common capability of all methods was one reason. In addition, the pilot study design had to consider practical limitations related to providing all HRA teams with the opportunity to observe and interact with the crews participating in the experiment. The impact of these limitations on the present results needs to be examined further and adjustments should be made in future studies as needed.

4.3 Upcoming Work: Pilot Study Second Phase and LOFW Study

In this first phase of this pilot study, a comparison of method predictions with the outcomes observed in two variants of an SGTR scenario has been performed, focusing on the qualitative predictions in the form of a) the factors driving performance and b) the operational expressions of these factors. The comparison has been limited to two of the nine HFEs defined for the two SGTR scenarios; these HFEs addressed the identification and isolation of the faulted steam generators in each of the two scenario variants.

Following this first phase, the following work completing the pilot study is planned for phase two:

- simulator data analysis for the remainder of the scenario,
- identification of the important factors driving performance in the remaining HFEs as well as the associated operational expressions,
- comparison of the qualitative predictions with the outcomes, and
- comparison of the quantitative HRA predictions as a measure of the difficulty associated with the complete set of HFEs.

While some of the insights on the study methodology are being incorporated into the second phase of the pilot study (completion of the comparisons based on the SGTR scenarios), this methodology will also be used for a complete study in phase three, using the loss of feedwater (LOFW) scenarios for which the simulator data collection has been performed.

5. REFERENCES

[1] R.L. Boring, S.M.L. Hendrickson, J.A. Forester, T.Q. Tran, and E. Lois: "Issues in Benchmarking Human Reliability Analysis Methods: A Literature Review," SAND2008-2619, Sandia National Laboratories, April 2008.

[2] Nuclear Energy Agency, "Errors of Commission in Probabilistic Risk Assessment," NEA/CSNI/R(2000)17, Report of the Working Group on Risk Assessment (WGRisk) of the Committee on the Safety of Nuclear Installations, OECD Nuclear Energy Agency, Paris, France, 2000.

[3] B. Kirwan: *A Guide to Practical Human Reliability Assessment,* Taylor & Francis, London, 1994.

[4] "Standard for Probabilistic Risk Assessment for Nuclear Power Plant Applications," ASME RA-S-2002 (including the Addenda to the Standard, RA-Sa-2003), American Society of Mechanical Engineers, New York, NY, April 5, 2002.

[5] B. Hallbert, R. Boring, D. Gertman, D. Dudenhoeffer, A. Whaley, J. Marble, J. Joe, E. Lois, Human Event Repository and Analysis (HERA) System – Overview, NUREG/CR-6903 Vol. 1, INL/EXT-06-11528, Washington DC, USA, July 2006.

[6] P.Ø. Braarud, H. Broberg, S. Massaiu, "Performance shaping factors and masking experiment 2006: Project status and planned analysis," Proc. Enlarged Halden Programme Group meeting, Storefjell, Norway, 11-16 March 2007.

[7] A. Kolaczkowski, J. Forester, E. Lois, S. Cooper: "Good Practices for Implementing Human Reliability Analysis (HRA)," NUREG-1792, U.S. Nuclear Regulatory Commission, Washington DC, USA, April 2005.

[8] J. P. Poloski, D. G. Marksberry, C. L. Atwood, W. J. Galyean, NUREG/CR-5750: Rates of Initiating Events at U.S. Nuclear Power Plants 1987-1995, NRC, December 1998, 55 + 227 pages

[9] P. E. MacDonald, V. N. Shah, L. W. Ward, P. G. Ellison, NUREG/CR-6365: Steam Generator Tube Ruptures, NRC, April 1996, 253 + 3 pages

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

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

[12] Bell, J., and A. Swain, "A Procedure for Conducting a Human Reliability Analysis for Nuclear Power Plants," NUREG/CR-2254, U.S. Nuclear Regulatory Commission, Washington, DC, May 1983.

[13] Parry, G., et al., "An Approach to the Analysis of Operator Actions in PRA," EPRI TR-100259, Electric Power Research Institute, Palo Alto, CA, 1992.

[14] Julius, J., J. Grobbelaar, D. Spiegel, and F. Rahn:, "The EPRI HRA Calculator® User's Manual, Version 3.0, Product ID #1008238, Electric Power Research Institute, Palo Alto, CA, May 2005.

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

[16] Gertman, D., Blackman, H., Marble, J., Byers, J., Haney, L., & Smith, C. (2005). The SPAR-H Human Reliability Analysis Method, NUREG/CR-6883. Washington, DC: U.S. Nuclear Regulatory Commission.

[17] V.N. Dang, B. Reer, S. Hirschberg: "Analyzing Errors of Commission: Identification and First Assessment for a Swiss Plant," Building the New HRA: Errors of Commission–from Research to Application, OECD NEA Workshop, Rockville, MD, USA, May 7-9, 2001 (published by OECD in 2002: NEA/CSNI/R(2002)3, 105-116).

[18] B. Reer, V.N. Dang: "The Commission Errors Search and Assessment (CESA) Method PSI Report Nr. 07-03," ISSN 1019-0643, May 2007, Paul Scherrer Institut, Switzerland.

[19] B. Reer, V.N. Dang, S. Hirschberg: "The CESA method and its application in a plant-specific pilot study on errors of commission," Reliability Engineering and System Safety 83 (2004), 187-205.

[20] B. Reer: "Outline of a Method for Quantifying Errors of Commission," LEA 09-302, Laboratory for Energy Systems Analysis, Paul Scherrer Institute, Villigen PSI, Switzerland, 2009.

[21] B. Reer, V.N. Dang: "Situational Features of Errors of Commission Identified from Operating Experience," LEA 09-303, Laboratory for Energy Systems Analysis, Paul Scherrer Institute, Villigen PSI, Switzerland, 2009.

[22] B. Reer: "An Approach for Ranking EOC Situations Based on Situational Factors," LEA 09-304, Laboratory for Energy Systems Analysis, Paul Scherrer Institute, Villigen PSI, Switzerland, 2009

[23] Beare, A.N., Gaddy, C., Singh, A., and Parry, G.W., An Approach for Assessment of the Reliability of Cognitive Response for Nuclear Power Plant Operating Crews, in Proceedings of Probabilistic Safety Assessment and Management, Beverly Hills, CA., Elsevier, February 1991

[24] Parry, G.W., et al., An Approach to the Analysis of Operating Crew Responses Using Simulator Exercises for Use in PSAs, presented at the OECD/BMU Workshop on special issues of level 1 PSA, Cologne, FRG, May 28, 1991

[25] Parry, G.W., Holy J., Kucera L., Human Reliability analysis - Analysis file 6T47AF01, NPP Temelin PSA documentation, Revision 1, March 1996

[26] Holy J., Kucera L., Human Reliability analysis - Analysis file 6T47AF01, NPP Temelin update PSA documentation, Revision 1, December 2001

[27] C. Bieder, E. Desmares, F. Cara, P. Le Bot, J.L. Bonnet: "MERMOS: New Issues on Human Failure in Emergency Operations of Nuclear Power Plants," Probabilistic Safety Assessment and Management, PSAM4, New York, USA, September 13-18, 1998.

[28] C. Bieder, P. Le Bot, E. Desmares, F. Cara, J.L. Bonnet: "MERMOS: EDF's New Advanced HRA Method" Probabilistic Safety Assessment and Management, PSAM4, New York, USA, September 13-18, 1998.

[29] P. Le Bot, C.Bieder, F. Cara: "MERMOS, a second generation HRA method : what it does and doesn't do," International Topical Meeting on Probabilistic Safety Assessment, PSA'99, Washington D.C., USA, August 22 – 26, 1999.

[30] C.Bieder, P. Le Bot, F. Cara: "What does a MERMOS analysis consist in?" International Topical Meeting on Probabilistic Safety Assessment, PSA'99, Washington D.C., USA, August 22 – 26, 1999.

[31] F. Cara, C.Bieder, P. Le Bot: "What are the CICAs ? Retrospective analysis of the Three Mile Island accident from the MERMOS viewpoint" International Topical Meeting on Probabilistic Safety Assessment, PSA'99, Washington D.C., USA, August 22 – 26, 1999.

[32] P. Le Bot, C. Bieder, S. Vidal: "Feedback from the actual implementation of the MERMOS method," Probabilistic Safety Assessment and Management, PSAM5, Osaka, Japan, November 27 - December 1, 2000.

[33] P. Le Bot, H. Pesme, F. Ruiz: "Methodological Validation of MERMOS through 160 analyses,," Probabilistic Safety Assessment, PSA'02, Detroit, MI, USA, October 6 – 9, 2002.

[34] P. Le Bot "Human reliability data, human error and accident models—illustration through the Three Mile Island accident analysis," in *Reliability Engineering & System Safety*, Volume 83, Issue 2, February 2004, Pages 153-167.

[35] P. Le Bot, A. Voicu: "Exploratory study of the impact of organisational factors on safety using Probabilistic Safety Assessments," presentation only, at Probabilistic Safety Assessment, PSA'05, San Fransisco, CA, USA, September 11-15, 2005,

[36] P. Le Bot: "Using expert judgement with MERMOS : from static assessment towards knowledge capitalisation," Seminar of the European Commission, The Use of Expert Judgment in Decision Making, Aix en Provence, France, 2005,

[37] P. Meyer, P. Le Bot, H. Pesme: "MERMOS : an extended second generation HRA method," Proc. 2007 IEEE 8th Human Factors and Power Plants (HFPP) and 13th Annual HPRCT Meeting, August 26-31, 2007, Monterey, CA, USA.

[38] H. Pesme, P. Le Bot, P. Meyer: "Little stories to explain Human reliability Assessment : a practical approach of the MERMOS method," Proc. 2007 IEEE 8th Human Factors and Power Plants (HFPP) and 13th Annual HPRCT Meeting, August 26-31, 2007, Monterey, CA, USA.

[39] P. Le Bot, P. Meyer, H. Pesme: "The CICA concept for use in the MERMOS method redifined by a new organisational reliability model," Proc. 2007 IEEE 8th Human Factors and Power Plants (HFPP) and 13th Annual HPRCT Meeting, August 26-31, 2007, Monterey, CA, USA.

[40] Pierre Le Bot: "Can nuclear safety engineering deal with the human and organizational aspects of error?" Symposium INSS Human factors, Tsuruga / Tokyo, Japan, 2007.

[41] Williams JC, (1986) HEART – A proposed method for assessing and reducing human error. In: Proceedings of the 9th Advances in Reliability Technology Symposium, University of Bradford, UK, 2-4 April, 1986, pp B3/R/1-B/3/R/13.

[42] P. Pyy, P., R. Himanen, A Praxis Oriented Approach for Plant Specific Human Reliability Analysis - Finnish Experience from Olkiluoto NPP. In: Cacciabue, P.C., and Papazoglou, I.A. (eds.), Proc. of the Probabilistic Safety Assessment and Management '96 ESREL'96 — PSAMIII Conference, Crete, June 24–26, 1996. Springer Verlag, London, 1996, pp. 882–887.

[43] J. Holmberg, J., and P. Pyy, An expert judgement based method for human reliability analysis of Forsmark 1 and 2 probabilistic safety assessment. In: Kondo, S. & Furuta, K. (eds.), Proc. of the 5th International Conference on Probabilistic Safety Assessment and Management (PSAM 5), Osaka, JP, 27 Nov. – 1 Dec. 2000. Vol. 2/4. Universal Academy Press, Tokyo, 2000, pp. 797–802.

[44] "Technical Basis and Implementation Guidelines for A Technique for Human Event Analysis (ATHEANA)," NUREG-1624, Rev. 1., USNRC, Washington, D.C., May 2000.

[45] "ATHEANA User's Guide, NUREG-1880, USNRC, Washington, D.C., June 2007

[46] Wondea Jung, et. al., "A Standard HRA Method for PSA in Nuclear Power Plant; K-HRA Method," KAERI/TR-2961/2005, 2005.

[47] Hollnagel E., Cognitive reliability and error analysis method CREAM, Elsevier Science Ltd., 1998

[48] HWR-844 rev 2, International HRA Empirical Study-Pilot Phase Report, OECD Halden Reactor Project, October 2009.

APPENDIX A
CREW SUMMARIES

APPENDIX A

CREW SUMMARIES

A.1 HFE 1A Success

Summary of Data Analysis of Crew H Scenario Base

HFE: HFE 1A: Failure of the crew to identify and isolate the ruptured steam generator in the base scenario

Short story of what happened in the selected part of the scenario (written after reviewing DVD, logs, and interview):

Identification:

03:40 SS: *"Activity nitrogen 16 SG, level, SG1."* Quick response to alarm.

04:17 SS decides to take a meeting of the form "decision with consultation." In the meeting it is decided that RO should check for applicable procedures. The SS says that they must reduce the load, and the meeting is never continued.

06:35 The crew performs manual reactor trip based on activity indication from sampling.

07:29 Safety injection actuated on order from SS.

10:23 Auxiliary feedwater is stopped to SG1 according to the note in E-0 step 12.

11:42 RO enters E-0 step 19, "Check if SGs are not ruptured." RO informs SS that they should transfer to E-3. SS says he will call chemistry.

12:27 RO starts E-3.

RO and ARO are working together in E-0. ARO's answers are quick and clear. RO reads the procedure clearly and reads notes and warnings.

Isolation:

13:30 RO starts isolation steps in E-3.

16:00 SG1 is isolated

RO asks SS to call a field operator to perform the local actions. RO misses one local action that is not normally performed at the operators' home plant, but when RO discovers this, it is corrected. The rest of the isolation is performed by RO and ARO without further problems.

Summary of the most influencing factors affecting performance:

Direct negative influences

Negative influences present

Neutral influences

- Time Pressure
  - No observation of perceived time pressure.
- Stress
  - In the interview, crew says they experienced a little stress in the beginning of the scenario but not after reactor trip, when they worked in the procedures.
- Complexity
  - All expected SGTR diagnosis information is available, as in a straightforward SGTR scenario.
- Procedures
  - Procedures match well with the available information and the scenario development. Poor layout of E-3 isolation step (step 3).
- Experience
  - SS is experienced, RO is experienced, ARO is less experienced.
- Communication
  - RO and ARO have good communication.
- Interface
  - The computerized interface is new to the crew, but is designed to resemble the home plant's functionality. Large screen display could facilitate overview.

Positive influences

- Training
  - SGTR training is normally held twice a year in the simulator.
- Work Processes
  - RO and ARO work well together. Procedure work is good.
- Team Dynamics
  - SS makes quick decisions, without much discussion. This is good as long as the decisions are good. SS calls a meeting, but doesn't keep the structure.

Summary of the (a) observed difficulty or ease that the crew had in performing the action of interest and (b) why the action was easy or difficult, based on observations from DVD review and interview with the crew:

The crew easily identified the tube rupture because they were well trained for the situation. They quickly made the decisions to trip the reactor. The SS generally made decisions quickly, without much discussion. The crew easily isolated the steam generator because they used a procedure that worked well for the situation.

Summary of Data Analysis of Crew M Scenario Base

HFE: HFE 1A: Failure of the crew to identify and isolate the ruptured steam generator in the base scenario

Short story of what happened in the selected part of the scenario (written after reviewing DVD, logs and interview):

Identification:

03:23 Detection of activity alarm. RO: "*Activity SG1 nitrogen 16.*" RO and SS check for other changes in process parameters and call them out.

04:04 SS: "*It looks like we have had a large tube rupture here.*"

04:39 Manual reactor trip: SS says that the level is dropping fast in the pressurizer. RO adds that they have extremely low pressurizer pressure. SS orders the trip.

06:41 Manual actuation of SI. SS monitors the pressurizer level, which is 13% and decreasing. SS suggests actuating SI. RO agrees, and comments that they know that they have a problem.

08:38 Crew is at E-0 step 12 when SS orders ARO to stop all feed water to SG1 and then to stop the steam-driven AFW pumps (as in E-3 step 3).

11:03 Crew decides to enter E-3 (from E-0 step 19).

SS is very involved in detailed tasks, often taking the role of RO (e.g., communicating events to all crew, giving orders directly to ARO). SS intervenes twice while the crew is in E-0 and orders steps from E-3 to ARO.

Isolation:

11:20 The crew has decided to enter E-3 (from E-0 step 19) when the SS orders ARO to dispatch an operator to isolate AFW pump steam supply valve (E-3 step 3d).

13:36 SG1 is isolated.

The crew is quick in identifying and isolating. The SS makes independent decisions without consulting/discussing with the other crew members. The SS also assigns tasks to ARO ahead of procedures.

Summary of the most influencing factors affecting performance:

Direct negative influences

Negative influences present

Neutral influences

- Scenario Complexity
  - All expected information to diagnose SGTR is available, as in a straightforward SGTR scenario.
- Procedures

- Procedures match well with the available information and the scenario development. Poor layout of E-3 isolation step (step 3).

- Communication
  - RO does not read warnings and notes. ARO answers clearly. SS communicates a lot, but takes over RO tasks.
- Interface
  - The computerized interface is new to the crew, but is designed to resemble the home plant's functionality. Large screen display could facilitate overview.
- Experience
  - SS is less experienced, RO is experienced, ARO is less experienced.
- Work Processes
  - SS gives direct orders to ARO, who is sometimes occupied and leaves RO working alone. SS assigns tasks to ARO ahead of procedures.

Positive influences

- Training
  - SGTR training is normally held twice a year in the simulator.
- Team Dynamics
  - The SS is very involved in the tasks, but at the same time keeps a good overview of the process situation and the crew's work progress.

Summary of (a) the observed difficulty or ease that the crew had in performing the action of interest and (b) why the action was easy or difficult, based on observations from DVD review and interview with the crew:

The crew experiences a good match between the procedures and the process situation, and easily works ahead of the procedures. In the interview they point to no difficulties, saying that they were well trained and that this was a familiar scenario. SS is involved in tasks and made decisions without consultations, but makes them correctly, which enabled a fast performance.

A.2 HFE 1A Failure

Summary of Data Analysis of Crew N Scenario Base

HFE: HFE 1A: Failure of the crew to identify and isolate the ruptured steam generator in the base scenario

Short story of what happened in the selected part of the scenario (written after reviewing DVD, logs, and interview):

Identification:

03:20 RO: "*Activity nitrogen 16, SG1.*" All crew members check process parameters and call out actions.

04:00 SS: "*We have to go to AOP 3 then*" (procedure for handling small tube leak up to max. 10 kg).

05:00 RO says pressurizer level cannot be maintained and wants to trip reactor. SS agrees and gives the order to trip at the same time as RO says *"I'll trip."*

07:20 SS and RO discuss decreasing PRZ level. SS: *"It is 11%, we actuate SI."*

11:00 ARO closes aux feed water to SG1 by order of the RO. When ARO is working on AFW, RO does not continue in the procedure, but checks the alarms.

11:50 ARO: *"I see indication for activity on the RMS for SG1 also."* SS: *"It is SG1 only?"* RO and ARO: *"Yes."*

17:00 RO: *"I enter E-3."* SS: *"Then we take a meeting first."*

19:10 RO starts E-3.

Good communication between RO and ARO. RO reads clearly and waits for ARO's answers. ARO answers well, repeating object and status. All crew members repeat received information. Procedure work is thorough and controlled, but not quick. ARO is sometimes slow in answering, and sometimes RO and SS take time to discuss. RO waits for ARO's actions, which slows them down. SS does not say that they are in a hurry, but interrupts sometimes with less important things.

Early on, the crew discusses the size of the leakage in SG1. They consider the procedure for handling a small tube leak, and SS suggests starting this procedure. Based on fast decrease in PRZ level, the crew decides to manually actuate reactor trip. When RO enters E-0 step 4, SI has not been actuated (PRZ > 10%). The crew uses some time to check the right column in step 4; SS and RO discuss actuating SI. SS calls for test of activity, opening of blow down and sampling valves; the crew closes aux feed to SG1 and checks activity indications on all SGs of the RMS picture. They also check indications on SGTR at an early point. There is a misunderstanding between RO and ARO about closing a steam valve, due to RO giving orders to close "the steam valve." RO means main steam valve, while ARO interprets this as steam valve to the steam-driven aux pump. The misunderstanding is discovered and corrected about a minute later. When completing E-0 step 19, RO suggests transferring to E-3. SS calls for a meeting before transferring to E-3. The meeting is very well conducted, and they settle on E-3 and isolating SG1. Calling a meeting when considering transferring from E-0 (to check status and coordinate strategy) is considered good practice at home plant.

<u>Isolation:</u>

19:10 RO starts E-3.

25:00 SG1 is isolated.

The crew checks all bullets of E-3 step 2 quickly and thoroughly to assure that the correct identification is SG1, and performs the isolation (E-3 step 3) well.

Summary of the most influencing factors affecting performance:

Direct negative influences

- Work Processes

- RO sometimes waits for ARO to take actions and does not continue alone in procedure. ARO does not use large screen display, but uses time to find the right display for checks. The way the crew works is generally good.

- Team Dynamics
  - SS does not focus on preventing overfilling SG. Calling a quick meeting at procedure transfer is considered good practice, but slows them down. SS has a good overview and the operators perform thorough checks and verifications, but because they are slow in this situation, they overfill the SG later.

Negative influences present

Neutral influences

- Interface
  - The computerized interface is new to the crew, but is designed to resemble the home plant's functionality. Large screen display could facilitate overview. No problems using the interface were observed, but ARO did not use the large screen efficiently.
- Scenario Complexity
  - All expected information to diagnose SGTR is available, as in a straightforward SGTR scenario.
- Procedures
  - Procedures match well with the available information and the scenario development. Poor layout of E-3 isolation step (step 3).
- Communication
  - Generally good, except one little misunderstanding.
- Experience
  - All crew members are experienced. They have not worked as a team for long.

Positive influences

- Training
  - SGTR training is normally held twice a year in the simulator.

Summary of (a) the observed difficulty or ease that the crew had in performing the action of interest and (b) why the action was easy or difficult, based on observations from DVD review and interview with the crew:

The crew had good indications of an SGTR in SG1. The indications were detected early and checked thoroughly. No specific difficulty with the diagnosis or the isolation was observed. The DVD analysis left the impression that both the procedures and the crew's additional assessment of the situation clearly indicated SGTR SG1. It seemed like the whole crew was well updated and coordinated on the situation and the chosen strategy, which also indicates that no difficulty was experienced.

The crew did not work as quickly as expected. The SS did not focus on speeding up the work and ARO was sometimes slow in responding, because the interface was new and ARO used the operating displays and not the large screen display. The work was done in a controlled, but not quick, manner.

A.3 HFE 1B Success

Summary of Data Analysis of Crew I Scenario Complex

HFE: HFE 1B: Failure of the crew to identify and isolate the ruptured steam generator in the complex scenario

Short story of what happened in the selected part of the scenario (written after reviewing DVD, logs, and interview):

Identification:

04:15 Automatic reactor trip. Reactor trip, SI, and steamline isolation are detected by RO, but they do not initially understand that they have a secondary break.

09:22 RO enters step 19 (Check if SGs are not ruptured). RO sees no indications of tube rupture. SS says OK. They continue in the procedure.

13:00 RO starts ES-1.1. RO opens the foldout page but does not read it carefully. The transfer to E-3 on high SG level is never detected during the scenario.

14:25 After stopping a high head SI pump, RO detects that the RC pressure is decreasing but thinks it is caused by cooldown.

14:52 RO discovers that the level is rising only in SG1. SS orders ARO to close AFW to SG1. ARO asks if they have tube rupture, but SS says no. The RCS pressure stabilizes, but the pressurizer level does not rise. RO says that they are consuming water somewhere. RO and SS discuss possible reasons for this.

17:30 SS orders a return to E-0. SS now leads the procedure work and quickly checks some selected steps in E-0. SS orders RO to do steps 18 and 19 (Check if SGs are not ruptured), but they still have no activity.

20:00 SS says that they could get water from somewhere else, so they close the normal feedwater bypass valve to make sure they have no flow to SG1. RO points out that they cannot increase the pressurizer level. (Strong indication that the level in SG1 comes from RCS.)

20:30 SS mentions E-3. SS tries to come up with a reason why they don't have activity, suggesting that they might not have a sampling flow even though the valve is open. SS decides to go to E-3, and orders the transfer.

22:35 RO starts E-3. This is a knowledge-based decision, because they have not found a direct order from the procedures to go to E-3.

SS is experienced. RO and ARO are less experienced. SS takes care of the alarm list and sometimes does ARO's and RO's jobs (answering the questions in the procedures and leading the procedural work, respectively). SS actively discusses the operation with the operators.

Most communication is between SS and RO. ARO is very quiet, and only responds to direct orders or questions. RO often seeks approval from SS.

The crew (SS) quickly turns ideas into actions. SS is good at prioritizing. When they restart E-0, SS only performs steps relevant for identification. Since SS makes most decisions alone, they do not lose much time in discussion.

Isolation:

24:00 RO starts E-3 step 3 (first isolation step). RO performs the isolation steps together with ARO. ARO writes down the local orders and then calls a field operator to carry them out. ARO misses the chance to close the turbine-driven AFW pump steam supply, but orders a field operator to do this locally. (At the crew's home plant, this action can only be done locally, which could explain this mistake.)

26:53 ARO has given the local orders to the field operator, and RO has finished the isolation steps in E-3.

In this situation, when there is no question about what the problem is, SS is no longer involved in the details. The isolation is quickly performed.

Summary of the most influencing factors affecting performance:

Direct negative influences

- Scenario Complexity
  - Steamline break is masking the tube rupture, and the lack of radiation indications complicates the diagnosis.
- Procedures
  - The E-0 procedure gives unclear guidance for diagnosing SGTR when indications of radiation are lacking. Poor layout of E-3 isolation step (step 3).
- Training
  - There was no training for this specific situation, and training for SGTR without radiation was held several years ago.

Negative influences present

Neutral influences

- Time pressure
  - No observations of time pressure.
- Stress
  - SS seems a bit confused by the absence of activity alarms, and tries to find reasons for it. Otherwise, the entire crew appears calm and structured in their work.
- Communication
  - SS leads the communication, primarily with RO, without having structured meetings. RO adds a lot of relevant information. ARO is passive and quiet.

- Work Processes
  - Good procedure reading, but little cooperation with ARO.
- Interface
  - The computerized interface is new to the crew, but is designed to resemble the home plant's functionality. Large screen display could facilitate overview.
- Experience
  - SS is experienced. RO and ARO are less experienced.

Positive influences

- Team Dynamics
  - SS is good at prioritizing and quickly orders important actions, like closing AFW to SG1. SS makes most decisions, which is efficient. SS is involved in details, but still has a good overview. ARO is a bit passive.

Summary of (a) the observed difficulty or ease that the crew had in performing the action of interest and (b) why the action was easy or difficult, based on observations from DVD review and interview with the crew:

The crew had difficulties with diagnosing the tube rupture because of isolated and failed activity sensors, and because the SGTR diagnosis step in E-0 only mentions activity, not SG level. They also had a hard time finding support in the procedures for transferring to E-3 because they did not thoroughly read the foldout page in ES-1.1.

The crew was relatively quick in their decision to transfer to procedure E-3 when they thought that they had SGTR, even though they did not find support for this in the procedure. One reason is that they generally work quickly: RO works alone and does not wait for answers from ARO, while SS quickly steps in and takes over the work when there are problems. SS is experienced and orders actions without much discussion.
The crew isolated the ruptured SG quickly because they generally work quickly in the procedures. RO focused on the steps in the procedure.

Summary of Data Analysis of Crew L Scenario Complex

HFE: HFE 1B: Failure of the crew to identify and isolate the ruptured steam generator in the complex scenario

Short story of what happened in the selected part of the scenario (written after reviewing DVD, logs, and interview):

<u>Identification</u>

11:00 RO is in E-0 step 19.

15:15 SGTR is mentioned for the first time. The identification occurs when the crew is in E-0 step 21 (Check if SI should be terminated). It is the ARO who first notices a "*large increase in SG1 level*"; however, the expected indication of radiation associated with an SG leakage is lacking, and the RO, who is continuing E-0, interprets a slightly decreasing RCS pressure as a stable level (response in step 21 point c is then assumed true), thus arriving at point e in step 21 (go to ES-1.1, "SI termination").

17:00 The SS brings up new confirmational evidence of tube rupture: pressure increase in SG1, higher than for the other SGs. SS also notices decreasing pressure in RCS (which excludes going to ES-1.1).

18:26 SS orders E-3.

19:13 The RO checks the foldout page of ES-1.1, which has a transfer point to E-3 based on SG level.

Isolation

19:35 RO starts E-3.

24:36 SG1 is isolated (ARO ends a call to FO ordering the local actions of step 3).

The crew missed point (d) in E-3 step 3 (Dispatch an operator to isolate air to AFW pump steam supply valve), possibly because this is not present at the home plant. On the other hand, this was surprising, since the RO and the ARO were sitting side by side and the ARO was writing down a list of necessary local actions. Another explanation, or a contributing factor for the action not to be performed, could be that it was forgotten because the RO was jumping from point to point in the procedure step, looking for the relevant ones to assign to the ARO, rather than strictly following the substeps progression (note: the sequence of points in step 3 does not separate or group control room actions from local actions). The SG was nonetheless isolated.

Summary of the most influencing factors affecting performance:

Direct negative influences

- Scenario Complexity
  - *Steamline break is masking the tube rupture, and the lack of the radiation indications complicates the diagnosis.*
- Procedures
  - The E-0 procedure gives unclear guidance for diagnosing SGTR when indications of radiation are lacking. Poor layout of E-3 isolation step (step 3).
- Training
  - There was no training for this specific situation, and training for SGTR without radiation was only held once several years ago, without the ARO.

Negative influences present

- Work Processes
  - Failure to take action: In the isolation, the crew misses point (d) in E-3 step 3 (Dispatch an operator to isolate air to AFW pump steam supply valve). There is a slight difference between home and simulated plants: the valve is closed manually at the home plant, but is closed by closing the air supply in the simulated plant. However, the action logic is the same.

Neutral influences

- Interface

- The computerized interface is new to the crew, but is designed to resemble the home plant's functionality. Large screen display could facilitate overview.

- Experience
  - SS is experienced, RO is experienced, ARO is less experienced.

Positive influences

- Communication
  - RO and ARO sit close to each other. RO reads the procedures aloud, and does most of the checks. When needed, RO asks the ARO to do the checks, and they communicate clearly. ARO also takes the initiative: communicates increase in SG1 level and checks for radiation before communicating its absence. There were no formal meetings, probably because none were needed, as the crew was well updated on process and strategy by good communication and procedure following.
- Team Dynamics
  - SS seems to keep a good oversight and intervenes when asked. He gives clear directions. RO has a questioning attitude and checks for alternative causes: (1) that SG1 increase is not due to AFW; (2) verifies that it is correct to enter the E-3 procedure by checking the foldout page of the ES-1.1 procedure.

Summary of (a) the observed difficulty or ease that the crew had in performing the action of interest and (b) why the action was easy or difficult, based on observations from DVD review and interview with the crew

The absence of radiation delayed the identification, though the crew was still swift in entering E-3. The plant condition first excluded entering ES-1.1 (RCS pressure slightly decreasing) when the crew was in E-0 step 21; the crew then investigated an alternative cause to the increasing level in SG1 by controlling its isolation from all feedwater. When feedwater was excluded as a cause for increasing SG level, the RO wanted to be sure that ES-1.1 (Stop if Safety Injection) was not the correct path and read the ES-1.1 foldout page. The foldout page pointed to E-3, based on SG level only (E-0 step 19, identification of SG tube rupture, is based on radiation indications only). It is not possible to state whether the SS looked ahead in E-0 at step 24 (a close transfer point to E-3, based on inability to control level in the ruptured SG) or just transferred to E-3 with his/her own knowledge. The decision was, however, confirmed by the RO when looking at the foldout page of ES-1.1.

Summary of Data Analysis of Crew M Scenario Complex

HFE: HFE 1B: Failure of the crew to identify and isolate the ruptured steam generator in the complex scenario

Short story of what happened in the selected part of the scenario (written after reviewing DVD, logs, and interview):

<u>Identification.</u>

11:50 The crew reaches step 21 (good work through E-0 until now). At this time, the RO observes a decrease in PRZ pressure and wants to move to step 22 from the step 21 right column (point c). The SS disagrees, probably attributing the PRZ pressure decrease to the cooldown. The SS talks about having a meeting, but does not initiate one.

15:00 SS orders the ARO to stop all AFW to SG1. SGTR is not mentioned.

16:00 A meeting is started at the RO's request. At 16:27, ARO reports that the SG1 level is rising, even though AFW was closed. This information is not pursued. SS does not follow the meeting procedure (he does not ask the crew for their analysis and suggestions), and instead presents his own plan to stop SI by transferring to ES-1.1.

17:42 SS gives orders to start ES-1.1.

20:00 RO stops one of the charging pumps (part of ES-1.1).

20:50 The SS reads to the crew the foldout page of ES-1.1, which gives instructions to go to E-3 if there is a rising level in one SG. The RO is still in ES-1.1, and starts a discussion about step 5. It takes about three minutes to start E-3 (at about 24:00).

<u>Isolation</u>

23:57 RO starts working with the E-3 procedure.

26:50 Isolation completed.

E-3 was followed in an appropriate and timely manner.

Summary of the most influencing factors affecting performance:

Direct negative influences

- Scenario Complexity
  - Steamline break is masking the tube rupture, and the lack of the radiation indications complicates the diagnosis.
- Procedures
  - The E-0 procedure gives unclear guidance for diagnosing SGTR when indications of radiation are lacking. Poor layout of E-3 isolation step (step 3).
- Training
  - There was no training for this specific situation, and training for SGTR without radiation was only held once several years ago.
- Team Dynamics.
  - *Supervision/management.* (a) From step 21 to the start of E-3, the SS gets very involved in tasks. SS also starts giving orders directly to ARO and calls FO. From this point forward the RO became passive (as the SS is actually working as RO); (b) *Meeting procedure not followed.* The SS calls a meeting but does not start it, seemingly occupied with own thoughts instead. Two minutes later the meeting is started, but the SS develops the strategy without consultation.

Negative influences present

- Stress
  - In the interview, the crew talks about stress caused by the SG1 level increase without radiation and by time pressure due to a late identification, as well as by the fact that the scenario started as a secondary side problem and later became an SGTR.

Neutral influences

- Interface
  - The computerized interface is new to the crew, but is designed to resemble the home plant's functionality. Large screen display could facilitate overview.
- Experience
  - SS is less experienced, RO is experienced, ARO is less experienced.

Positive influences

- Communication
  - Good communication and interaction between RO and ARO until E-0 step 19. ARO gives clear answers and at step 9 promptly informs RO of steamline isolation. Communication is disrupted by SS involvement from step 21 until E-3 is started. Good again under isolation.

Summary of (a) the observed difficulty or ease that the crew had in performing the action of interest and (b) why the action was easy or difficult, based on observations from DVD review and interview with the crew:

The absence of radiation delayed the identification. The crew noticed a rising level in SG1, but never discussed a tube rupture because of the lack of radiation indications. The SS got too involved in the process and stopped the RO to make a correct decision, that is, continuing to E-0 step 22, which would likely have sped up identification (probably by entering E-3 from step 24). The SS also disrupted crew interaction and communication by silencing RO and ARO and centralizing all decisions. The isolation was well performed.

A.4 HFE 1B Failure

Summary of Data Analysis of Crew C Scenario Complex

HFE: HFE 1B: Failure of the crew to identify and isolate the ruptured steam generator in the complex scenario

Short story of what happened in the selected part of the scenario (written after reviewing DVD, logs, and interview):

<u>Identification:</u>

04:02 Crew detects reactor trip. They do not seem to understand that they have a secondary break. SI is actuated automatically and detected shortly afterwards.

12:12 ARO reports in E-0 step 19 (Check if SGs are not ruptured) that the level is rising in SG1, but not in SG2 or SG3, though the crew still leaves step 19 when they do not have activity readings.

13:53 RO has entered E-0 step 21 (Check if SI should be terminated) when ARO says that he will close auxiliary feedwater to SG1 because the level rises quickly. This leads to a number of checks in which the SS is also involved, verifying that main feedwater is closed and that sampling valves from SG1 are open. The SS suggests calling out chemistry, but this is not done.

18:03 RO suggests that they transfer to procedure E-3, but the SS argues that they could also go to ES-1.1 (SI termination) and have a new check there. He does not, however, find the criteria on the ES-1.1 foldout page that would lead them directly to E-3. The SS is hesitant about what to do, and the crew is stuck in discussions.

19:45 SS orders RO to go to ES-1.1, but then says that they can have a meeting first. (Note: it would have been better to take a meeting earlier, before making a decision.)

20:25 Meeting starts and SS changes his mind and decides to go to E-3. SS: "*It doesn't matter how we do it. We can take E-3 directly. We were in step 19, and we say that we have a problem there. We take the E-3 there, instead of going to ES-1.1 now.*"

24:20 RO starts procedure E-3. The transfer to E-3 is a knowledge-based decision since they do not find any procedure step that directly leads them to E-3.

RO does not read aloud the notes and warnings in procedure E-0, and often answers his own questions, not waiting for ARO to answer.

When the problem with the rising SG level begins, there are a lot of discussions, but no structured meeting for decision making. There is a good use of redundant instrumentation, but the crew is stuck in checking, without applying what they know to make a decision. They remain in E-0 step 19 (Check if SI should be terminated) for 10 minutes. When RO says that he wants to go to E-3, SS stops him. SS wants to continue in procedures, but RO does not agree. SS hesitates when procedures tell them to go to ES-1.1, as they feel that they should go to E-3.

<u>Isolation:</u>

25:25 RO starts E-3 step 2 (Identify ruptured SG).

26:15 RO orders ARO to close the turbine-driven AFW pump steam supply valve in the control room and then send a field operator to do a local action. ARO notes the actions on a piece of paper, but does not close the valve immediately. ARO asks RO about other local actions a couple of substeps ahead in the procedure. RO reads these actions and misses the verifications in two of the isolations's substeps 3e (Verify lowdown isolation valves from ruptured SG–CLOSED) and 3f (Verify feed water isolation for ruptured SG). These valves are already closed.

27:09 SS suggests that they close manually operated valves to ensure that SG1 is isolated. He also reminds RO that they are still unsure of the tube rupture diagnosis.

28:00 Isolation steps in E-3 are completed, except for local actions. RO continues in procedure E-3. ARO calls chemistry to sample SGs. (Note: he should have prioritized the local actions that include closing the turbine-driven AFW pump steam supply valve.

It would have been better if the shift supervisor called chemistry himself instead of asking ARO, so as not to delay the operators' work in E-3).

31:10 ARO calls field operator to do local actions for SG isolation.

32:20 ARO closes the turbine-driven AFW pump steam supply valve.

33:15 ARO ends the conversation with the field operator. SGTR is isolated (according to HFE 1B).

The communication between RO and ARO is clear. ARO notes down the local actions to be given to a field operator; RO, however, seems not to read the procedure carefully, and leaves out all warnings and notes, as well as two sub steps of the isolation.

SS interrupts the crew's work with a potentially good idea that ultimately delays the work. He also comments that they are not sure that they have a tube rupture. Instead of helping the crew to focus on quickly isolating the steam generator, he slows them down.

ARO calls out personnel to verify the tube rupture before initiating local actions to isolate ruptured SG. ARO also delays an important isolation action in the control room.

Summary of the most influencing factors affecting performance:

Direct negative influences

- Scenario Complexity
  - Steamline break is masking the tube rupture, and the lack of radiation indications complicated the diagnosis.
- Procedures
  - The E-0 procedure gives unclear guidance for diagnosing SGTR when indications of radiation are lacking. Poor layout of E-3 isolation step (step 3).
- Training
  - There was no training for this specific situation, and training for SGTR without radiation was held several years ago.
- Team Dynamics
  - SS very cautious, and has difficulties giving clear directions. The crew members have not been working together as a team for a long time.
- Work Processes
  - There were some observations that the procedure reading was less than adequate.

Negative influences present

- Stress
  - The crew experiences stress from rising SG levels and not knowing what to do, while the procedures lead them in a direction that doesn't feel right. SS appears indecisive.
- Time pressure

- No observation of operators directly stating that they feel time pressure during the scenario. In the interview, however, they mention feeling time pressure.

Neutral influences

- Interface
  - The computerized interface is new to the crew, but is designed to resemble the home plant's functionality. Large screen display could facilitate overview.
- Communication
  - Both good and bad examples.
- Experience
  - All crew members are experienced.

Positive influences

Summary of (a) the observed difficulty or ease that the crew had in performing the action of interest and (b) why the action was easy or difficult, based on observations from DVD review and interview with the crew:

The crew had trouble deciding on a course of action, even after finding a suspected tube rupture and using the available information to support this diagnosis. It was difficult for the crew to transfer to E-3 because they did not have the indications that normally support this decision (radiation), and because the procedure led them to ES-1.1 (SI termination). They did not find a direct transition from the procedures to E-3. SS hesitated, and did not initiate a meeting quickly enough to analyze the situation and make a decision.

When working in the isolation step in E-3, RO and ARO were not focused on quickly isolating the ruptured SG. This might be the result of inadequate training, but they were also distracted by SS, who might have slowed the crew down by pointing out that the diagnosis was not certain and ordering additional actions.

When isolating the ruptured SG, RO missed two verifications because of inadequate procedure reading, and because ARO's questions about actions further down in the procedures distracted RO.

ARO delayed the action to close the turbine-driven AFW pump steam supply valve in the control room, possibly because the operators' home plant only does this action locally, not in the control room.

Summary of Data Analysis of Crew F Scenario Complex

HFE: HFE 1B: Failure of the crew to identify and isolate the ruptured steam generator in the complex scenario

Short story of what happened in the selected part of the scenario (written after reviewing DVD, logs and interview):

Identification:

05:40 RO: "*We have Reactor Trip, and I take E-0.*"

11:20 RO is in E-0 step 19.

13:15 RO is in E-0 step 21c. RO interprets RCS pressure as stable, and E-0 step 21c then leads to transfer to procedure ES-1.1 (SI Termination).

13:40 SS calls for a meeting.

15:30 RO: "*Why does it increase more in SG1 than in the other ones?*" ARO: "*I have decreased aux feedwater.*" RO: "*OK, I'll continue.*"

17:20 SS: "*OK, we take a new meeting.*"

18:40 ARO orders FO to check if they are blowing steam over the roof.

18:50 RO: "*I ask for opening of sampling SG1.*"

19:10 SS orders RO to go back to E-0 for a new diagnosis.

21:17 RO: "*You don't think we have an SGTR in SG1?*" Crew does not discuss SGTR further.

25:00 RO is at E-0 Step 14 right column c, and performs steamline isolation (note: they already have steamline isolation).

26:30 ARO: "*Important information from FO, the whole turbine building is full of steam, but it is disappearing now. I shall not speculate if it is in connection with your steam isolation. The impression I got was that the whole building had been full of steam, but that it has started to clear away now.*"

27:20 SS: "*Have we seen radiation indications?*" ARO: "*No, it seems like a secondary break then.*" SS: "*But why such a high level in SG 1?*"

29:00 RO is in step 19, and concludes that there are no conditions for SGTR.

29:30 SS: "*Level in SG1 still increasing and we transfer to E-3.*" ARO: "*But we have no activity indication.*" RO talks about pressure, temperature in RC returning, RO is hesitant to follow SS's order on transferring to E-3.

31:00 RO transfers to E-3. The transfer to E-3 is a knowledge-based decision since they do not find any procedure step that directly leads them to E-3.

The checks are not adequately performed in E-0 step 19 (SGTR activity). RO continues with the next step, step 20, before ARO reports on step 19. ARO does not report back clearly on step 19. (Note: this does not affect the diagnosis at this point, but it is not a good way to work.) In E-0 step 21 (about Terminating SI), the crew performs an incorrect assessment of RCS pressure. They have slightly decreasing pressure, but interpret the RCS pressure as stable. The RO suggests transferring to ES-1.1 (SI Termination). At this point the SS calls for a meeting. This is good work practice when transferring from E-0, especially in this situation, where the crew seems unsure about what actually has happened and whether they are on track. However, the meeting does not go well: it does not account for the reported increase in SG level, and it does not lead to a plan or to clear individual orders. The crew partially follows the meeting procedure, but there is no clear ending and no real result, other than to continue with the procedure ES-1.1. After the meeting, RO and ARO continue the discussion about the SG1 and PRZ levels' unexpected behavior. The SS calls for a new meeting, but it dissolves into a combination of discussion and work. They discuss both cooling the RCS and losing level from the RCS as potential explanations for the decreasing PRZ level. ARO calls FO to see if they let out steam, and RO calls for opening of sampling to SG1. The situation is unclear for the crew, and the SS orders RO to return to E-0 for a new diagnosis. The second time in E-0 they close the feedwater to SG1. The RO mentions SGTR to the ARO. In E-0 step 14 (about RCS temperature), RO performs steam isolation as described in the right column of the procedure step (note: they already have steam isolation from the beginning of the scenario). Shortly after the isolation, they receive reports from the FO that there has been steam in the turbine building, but that it is clearing away. Together with an increasing PRZ pressure at this point, this is interpreted as a secondary break that they have now isolated. The SS questions the high SG1 level, which has no explanation. The RO continues in E-0 step 19 (SGTR activity check) and concludes that there is no indication of an SGTR. The SS focuses on the SG1 level and wants to transfer to E-3. E-3 does not fit the RO's or the ARO's understanding of the situation, and some time is used to talk about the indications that don't seem to indicate an SGTR at this point. After a brief talk/discussion, the RO transfers to E-3.

Isolation:

31:00 RO transfers to E-3.

32:30 Crew has decided that it is SG1 that will be isolated.

36:10 SG1 isolated according to completed step 3g. RO continues with step 4.

In E-3 step 2, the RO takes some time to check all items of step 2 and discusses whether it is sufficient to have one indication and then move on in the procedure. The SS pointed out that one indication is sufficient. The remaining part of the isolation went OK.

Summary of the most influencing factors affecting performance:

Direct negative influences

- Scenario Complexity
  - Steamline break is masking the tube rupture, and the lack of radiation indications complicates the diagnosis. The missing activity indications stop the crew from diagnosing SGTR. The simultaneous occurrence of the crew performing steam isolation and getting the information from the FO about steam in the turbine building supports the crew's incorrect assessment of a secondary break.
- Work Processes
  - Poor meeting conduct by the whole crew. Meetings do not have definitive endings, and do not lead to concrete plans. Self-checks are inaccurate, and miss the transfer to E-3 on the ES-1.1 foldout page.
- Team Dynamics
  - SS gives few direct orders, and the orders are not clear.
- Procedures
  - Procedure E-0 gives unclear guidance for diagnosing SGTR when indications of radiation are lacking. Poor layout of E-3 isolation step (step 3).
- Training

- There is no training for this specific situation, and training for SGTR without radiation was held several years ago.

Negative influences present

- Team Dynamics
  - Supervisor tries to bring up the increasing SG1 level, but does not come through with the message. The crew members have worked together for some time, but not for long. SS is new in the crew.
- Stress
  - In the interview the operators say that they experienced stress because the situation was unclear.

Neutral influences

- Interface
  - The computerized interface is new to the crew, but is designed to resemble the home plant's functionality. Large screen display could facilitate overview.
- Experience
  - All crew members are experienced.

Positive influences

Summary of (a) the observed difficulty or ease that the crew had in performing the action of interest and (b) why the action was easy or difficult, based on observations from DVD review and interview with the crew:

The crew has trouble arriving at a diagnosis. The missing activity indication makes them hesitant to diagnose SGTR, and they have trouble combining the available indications into a coherent picture; in the interview, they mention difficulties in linking the different events together. Hypotheses about secondary break seem to prevent the crew from adequately accounting for the high SG1 level. The crew has difficulty establishing a good work process for an unclear situation.

Summary of Data Analysis of Crew J Scenario Complex

HFE: HFE 1B: Failure of the crew to identify and isolate the ruptured steam generator in the complex scenario

Short story of what happened in the selected part of the scenario (written after reviewing DVD, logs, and interview):

Identification:

04:18 Crew detects reactor trip.

10:30 Crew is at E-0 step 19 (SGTR identification).

12:25 RO: "*We are starting to get a slightly rising level in one SG.*"

13:15 RO and ARO conclude in E-0 step 21 (Check if SI can be terminated) that the RCS pressure is decreasing and continue in procedure E-0.

13:20 ARO reports 23% in SG1, and 0% percent level in the other two.

14:40 RO reads E-0 step 24b "Control feed flow to maintain SG level between 10% and 50%." ARO does not answer. RO to SS: "*Seems like we could have a feedwater break.*" RO does not read the right column in step 24b, which would direct them to E-3. RO and SS have a small consultation about the situation. The consultation leads them to send an FO to look for indications of feedwater leakage.

15:50 RO: "*No indication of radiation this time either?*" (RO is at step 25).

18:15 RO is at the last step of E-0, step 32, which is to return to step E-0 14 for new checks.

19:20 FO calls SS and reports that there has been steam in the turbine building (due to the steamline break), and it is now vanishing.

20:10 RO to SS: "*Now we have indication for secondary break, and we go to E-2.*" SS: "*Yes, OK.*"

25:49 Crew talks about the level in SG1 and says it can be an SGTR.

27:00 SS: Points to E-2 step 7 (which is transfer to E-3). SS: "*We take E-3.*"

27:30 Decision to transfer to E-3. This is a knowledge-based decision because SG level is not mentioned in E-2 step 7.

SS takes over the alarm handling on request from the RO (good division of work). In the beginning of the scenario before the crew enters E-0, the RO says, "*It feels like a secondary break.*" When performing E-0 step 19, the crew uses some time to check this step. RO asks ARO twice if there is no activity indicated, and RO asks if sampling valves are open. The crew detects the increasing SG1 level, accompanied by zero level in SG2 and SG3. RO talks again about a possible feedwater break. At E-0 step 25 (secondary radiation), it sounds like RO rules out SGTR based on absence of activity indications. At E-0 step 25b, the crew misjudges the use of the appendix pointed out in the procedure step. They order an FO to apply the appendix, while the main part of the appendix is control room actions. The appendix contains a check of increasing SG level as criteria for SGTR diagnosis. It seems that when SGTR is ruled out, the hypothesis about a secondary break is accepted to explain the differences in the SG levels. When the FO reports back that there has been steam in the turbine building, they decide to transfer to E-2 (Isolation of SG with secondary break). Following the E-2 procedure, no SG with secondary break can be identified. The crew is "stuck" in E-2. RO once more rules out SGTR based on absence of activity indication. The SS suggests transferring to E-3 based on E-2 step 7, which contains a transfer to E-3.

The crew comments on and partly discusses process status and strategy while working through the E-0 procedure. One alternative way of working would be to call for a quick meeting and get an overview of process status and strategic options in a more formal and explicit way. At the point of transferring from E-0 to E-2, and from E-2 to E-3, there is neither checking of the basis for transfer nor evaluation of the chosen strategy. This is the SS's responsibility, but the RO should/could also have suggested such an activity.

<u>Isolation:</u>

28:20 RO is at E-3 step 2. RO says that SG1 is to be considered ruptured.

36:40 RO: SG main steam lines and bypass closed. (SG is isolated).

The work in step 3 is slow. The crew should speed up to proceed to the pressure balance. They use time taking samples (both RO and ARO, communicate unclearly with each other and with the field operator). ARO is inexperienced, and RO sometimes walks over to ARO's work station to help. The SS should have directed the crew to complete step 3 more quickly so that they could get ahead in procedure E-3, towards the pressure balance. Instead, SS asks for details about the ongoing work.

Summary of the most influencing factors affecting performance:

Direct negative influences

- Scenario Complexity
  - Steamline break is masking the tube rupture, and the lack of the radiation indications complicates the diagnosis. The missing activity indications prevented the crew from diagnosing SGTR until very late.
- Procedures
  - The E-0 procedure gives unclear guidance for diagnosing SGTR when indications of radiation are lacking. Poor layout of E-3 isolation step (step 3).
- Work Processes
  - There was a tendency to follow procedures too literally without bringing together the crews' assessment of the situation, and the procedures were not adequately used to check the basis for transferring between procedures.
- Training
  - There was no training for this specific situation, and training for SGTR without radiation was held several years ago. Not all crew members had participated in the previous training on SGTR without radiation.

Negative influences present

- Team Dynamics
  - The SS should have contributed more actively to the difficulties and to the decisions suggested by the RO. The SS performance in maintaining overview, analyzing the situation, and giving directions was less than adequate.
- Communication
  - While working on the isolation, RO and ARO talk past each other, and the orders to the field operator are initially not what they had intended.

Neutral influences

- Time pressure
  - No observation of perceived time pressure.

- Experience
  - SS is experienced. RO and ARO are less experienced.
- Stress
  - No observation of stress.
- Interface
  - The computerized interface is new to the crew, but is designed to resemble the home plant's functionality. Large screen display could facilitate overview.

Positive influences

Summary of (a) the observed difficulty or ease that the crew had in performing the action of interest and (b) why the action was easy or difficult, based on observations from DVD review and interview with the crew:

The crew had difficulties identifying a diagnosis while working through the E-0 procedure, as well as with using the increasing SG level as an indication of SGTR. The missing activity indication was heavily used to rule out the SGTR option, and an early hypothesis about secondary break seems to have led to a fixation on this interpretation. In the absence of a clear diagnosis, the indications were interpreted to confirm a secondary break hypothesis, which resulted in a strategy to "get on with the work." These difficulties were exaggerated by inadequate supervision that did not initiate activities to question or evaluate the current hypothesis and strategy. RO also made two errors in the procedure reading: not reading the right side column of E-0 step 24 and not initiating the control room actions of appendix 22.

The crew has difficulties focusing on quickly isolating the ruptured SG when working in E-3. They become fixated on opening sampling, and, because of poor communication, this takes some time. There were also insufficient directions from the SS.

APPENDIX B

FORM A

APPENDIX B

FORM A

Empirical Test–Item 8. Form A for HRA Team Response **HRA TEAM:** ***team ID***

Ver. 2007-01-21 **METHOD:** ***method name***

HFE: ***HFE ID***

Instructions: For each HFE, fill out a Form A. There are nine HFEs in the two SGTR scenario variants: 1A, 2A, 3A, 4A, 5B1, and 5B2 (see item 5 for the definitions). ***Please make sure that each page has your team ID, method name, the HFE ID (in the header), and a page number.***

Item 1) HEP = _________(provide a mean and uncertainty measures)

Item 2) Provide a summary of the most influencing factors on the crews' behavior with respect to this HFE and why they are important. The description should be in terms of the PSFs, causal factors, and other influence characterizations explicitly identified through the HRA method being used.

Factors relevant to the success and/or failure of the HFE should described, but the factors that may drive the crews to fail are of particular interest. The discussion should reflect the basis for the HEP obtained for the HFE.

This discussion should be in terms of the "factors" or characterizations explicitly identified as important from the application of the HRA method. Use the terminology of the HRA method.

Item 3) Provide a qualitative assessment discussing (a) the perceived difficulty or ease that you predict the crew will have in performing the action of interest (transferring to E-3 and initiating the actions therein) and (b) why the action should be easy or difficult, based on insights from using the HRA method.

Explicitly discuss the difficulties associated with the HFE in operational or scenario-specific terms. How will the driving factors be manifested in the crews' performances? If there are specific factors you expect to see in the crews' performances, such factors should be stated here.

The statements you make here may be conditional. Crews with characteristic ___ would be expected to ______.

You may also include in your assessment other predictions that are not directly based on the application of the HRA method. Please clearly identify these parts of your assessment and discuss.

APPENDIX C

FORM B

APPENDIX C

FORM B

Empirical Test–Response Form B **HRA TEAM:** ***team ID***

METHOD: Method name

2007-03-21

EMPIRICAL TEST OF HRA METHODS

RESPONSE FORM B ("STRUCTURED RESPONSE")

In contrast to Form A's open-ended questions, Response Form B is structured. It is used by the HRA Teams (the teams applying the methods) to report their predictions, and is also used in parts of the data analyses of the Halden simulator experiments. Its purpose is to allow the predictions and the experiment outcomes to be expressed in a common language.

The structure is based on the HERA worksheets (NUREG/CR-6903). It has been simplified by eliminating many items that are not likely to be relevant to reporting the results of predictive analyses or to operations in the control room (particularly in the PSF-related Sections 7 and 8, in Part 2 of the form); when possible, interpret the items according to a predictive perspective, and ignore those items that do not or cannot apply. Note also that this form's structure will be used in the analysis of the crew performance in HAMMLAB (simulator outcomes and findings); in that context, HERA's orientation to reporting experienced events is expected to be useful.

The HRA teams are **asked to express their results separately in "free-form"**—namely, in their response to the separate Response Form A—in:

Item 2) "summary of most influencing factors in the method's terminology"

Item 3) "qualitative assessment in operational or scenario-specific terms"

IMPORTANT for HRA Teams: You are asked to connect your response on these worksheets to the results obtained with the method you are applying; the "comment" space next to the response is intended for this purpose.

However, HRA team predictions that are not supported by the application of the method can still be reported on the worksheet. For instance, your team may predict that a given factor is important but not identify a means within the context of the method to express this importance in the application. In these cases, indicate this by noting "not supported by the method," "somewhat supported by method," or a similar comment.

Also, you may leave fields open or state N/A if a field is not relevant for your method.

Overview of the Response Form B

Part 1. Results on the scenario, all HFEs in the scenario, and their relationship. **Use one Part 1 per scenario variant.**

Part 2. Results for **each HFE. Use a separate Part 2 for each HFE.**

Please fill in the "Team ID," "Method Name," and "HFE" in the page headers (there are three section headers).

Response Form B–Part 1

(one Part 1 per scenario variant)

1. Plant and Event Overview

"Plant": *Test subjects in HAMMLAB. PWR simulator.*

Scenario variant: SGTR scenario–base variant OR complex variant.

2. General Trends across HFEs

General Trends ***No general trends identified***

Indicate any strong, overarching trends or context in the scenario, across the HFEs, and provide a detailed explanation. This section is optional and only used when an issue is seen repeatedly throughout the scenario, to highlight the trend that may not be readily evident across the "part 2"s for each HFE.

☒	Trend	Comment
	Procedures (e.g., repeated failure to use or follow procedures)	
	Workarounds (e.g., cultural acceptance of workarounds contributes to multiple subevents)	
	Strong mismatch (see footnote [18] and indicate which type in comment column)	
	Deviation from previously analyzed or trained scenarios	
	Extreme or unusual conditions	
	Strong preexisting conditions	
	Misleading or wrong information, such as plant indicators or procedures	
	Information rejected or ignored	
	Multiple hardware failures	
	Work transitions in progress	
	Poor safety culture	
	Configuration management failures, including drawings and tech specs, such as incorrect room penetrations, piping, or equipment configurations	
	Failure in communication or resource allocation	
	Other (discuss)	

[18] "Strong mismatch" refers to mismatches a) between operator expectations and evolving plant conditions; b) between communications goals compared to practice; c) between complexity and speed of event compared to training and procedural support; or d) between operator mental model and actual event progression.

3. Dependencies Among HFEs

Table 3.1: HFE Dependency Table[19]

Place the HFE IDs on the top row and in the left column of the pyramid table.

You may ***EITHER***

- *place an "**x**" to indicate that the HFEs are not independent (non-zero level of dependence up to complete dependence)* ***OR***
- *indicate the level with* ***L/M/H/C*** *(for low/medium/high/complete dependence), leaving the cell empty for zero dependence.*

In Table 3.2, comment on the factors that caused the HFEs to be dependent. Use more sheets if needed.

		Subsequent HFE							
	HFE ID	**1**	**2**	**...**					
Preceding HFE	**1**								
	2								
	...								

[19] This HFE dependency table may also be used to report dependencies within HFEs, if the HRA team has decomposed a predefined HFE. In such cases, the dependencies would be across sub-HFEs.

Table 3.2–Dependency Factors for Dependent HFEs in Table 3.1

Add rows if needed.

Dependency Factors to be used in Table 3.2

- Similar task
- Same person/people
- Close in time
- Same location/same equipment
- No independent oversight
- Same cues
- Action prompts next incorrect action
- Similar environmental conditions
- Unreliable system feedback
- Prior human failures on same equipment
- Lack of intervening human success
- Cultural dependency
- Mindset
- Work Practices
- Other (explain)

Row HFE	Column HFE	Affects >1 subsequent HFE ☒	Comment (refer to the dependency factors listed above to justify the assessed level of dependency)

Response FORM B–Part 2

(one Part 2 per HFE)

4. Activities particularly important for success/failure of this HFE

Table 4: Human Cognition/Activity Type Associated with the HFE

What types of activities are particularly important for the success or failure of this HFE (the performance challenges for this HFE that are associated with these activities)? Use the comments column to explain the specific challenges. You may refer to your Form A response.

☒	Activity Type	Comment
	Detection: Detection or recognition of a stimulus (e.g., a problem, alarm, etc.)	
	Interpretation: Interpretation of the stimulus (e.g., understanding the meaning of the stimulus)	
	Planning: Planning a response to the stimulus	
	Action: Executing the planned response	
	Indeterminate	

5. Dominant Error Type for the HFE

Table 5.1: Commission/Omission

Will there be a dominant type for the failure of the HFE (in terms of omission/commission)? If both EOC and EOO are potentially significant in connection with the failure of the HFE, please discuss. If only one type is likely, check just one. Cross-reference your Form A response.

☒	Commission/Omission	Comment
	Error of Commission: An incorrect, unintentional, or unplanned action is an error of commission.	
	Error of Omission: Failure to perform an action is an error of omission.	
	Neither dominates	
	Not addressed by analysis	

Table 5.2: Slip/Lapse/Mistake/Circumvention

Will there be a dominant type of failure for the HFE (in terms of the slip or lapse/mistake/circumvention taxonomy)? You may select more than one type. In your comment, indicate the sub-type (from the list in the table). Cross-reference your Form A response.

☒	Error Type (and sub-types)	Comment
	Slip or lapse: - Response implementation error - Unconscious wrong action or failure to act, wrong reflex, wrong instinctive action - Wrong action or lack of action due to omission of intentional check, insufficient degree of attention, unawareness - Strong habit intrusion, unwanted reversion to earlier plan - Continuation of habitual sequence of actions - Failure to act because focal attention is elsewhere, failure to attend to need for change in action sequence - Omission of intentional check after task interruption - Interference error between two simultaneous tasks - Confusion error (wrong component, wrong unit), spatial disorientation (wrong direction), check on wrong object - Omission of steps or unnecessary repetition of steps in (unconscious) action sequence - Task sequence reversal error - Other slip or lapse (explain)	
	Mistake: - Misdiagnosis, misinterpretation, situation assessment error - Wrong mental model, wrong hypothesis - Failure to detect situation, information overload (indications not noticed, acted upon) - Use of wrong procedure - Misunderstood instructions/information - Lack of specific knowledge - Tunnel vision (focus on limited number of indications, lack of big picture) - Overreliance on favorite indications - Not believing indications/information (lack of confidence) - Mindset/preconceived idea/confirmation bias/overconfidence (failure to change opinion, discarding contradictory evidence) - Overreliance on expert knowledge - Other mistake (explain)	
	Circumvention: In spite of a good understanding of the system (process, procedure, specific context), an intentional breaking of known rules, prescriptions, etc. occurred without malevolent intention. - Administrative control circumvented or intentionally not performed - **Required procedures, drawings, or other references not used** - **Intentional shortcuts in prescribed task sequence** - **Unauthorized material substitution** - **Situations that require compromises between system safety and other objectives (production, personal or personnel safety, etc.)** - Intentional disregard of safety prescriptions/concerns - Other circumvention (explain)	
	Other error type (refer to relevant section of open-ended Form A)	
	Not addressed by analysis	

6. Contributory Plant Conditions

Table 6: Plant Conditions

*Indicate with an "**x**" the plant conditions that contribute to this HFE and/or influence the decisions and/or actions of personnel. Leave a detailed comment, with reference to the documentation of your HRA.*

☒	Plant Condition	Comment
	Equipment failure/malfunction	
	System/train/equipment unavailable	
	Instrumentation problems/inaccuracies	
	Control problems	
	Plant/equipment not in a normal state	
	Plant transitioning between power modes	
	Loss of electrical power	
	Reactor scram/plant transient	
	Other *(identify in comment column)*	
	None/Not Applicable/Indeterminate	

7. Contributory Factors–Positive

Tables 7.1 and 7.2 address the factors identified as important positive drivers of performance. Note that Tables 8.1 and 8.2 address the important negative drivers of performance. For more details on the definitions of these contributory factors, please refer to the Form B Coding Definitions.

Positive Contributory Factors/PSF

Table 7.1 is used to identify the positive contributory high-level factors and their relative importance. For each of these high-level factors, Table 7.2 is used to identify the specific lower-level factors that contribute to this weighting. You will probably find it easier to complete the check-off table (Table 7.2) before assigning these points in Table 7.1.

Note that Section 8 addresses negative contributory factors.

A point system is used to indicate the relative importance of the contributory factors. Examples:

- If one factor is important, assign 100 points.
- If two factors are equally important, assign 50 points to each.
- If two factors are important but one is more so than the other, assign the points accordingly (e.g., 80 points to one and 20 to the other).
- You may allocate points to "other positive PSFs" that you identify in the comment column and discuss in Table 7.2.

The total of the positive points should sum to 100. *This point system allows us to distinguish between an analysis that indicates that only one factor is important, as opposed to an analysis that indicates multiple important factors.*

Table 7.1: Positive Contributory Factors and Relative Importance

Indicate the relative importance of the positive contributory factors by allocating up to 100 points among the PSFs in the following table.

☒	PSF	Positive Points	Comment
	Available Time		
	Stress and Stressors		
	Complexity		
	Experience and Training		
	Procedures and Reference Documents		
	Ergonomics and HMI		
	Fitness for Duty/Fatigue		
	Work Processes		
	Communication		
	Environment		
	Team Dynamics/Characteristics		
	OTHER Positive PSFs not falling in above *(discuss)*		
		TOTAL: 100 points	

Table 7.2: Identified Positive Factors (Sub-Factors of PSFs)

*Indicate the sub-factors that are the basis for the points allocated to the PSFs in Table 7.1. Check all sub-factors that apply to the HFE. Type "**x**" where* □ *appears.*

Indicate whether the sub- factor is selected based on the application of the method or if it is an inference (qualitative analysis or expert judgment). Leave a detailed comment, with reference to your Form A response. This table continues on the next page.

PSF	Positive Contributory Factor	Application of HRA Method vs Inference	Comment
Available Time	□ More than enough time, given the context	□ Method □ Inferred	
	□ Other: *discuss in comment field*	□ Method □ Inferred	
Stress and Stressors	□ Enhanced alertness/no negative effects	□ Method □ Inferred	
	□ Other: *discuss in comment field*	□ Method □ Inferred	
Complexity	□ Failures have single vs. multiple effects	□ Method □ Inferred	
	□ Causal connections apparent	□ Method □ Inferred	
	□ Dependencies well defined	□ Method □ Inferred	
	□ Few or no concurrent tasks	□ Method □ Inferred	
	□ Action straightforward with little to memorize and with no burden	□ Method □ Inferred	
	□Other: *discuss in comment field*	□ Method □ Inferred	
Experience and Training	□ Frequently performed/well-practiced task	□ Method □ Inferred	
	□ Well qualified/trained for task	□ Method □ Inferred	
	□ Other: *discuss in comment field*	□ Method □ Inferred	

PSF	Positive Contributory Factor	Application of HRA Method vs Inference	Comment
Procedures and Reference Documents	☐ Guidance particularly relevant and correctly directed toward the correct action or response	☐ Method ☐ Inferred	
	☐ Other: *discuss in comment field*	☐ Method ☐ Inferred	
Ergonomics and HMI	☐ Unique features of HMI were particularly useful to this situation	☐ Method ☐ Inferred	
	☐ Other: *discuss in comment field*	☐ Method ☐ Inferred	
Fitness for Duty/Fatigue	☐ Optimal health/fitness was key to the success	☐ Method ☐ Inferred	
	☐ Other: *discuss in comment field*	☐ Method ☐ Inferred	
Work Processes	☐ Other: *discuss in comment field*	☐ Method ☐ Inferred	
Supervision/Management	☐ Clear performance standards	☐ Method ☐ Inferred	
	☐ Supervisors properly involved in task	☐ Method ☐ Inferred	
	☐ Supervisors alerted operators to key issue that they had missed	☐ Method ☐ Inferred	
	☐ Other: *discuss in comment field*	☐ Method ☐ Inferred	
Conduct of Work	☐ Quick identification of key information was important to success	☐ Method ☐ Inferred	
	☐ Error found by second checker, second crew, or second unit	☐ Method ☐ Inferred	
	☐ Important information easily differentiated	☐ Method ☐ Inferred	
	☐ Determining appropriate procedure to use in unique situation was important to success	☐ Method ☐ Inferred	
	☐ Complex system interactions identified and resolved	☐ Method ☐ Inferred	
	☐ Remembered omitted step	☐ Method ☐ Inferred	
	☐ Difficult or potentially confusing situation well understood	☐ Method ☐ Inferred	
	☐ Safety implications identified and understood in a way that was important to success	☐ Method ☐ Inferred	
	☐ Acceptance criteria understood and properly applied to resolve difficult situation	☐ Method ☐ Inferred	
	☐ Proper post-modification testing identified and ensured resolution of significant problem	☐ Method ☐ Inferred	
	☐ Other: *discuss in comment field*	☐ Method ☐ Inferred	
Communication	☐ Communications practice was key to avoiding severe difficulties	☐ Method ☐ Inferred	
	☐ Other: *discuss in comment field*	☐ Method ☐ Inferred	
Environment	☐ Environment was particularly important to success	☐ Method ☐ Inferred	
	☐ Other: *discuss in comment field*	☐ Method ☐ Inferred	
Team Dynamics/Characteristics	☐ Extraordinary teamwork and / or sharing of work assignments was important to success	☐ Method ☐ Inferred	
	☐ Exceptional coordination/communications clarified problems during event	☐ Method ☐ Inferred	
	☐ Other: *discuss in comment field*	☐ Method ☐ Inferred	
OTHER Positive PSF Not Falling In Above		☐ Method ☐ Inferred	

8: Contributory Factors–Negative

Table 8.1 is used to identify the negative contributory high-level factors and their relative importance. For each of these high-level factors, Table 8.2 is used to identify the specific lower-level factors that contribute to this importance. You will probably find it easier to complete the check-off table (Table 8.2) before assigning these points in Table 8.1.

A point system is used to indicate the relative importance of the contributory factors. Examples:

- If one factor is important, assign 100 points.
- If two factors are equally important, assign 50 points to each.
- If two factors are important but one is more so than the other, assign the points accordingly (e.g., 80 points to one and 20 to the other).
- You may allocate points to "other negative PSFs" that you identify in the comment column and discuss in Table 8.2.

The total of the negative points should sum to 100. *This point system allows us to distinguish between an analysis that indicates that only one factor is important, as opposed to an analysis that indicates multiple important factors.*

Table 8.1: Negative Contributory Factors and Relative Importance

Indicate the relative importance of the negative contributory factors by allocating up to 100 points among the PSFs.

☒	PSF	Negative Points	Comment
	Available Time		
	Stress and Stressors		
	Complexity		
	Experience and Training		
	Procedures and Reference Documents		
	Ergonomics and HMI		
	Fitness for Duty/Fatigue		
	Work Processes		
	Communication		
	Environment		
	Team Dynamics/Characteristics		
	OTHER NEGATIVE PSFS not falling in above		
		TOTAL: 100 points	

Table 8.2: Identified Negative Factors (Sub-factors of PSFs)

*Indicate the sub-factors that are the basis for the negative points allocated to the PSFs in Table 8.1. Check all sub-factors that apply to the HFE. Type "**x**" where □ appears.*

Indicate whether the lower-level factor is selected based on the application of the method (e.g., by a checklist) or if it is an inference (qualitative analysis or expert judgment). Leave a detailed comment, with reference to your Form A response. This table continues over several pages. Note: LTA stands for "less than adequate."

PSF	Negative Contributory Factor	Application of HRA of Method vs Inference	Comment
Available Time	□ Limited time to focus on tasks	□ Method □ Inferred	
	□ Time pressure to complete task	□ Method □ Inferred	
	□ Inappropriate balance between available and required time	□ Method □ Inferred	
	□ Other: *discuss in comment field*	□ Method □ Inferred	
Stress and Stressors	□ High stress	□ Method □ Inferred	
	□ Other: *discuss in comment field*	□ Method □ Inferred	
Complexity	□ High number of alarms	□ Method □ Inferred	
	□ Ambiguous or misleading information present	□ Method □ Inferred	
	□ Information fails to point directly to the problem	□ Method □ Inferred	
	□ Difficulties in obtaining feedback	□ Method □ Inferred	
	□ General ambiguity of the event	□ Method □ Inferred	
	□ Extensive knowledge regarding the physical layout of the plant is required	□ Method □ Inferred	
	□ Coordination required between multiple people in multiple locations	□ Method □ Inferred	
	□ Scenario demands that the operator combine information from different parts of the process and information systems	□ Method □ Inferred	
	□ Worker distracted/interrupted (W2 198)	□ Method □ Inferred	
	□ Demands to track and memorize information	□ Method □ Inferred	
	□ Problems in differentiating important from less important information	□ Method □ Inferred	
	□ Simultaneous tasks with high attention demands	□ Method □ Inferred	
	□ Components failing have multiple rather than single effects	□ Method □ Inferred	
	□ Weak causal connections exist	□ Method □ Inferred	
	□ Loss of plant functionality complicates recovery path	□ Method □ Inferred	
	□ System dependencies are not well defined	□ Method □ Inferred	
	□ Presence of multiple faults	□ Method □ Inferred	
	□ Simultaneous maintenance tasks required or planned	□ Method □ Inferred	
	□ Causes equipment to perform differently during the event	□ Method □ Inferred	
	□ Subevent contributes to confusion in understanding the event	□ Method □ Inferred	
	□ Other: *discuss in comment field*	□ Method □ Inferred	
Experience and Training	□ Fitness for Duty (FFD) training missing/less than adequate (LTA) (F 124)	□ Method □ Inferred	
	□ Training LTA (T 100)	□ Method □ Inferred	
	□ Training process problem (T 101)	□ Method □ Inferred	
	□ Individual knowledge problem (T 102)	□ Method □ Inferred	
	□ Simulator training LTA (T4 103)	□ Method □ Inferred	

PSF	Negative Contributory Factor	Application of HRA of Method vs Inference	Comment
	☐ Work practice or craft skill LTA (W2 188)	☐ Method ☐ Inferred	
	☐ Not familiar with job performance standards	☐ Method ☐ Inferred	
	☐ Not familiar/well-practiced with task	☐ Method ☐ Inferred	
	☐ Not familiar with tools	☐ Method ☐ Inferred	
	☐ Not qualified for assigned task	☐ Method ☐ Inferred	
	☐ Training incorrect	☐ Method ☐ Inferred	
	☐ Situation outside the scope of training	☐ Method ☐ Inferred	
	☐ Other: *discuss in comment field*	☐ Method ☐ Inferred	
Procedures and Reference Documents	☐ No procedure/reference documents (P 110)	☐ Method ☐ Inferred	
	☐ Procedure/reference document technical content LTA (P 111)	☐ Method ☐ Inferred	
	☐ Procedure/reference document contains human factors deficiencies (P 112)	☐ Method ☐ Inferred	
	☐ Procedure/reference document development and maintenance LTA (P 113)	☐ Method ☐ Inferred	
	☐ Procedures do not cover situation	☐ Method ☐ Inferred	
	☐ Other: *discuss in comment field*	☐ Method ☐ Inferred	
Ergonomics and HMI	☐ Alarms/annunciators LTA (H1)	☐ Method ☐ Inferred	
	☐ Controls/input devices LTA (H2)	☐ Method ☐ Inferred	
	☐ Displays LTA (H3)	☐ Method ☐ Inferred	
	☐ Panel or workstation layout LTA (H4)	☐ Method ☐ Inferred	
	☐ Equipment LTA (H5)	☐ Method ☐ Inferred	
	☐ Tools and materials LTA (H6)	☐ Method ☐ Inferred	
	☐ Labels LTA (H7)	☐ Method ☐ Inferred	
	☐ Other: *discuss in comment field*	☐ Method ☐ Inferred	
Fitness for Duty/Fatigue	☐ Working continuously for considerable number of hours	☐ Method ☐ Inferred	
	☐ Working without rest day for considerable time	☐ Method ☐ Inferred	
	☐ Unfamiliar work cycle	☐ Method ☐ Inferred	
	☐ Frequent changes of shift	☐ Method ☐ Inferred	
	☐ Problems related to night work	☐ Method ☐ Inferred	
	☐ Circadian factors/individual differences (F 127)	☐ Method ☐ Inferred	
	☐ Impairment (F 129)	☐ Method ☐ Inferred	
	☐ Other: *discuss in comment field*	☐ Method ☐ Inferred	
Work Processes			
Planning/Scheduling	☐ Inadequate staffing/task allocation (W1 181)	☐ Method ☐ Inferred	
	☐ Scheduling and planning LTA (W1 180)	☐ Method ☐ Inferred	
	☐ Other: *discuss in comment field*	☐ Method ☐ Inferred	
Supervision/Management	☐ Inadequate supervision/command and control (O1 130)	☐ Method ☐ Inferred	
	☐ Management expectations or directions less than adequate (O1 131)	☐ Method ☐ Inferred	
	☐ Duties and tasks not clearly explained/work orders not clearly given	☐ Method ☐ Inferred	
	☐ Progress not adequately monitored	☐ Method ☐ Inferred	
	☐ Frequent task reassignment	☐ Method ☐ Inferred	
	☐ Pre-job activities (e.g., pre-job briefing) LTA (W1 183)	☐ Method ☐ Inferred	
	☐ Safety aspects of task not emphasized	☐ Method ☐ Inferred	
	☐ Informally sanctioned by management	☐ Method ☐ Inferred	
	☐ Formally sanctioned workarounds cause problems	☐ Method ☐ Inferred	
	☐ Other: *discuss in comment field*	☐ Method ☐ Inferred	
Conduct of Work	☐ Self-check LTA (W2 197)	☐ Method ☐ Inferred	

PSF	Negative Contributory Factor	Application of HRA of Method vs Inference	Comment
	☐ Improper tools or materials selected/provided/used	☐ Method ☐ Inferred	
	☐ Necessary tools/materials not provided or used	☐ Method ☐ Inferred	
	☐ Information present but not adequately used	☐ Method ☐ Inferred	
	☐ Failure to adequately coordinate multiple tasks/task partitioning/interruptions	☐ Method ☐ Inferred	
	☐ Control room sign-off on maintenance not performed	☐ Method ☐ Inferred	
	☐ Tag-outs LTA (W1 184)	☐ Method ☐ Inferred	
	☐ Second independent checker not used or available	☐ Method ☐ Inferred	
	☐ Work untimely (e.g., too long/late) (W2 192)	☐ Method ☐ Inferred	
	☐ Log-keeping or log review LTA (W2 195)	☐ Method ☐ Inferred	
	☐ Independent verification/plant tours LTA (W2 196)	☐ Method ☐ Inferred	
	☐ Procedural adherence LTA (W2 185)	☐ Method ☐ Inferred	
	☐ Failure to take action/meet requirements (W2 186)	☐ Method ☐ Inferred	
	☐ Action implementation LTA (W2 187)	☐ Method ☐ Inferred	
	☐ Recognition of adverse conditions/questioning LTA (W2 189)	☐ Method ☐ Inferred	
	☐ Failure to stop work/non conservative decision making (W2 190)	☐ Method ☐ Inferred	
	☐ Non-conservative action (W2 193)	☐ Method ☐ Inferred	
	☐ Failure to apply knowledge	☐ Method ☐ Inferred	
	☐ Failure to access available ☐ Methods of information	☐ Method ☐ Inferred	
	☐ Situational surveillance not performed	☐ Method ☐ Inferred	
	☐ Incorrect parts/consumables installed/used	☐ Method ☐ Inferred	
	☐ Incorrect restoration of plant following maintenance/isolation/testing	☐ Method ☐ Inferred	
	☐ Independent decision to perform work around or circumvent	☐ Method ☐ Inferred	
	☐ Other: *discuss in comment field*	☐ Method ☐ Inferred	
Problem Identification	☐ Problem not completely or accurately identified (R1 140)	☐ Method ☐ Inferred	
	☐ Problem not properly classified or prioritized (R1 141)	☐ Method ☐ Inferred	
	☐ Other: *discuss in comment field*	☐ Method ☐ Inferred	
Communication	☐ No communication/information not communicated (C 160)	☐ Method ☐ Inferred	
	☐ *Misunderstood or misinterpreted* information (C 51)	☐ *Method* ☐ *Inferred*	
	☐ Communication not timely (C 52)	☐ Method ☐ Inferred	
	☐ Communication content LTA (C 53)	☐ Method ☐ Inferred	
	☐ Communication equipment LTA (C 162)	☐ Method ☐ Inferred	
	☐ Other: *discuss in comment field*	☐ Method ☐ Inferred	
Environment	☐ Temperature/humidity LTA (H10 71)	☐ Method ☐ Inferred	
	☐ Lighting LTA (H10 72)	☐ Method ☐ Inferred	
	☐ Noise (H10 73)	☐ Method ☐ Inferred	
	☐ Radiation (H10 74)	☐ Method ☐ Inferred	
	☐ Work area layout or accessibility LTA (H10 75)	☐ Method ☐ Inferred	
	☐ Postings/signs LTA (H10 76)	☐ Method ☐ Inferred	

PSF	Negative Contributory Factor	Application of HRA of Method vs Inference	Comment
	☐ Task design/work environment LTA (F 126)	☐ Method ☐ Inferred	
	☐ Other: *discuss in comment field*	☐ Method ☐ Inferred	
Team Dynamics / Characteristics	☐ Supervisor too involved in tasks, inadequate oversight	☐ Method ☐ Inferred	
	☐ Crew interaction style not appropriate to the situation	☐ Method ☐ Inferred	
	☐ Team interactions LTA (W2 191)	☐ Method ☐ Inferred	
	☐ Other: *discuss in comment field*	☐ Method ☐ Inferred	
OTHER Negative PSFs not falling in above		☐ Method ☐ Inferred	

End of Form B, Part 2 for this HFE.

APPENDIX D

FORM B INSTRUCTIONS

APPENDIX D

FORM B INSTRUCTIONS

Empirical Test of HRA Methods

RESPONSE FORM B INSTRUCTIONS

Introduction

These instructions are excerpted and extended from the Human Event Repository and Analysis (HERA) Coding Manual (NUREG/CR-6903, Volume 2, in press). As noted on the worksheets, they represent a simplified version of the HERA worksheets, optimized for reporting the findings from HRA methods in a standardized format. HERA has, among other purposes, been designed to encompass a wide range of performance shaping factors that are relevant to control room operations in nuclear power plants. The purpose of using these worksheets is to facilitate the comparison of HRA methods; however, it is realized that some items in the worksheets may be more or less applicable to particular HRA methods and their accompanying analyses. These instructions provide clarification on terms used in the worksheets. Due to the HERA origin of these instructions, they may include terms that do not appear on Form B. If you wish to use such terms, use the field "Other: *discuss in comment field.*"

General Instructions (From Form B)

IMPORTANT FOR HRA TEAMS: You are asked to connect your response on these worksheets to the results obtained with the method you are applying; the "comment" space next to the response is intended for this purpose.

However, HRA team predictions that are not supported by the application of the method can still be reported on the worksheet; for instance, your team may predict that a given factor is important but not identify a means within the context of the method to express its importance in the application. In these cases, indicate this by noting "not supported by the method," "somewhat supported by method," etc.

Also, you may leave fields open or state N/A if a field is not relevant for your method.

Part 1. Results for the scenario, all HFEs in the scenario, and their relationship. Use one Part 1 per scenario variant.

Part 2. Results for each HFE. Use a separate Part 2 for each HFE.

Please fill in the "Team ID," "Method Name," and "HFE" in the page headers (there are three section headers) for this HFE.

FORM B, PART 1, Section 1

Indicate by checking or circling whether the scenario variant involved the base variant or the complex variant.

FORM B, PART 1, Section 2

Circle or check either *General Trends* or *No general trends identified* as applicable.

Denote with a checkmark the predicted trends in the scenario and specifically document these in the *Comment* field. This section is used to illustrate any strong, overarching trend(s), issue(s), or context(s) across HFEs, and should be completed when an issue is seen repeatedly throughout the event, to highlight the trend that may not be readily evident from the separate Part 2 coding. For example, if the HRA method suggests that the scenario involves multiple instances of crews performing workarounds rather than fixing a problem, or if there is a cultural influence that affects all subevents, it could be documented and explained here. Not all items are applicable to the simulator scenario.

FORM B, PART 1, Section 3

The dependency matrix in Worksheet Table 3.1 resembles a correlation matrix, whereby each HFE is listed on both the horizontal and vertical axes and each HFE is paired with another. The approach to dependency in the worksheets offers analysts the opportunity for scaled or nonparametric dependency estimation. Scaled dependency should be listed as zero (Z), low (L), medium (M), high (H), or complete (C) dependency, as covered by the method. The analyst should map the method's dependency scale to this scale as appropriate, providing the method-native value in parentheses. Nonparametric dependency should be indicated simply by a checkmark in the rubric, to indicate that dependency exists between the two HFEs. Nonparametric dependency does not attempt to capture the degree of dependency, just whether or not dependency exists between two HFEs.

For dependent HFEs, the analyst should provide an explanation in Worksheet Table 3.2 regarding the dependency factors. Common dependency factors are listed in the columns above the table, but the analyst should document only those factors that are considered in the HRA method used. The analyst should also indicate (by checking the appropriate box) whether dependency is in effect in more than one subsequent HFE.

FORM B, PART 2, Section 4

For each HFE, indicate in Worksheet Table 4 any human cognition or activity types that are associated with the HFE. This section considers the steps in human information processing or decision making, allowing the analyst to indicate the type of activity the operator is engaged in during the HFE. Indicate and document the following:

- Detection or recognition of a condition or change in situation (e.g., a problem or alarm)
- Interpretation of the condition or change in situation
- Planning a response to the situation
- Executing the response (action)

This information is useful for a variety of HRA methods, which provide separate performance shaping factor weighting values for each category. If the HRA method's treatment of cognition types does not directly map to those provided in Worksheet Table 4, an explanation of the mapping should be provided in the *Comment* field, as appropriate.

FORM B, PART 2, Section 5

This section utilizes two separate error taxonomies for classifying the HFE. The analyst should check the appropriate boxes and document relevant considerations in the *Comment* fields.

Error of Omission/Error of Commission

An *Error of Commission* is an incorrect, unintentional, or unplanned action. This occurs when a person makes an overt action, or performs an incorrect action. An error of commission typically leads to a change in plant or system configuration, resulting in a degraded plant or system state. Examples include terminating running safety-injection pumps, closing valves, and blocking automatic initiation signals.

An *Error of Omission*, on the other hand, is a failure to take a required action, which typically leads to an unchanged or inappropriately changed plant or system configuration, resulting in a degraded plant or system state. Examples include failures to initiate standby liquid control system, start auxiliary feedwater equipment, and isolate a faulted steam generator.

The analyst should indicate which of these two error types applies to the HFE under consideration and provide an appropriate explanatory comment. If *neither dominates*, this should be noted. If this taxonomy is *not addressed by analysis*, this should likewise be noted.

Slip/Lapse/Mistake/Circumvention

This error taxonomy is related to Rasmussen's cognitive control framework, but has been expanded to include circumventions. It is possible for an HFE to involve more than one category of error, so the analyst should select all options that apply. For example, it is common for a circumvention to be made based on an incorrect understanding of the situation (mistake). Definitions are provided below, while examples are embedded directly in Worksheet Table 5.2.

Slips or *Lapses* are the category of errors that occur when a person intends to take the correct action, but either takes a wrong action (a slip) or fails to take the action they intended (a lapse). A slip or a lapse is an unconscious unintended action or failure to act, resulting from an attention failure or a memory failure in a routine activity. In spite of a good understanding of the system (process, procedure, and specific context) and the intention to perform the task correctly, an unconscious unintended action or a failure to act occurs, or a wrong reflex or inappropriate instinctive action takes place. Simple examples would include turning the wrong switch when the correct one is located next to it or inadvertently leaving out a step in a procedure when the intention was to complete the step.

Mistakes are the category of errors that occur when a person is following a plan diligently, but the plan is inadequate for him/her to achieve his/her goal. A mistake occurs when an intended action results in an undesired outcome. Mistakes can be rule-based, as when an inappropriate rule or procedure is selected for a situation or when a good rule is misapplied, or knowledge-based, as when the situation is not fully understood and no rules are available to aid operators in solving the problem.

Circumventions are the category of errors that occur when, in spite of a good understanding of the system (process, procedure, specific context), a person deliberately violates rules, prescriptions, etc, without malevolent intention, usually with the intention of maintaining safe or efficient operations. It is possible for the outcome of such a circumvention to be successful—for example, if the rules did not apply or did not work and creative problem-solving was required—in which case the subevent would likely not be classified as an HFE. However, it is often the case that such circumvention results in a degraded plant condition.

Other error type. When the HRA method uses another error classification, this should be elaborated upon in this section.

Not addressed in analysis should be indicated if the HRA method does not consider this classification scheme.

FORM B, PART 2, Section 6

Identify any plant and equipment conditions that contributed to the HFE. This list, based partially on Halden Reactor Project Report HWR-521 (Braarud, 1998), summarizes plant conditions that contribute to the HFE or influence the decisions and actions of the personnel. If significant plant factors are at play in the HFE but are not listed, the analyst may specify *Other* and provide details in the corresponding text entry field. If *No* conditions apply, this should likewise be indicated.

FORM B, PART 2, Sections 7 and 8

Sections 7 and 8 collect information about PSFs that influenced the HFE. PSFs provide a means of tracing either the detrimental or positive effect on human performance. Section 7 captures positive contributors of performance, regardless of the outcome of the HFE. Section 8 captures negative contributors to performance in the HFE. The eleven HERA PSFs were developed based on the *Good Practices for Implementing HRA* (Kolaczkowski et al., 2005). To the extent reasonable, the analyst is asked to map the HRA method's PSFs to the eleven PSFs provided in Worksheet Tables 7.2 and 8.2; it is recognized, however, that this may not always be feasible. Analysts are encouraged to document additional PSFs in the *Other* fields as required.

The eleven PSFs considered in the worksheets include the following factors.

Available Time refers to the time available to complete a task. In the worksheets, available time considers the time available versus the time required to complete an action, including the impact of concurrent and competing activities.

Stress and Stressors are broadly defined to describe the mainly negative, though occasionally positive, arousal that impacts human performance. A small amount of stress can be beneficial and enhance performance, though it more often contributes to performance detriments. When evaluating the impact of stress as a PSF, analysts should consider workload, task complexity, time pressure, and perceptions of pressure or threat, familiarity with the situation at hand, physical stressors, such as those imposed by environmental conditions (e.g., high heat, noise, poor ventilation, poor visibility, or radiation). Clearly, stress is context-dependent; it is not independent of other PSFs. If other PSFs, such as available time, complexity, training, or fitness for duty are poor, it is probable that stress is elevated. Analysts should consider the situation as a whole, including the other relevant PSFs, when assessing stress as a PSF.

Complexity refers to the difficulty of the task in the given context. It considers how ambiguous or familiar the situation or task is, the number of separate inputs that occur to the operator simultaneously and possible causes, the mental effort and knowledge required, the clarity of cause-and-effect relationships in task performance and system response, the number of actions required in a certain amount of time, and the physical effort or precision required. It also considers the environment in which the task is to be performed, any special sequencing or coordination that is required (e.g., if it involves multiple persons in different locations), the presence and number of parallel tasks or other distractions, and the presence and quality of indications.

Experience and Training includes years of experience for the individual or the crew, specificity of training to the work being performed, quality of training, and amount of time since training. This also includes the frequency of an activity (e.g., routinely vs. rarely) and an operator's familiarity or experience with a specific task or situation.

Procedures and Reference Documents refers to the availability, applicability, and quality of operating procedures, guidance, or reference documents, or best practices for controlling work quality for the tasks under consideration. It can also refer to policies and rules or regulations that govern work at a plant. When assessing the influence of procedures and reference documents on a subevent, analysts should consider the degree to which the available procedures clearly and unambiguously address the situation at hand, completeness, accuracy, the degree to which procedures assist the crew in making correct diagnoses, the extent to which they have to rely on memory, and how easy or difficult the procedure is to read, follow, or implement.

Ergonomics and Human-Machine Interface (HMI) is a broad category that encompasses all aspects of how people interact with the plant systems, equipment, data or information interfaces, instrumentation, and other aspects of their environment. Included in this PSF are the availability and clarity of instrumentation, the quality and quantity of information available from instrumentation, the layout of displays and controls, the ergonomics of the control room or work location, the accessibility and operability of the equipment to be manipulated (e.g., manually opening a valve requires an operator to climb over pipes and use a tool from an awkward position), and the extent to which special physical fitness requirements, tools, or equipment are needed to perform a task. The adequacy or inadequacy of computer software is also included in this PSF.

Fitness for Duty/Fatigue refers to whether or not the individual performing the task is physically and mentally fit to perform the task at that time. This includes such considerations as fatigue, illness, drug use (legal or illegal), physical and mental health, overconfidence, personal problems, time of day, and work schedule.

Work Processes refers to aspects of working, including intra-organizational collaboration, safety culture, work planning, communication, and management support and policies. The Work Processes PSF is divided into four sub-categories:

- *Planning and Scheduling*: Those contributing factors to a subevent that involve planning work activities and scheduling. Work planning includes work package development and ensuring that personnel have enough resources (e.g., tools, materials, or funding) to perform their work. Scheduling includes ensuring that sufficient and appropriate personnel are available to work. It also includes ensuring that personnel do not work too much overtime.
- *Supervision and Management*: Contributing factors to a subevent that involve

supervision of work and organizational or management issues. This includes such factors as command and control, quantity, quality, and appropriateness of supervision, whether work orders or instructions are given clearly, management emphasis on safety, weaknesses and strengths in organizational attitudes and administrative guidance, and organizational acceptance of workarounds.

- *Conduct of Work*: Contributing factors to a subevent that involve performance of work activities, at both the individual and group levels. This includes such factors as procedural adherence, whether work is done in a timely manner, appropriate or inappropriate use of knowledge and available information, recognition of adverse conditions, ability to coordinate multiple tasks, and proper use of tools and materials.
- *Problem Identification and Resolution (PIR)*: All contributing factors to a subevent that involve identifying and resolving problems at a plant. This includes factors such as classification of issues, root cause development, planning and implementation of corrective actions, review of operating experience, trending of problems, individuals' questioning attitudes and willingness to raise concerns, and preventing and detecting retaliation.

Communication refers to the quality of verbal and written interaction between personnel working together as a crew in the simulator. This includes whether the content of communications are clear, complete, are verified and managed in such a way to ensure their receipt and comprehension, as well as whether one can be easily heard.

Environment refers to external factors, such as ambient noise, temperature, lighting, weather, etc, which can greatly influence the personnel ability to carry out their prescribed tasks.

Team Dynamics and Characteristics refers to the crew interaction style and whether it is appropriate to the situation at hand. At first glance, some aspects of this factor are related to the Communication PSF, such as the quality of communication strategies used by the crew and the supervision and conduct of work subcategories of the Work Processes PSF (e.g., style of supervision and procedural adherence). However, this PSF is specific to characterizing the crew as a whole and how the dynamics within or between teams influence performance and event response. Specifically, team dynamics and characteristics include such aspects as the degree to which independent actions are encouraged or discouraged, supervision style (e.g., democratic or authoritarian), the presence of common biases or informal rules, such as how procedural steps are to be interpreted or which steps can be skipped, how well the crew ensures that everyone stays informed of activities or plant status, and the crew's overall approach in responding to an event, whether it's aggressive or slow and methodical (Kolaczkowski et al., 2005). It is important to note that the worksheets do not identify any one type of crew interaction style as "better" than others; the effect of crew characteristics is largely dependent on the situation under analysis and whether the crew dynamics were appropriate to that situation.

Tables 7.1 and 8.1

In these tables, the analysts should indicate the relative importance or weighting of the positive or negative contributory factors, respectively. These tables also allow the analyst to specify which PSFs were considered. The total weighting should equal 100. As noted in the worksheet instructions:

- If only one PSF is important, assign 100 points to that PSF and 0 points to all remaining PSFs.
- If two PSFs are equally important, assign 50 points to each and 0 points to all remaining PSFs.
- And so on.

The point system allows the empirical study comparison team to distinguish between an analysis that considers only one PSF and an analysis that considers multiple PSFs.

Other PSFs may be added as necessary, and should be weighted as appropriate.

Tables 7.2 and 8.2

Each PSF has a number of corresponding contributory factors or PSF details that may come into play when assigning the PSF. The analyst should indicate when one or more factors apply and denote whether this detail was explicitly covered by the method or merely inferred (e.g., through expert judgment). This list is not meant to be exhaustive, but rather represents a list of factors derived from operating experience. In the case where a PSF applies but is not covered by one or more pre-specified contributory factors, the analyst should indicate *other* and discuss it in the *Comment* field.

The contributory factor descriptions are designed to be self-explanatory. In the case where an analyst is not absolutely sure if a contributory factor applies, he/she may indicate this in the *Comment* field.

NRC FORM 335 (9-2004) NRCMD 3.7 U.S. NUCLEAR REGULATORY COMMISSION **BIBLIOGRAPHIC DATA SHEET** *(See instructions on the reverse)*	1. REPORT NUMBER (Assigned by NRC, Add Vol., Supp., Rev., and Addendum Numbers, if any.) NUREG/IA-0216 Vol.1
2. TITLE AND SUBTITLE International HRA Empirical Study-Pilot Phase 1 Report Description of Overall Approach and First Results from Comparing HRA Methods to Simulator Data	3. DATE REPORT PUBLISHED MONTH: November YEAR: 2009 4. FIN OR GRANT NUMBER N6129
5. AUTHOR(S) E. Lois, G. Parry, Nuclear Regulatory Commission Andreas Bye, Halden Reactor Project Vinh N. Dang, Paul Scherrer Institute, Switzerland John Forester, Ron Boring, Sandia National laboratories Jeff Julius, Scientech	6. TYPE OF REPORT Technical 7. PERIOD COVERED *(Inclusive Dates)*

8. PERFORMING ORGANIZATION - NAME AND ADDRESS *(If NRC, provide Division, Office or Region, U.S. Nuclear Regulatory Commission, and mailing address; if contractor, provide name and mailing address.)*

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

INSTITUTT FOR ENERGITEKNIKK
OECD Halden Reactor Proje ct
P.O. Box 173, NO-17 51 Halden, Norway

Sandia National laboratories
P.O. Box 5800, MS0748
Albuquerque, NM, 87185-0748

9. SPONSORING ORGANIZATION - NAME AND ADDRESS *(If NRC, type "Same as above"; if contractor, provide NRC Division, Office or Region, U.S. Nuclear Regulatory Commission, and mailing address.)*

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

INSTITUTT FOR ENERGITEKNIKK
OECD Halden Reactor Proje ct
P.O. Box 173, NO-17 51 Halden Norway

10. SUPPLEMENTARY NOTES

Erasmia Lois, NRC Project Manager

11. ABSTRACT *(200 words or less)*

Volume 1 of NUREG/IA-0216 documents the Pilot Phase of the International Human Reliability Analysis (HRA) Empirical Study. This three-phase study is a multinational, multiteam effort supported by the O rganization for Economic Cooperation and Development (OECD) Halden Reactor Project (HRP), the Swiss Federal Nuclear Insp ectorate, the U.S. Electric Power Research Institute, and the U.S. Nuclear Regulatory Commission (NRC). The obj ective of this study is to develop an empirical understanding of the performance, strengths, and weaknesses of different HRA me thods used to model human response to accident sequences in probabilistic risk assessments (PRAs). The empirical basi s was developed through experiments performed at the HRP's research simulator, with real crews responding to accide nt situations similar to those modeled in PRAs. The scope of the study is limited to HRA methods thought appropriate for use in PRAs evaluating internal events during full power operations of current light water reactors. The Pilot Phase consisted of developing, testing, and revising the study's methodology and design. Phase 2 consists of the comparison of HRA predictions for all nine steam generator tube rupture human actions. Phase 3 consists of the comparison of four loss-of-feedwater hu man actions, as well as documentation of the overall study results. The results of the Empirical Study will provide a techn ical basis for improving individual methods, improving existing guidance documents for performing and reviewing HRAs (e.g., NUREG-1792, HRA Good Practices), and developing additional guidance and training materials for implementing individu al methods.

12. KEY WORDS/DESCRIPTORS *(List words or phrases that will assist researchers in locating the report.)*	
human reliability analysis probabilistic risk assessment human performance HRA data simulator data empirical data empirica study	13. AVAILABILITY STATEMENT unlimited 14. SECURITY CLASSIFICATION *(This Page)* unclassified *(This Report)* unclassified 15. NUMBER OF PAGES 16. PRICE

UNITED STATES
NUCLEAR REGULATORY COMMISSION
WASHINGTON, DC 20555-0001

OFFICIAL BUSINESS