*Proceedings of the 33rd European Safety and Reliability Conference (ESREL 2023)*
*Edited by* Mário P. Brito, Terje Aven, Piero Baraldi, Marko Čepin and Enrico Zio

doi: 10.3850/978-981-18-8071-1_P644-cd

# Review of human error assessment methods suitable for the design of maritime remote control rooms and processes

Danilo Taverna Martins Pereira de Abreu[1], Marcelo Ramos Martins[1], Marcos Coelho Maturana[1], Jakub Montewka[2], Marilia Abilio Ramos[3]

[1] *Department of Naval and Ocean Engineering, Analysis, Evaluation and Risk Management Laboratory (LabRisco), University of São Paulo, Brazil. E-mail: danilo.abreu@usp.br, mrmartin@usp.br, marcos@labrisco.usp.br*

[2] *Ship Design Department, Institute of Ocean Engineering and Ship Technology, Gdansk University of Technology, Gdańsk, Poland. E-mail: jakub.montewka@pg.edu.pl*

[3] *The B. John Garrick Institute for the Risk Sciences, UCLA, Los Angeles, USA. E-mail: marilia.ramos@ucla.edu*

Shipping is facing numerous innovations nowadays that, if pursued, could significantly change how ships are designed, operated and navigated. One of these innovations is the remote control of ships. In this new context, decisions are made outside the controlled vessel, from a remote control centre, and with limited awareness of the vessel and surrounding conditions. To ensure operator performance in remote control centres, they must be designed with a human-centred approach. To this end, in addition to adopting human factors principles, appropriate human reliability analysis (HRA) techniques must be used to reduce human error probability. Currently, several HRA methods exist, both in the scientific literature and in industry standards. However, most methods were developed or tested in domains other than remote maritime. In the lack of an HRA method specific to remote maritime operations, modified versions of nuclear and aviation-based methods and tools are applied. Therefore, aiming at overcoming these limitations, the objectives of this paper are threefold: 1) to review existing methods suitable for designing remote control centres and processes; 2) to shortlist methods found applicable in the maritime context; 3) to elaborate on overall method requirements and research directions.

*Keywords*: autonomous shipping, human reliability analysis, maritime transportation, risk assessment

## 1. Introduction

Maritime Autonomous Surface Ships (MASS) are vessels anticipated to operate in national and international waters at various levels of autonomy, ranging from supervised navigation, through remote control to full autonomy, (Fan et al., 2022). In the most advanced concept designs, human operators are allocated in a shore control center (SCC) and can monitor a group of ships simultaneously, intervening when needed (e.g., in case of automation failure).

Since human intervention is expected to occur in critical situations, it is important to ensure proper and safe interaction between humans and machine. In this sense, several authors are contributing to understanding the human contribution to the risk and safety of MASS operations. Ramos et al. (2018) developed a comprehensive task analysis to facilitate the identification of human-machine errors and applied it to a case study involving autonomous ship collisions. In a more generic sense, Chang et al. (2021) presented a risk assessment for the operation of MASS, which identified human error as an important contributor to the risk. Zhang et al. (2020) developed a probabilistic model of human error assessment using the Technique for Human Error-Rate Prediction (THERP) combined with Bayesian networks.

Despite the recent advances, there is a lot of room for improvement in understanding human performance aspects and developing analysis methods for maritime remote control rooms. Human errors in these operations can be modeled and quantified through human reliability analysis (HRA) methods, which can also provide insights into operational and design aspects to enhance operators' performance and reduce errors' likelihood. Yet, most of the HRA methods have been developed or tested for the nuclear industry. SCC operations can significantly differ from Nuclear Power Plants (NPPs), including unique failure modes and causes of human failures, and these factors may not be correctly modeled and quantified through current HRA methods (Ramos and Mosleh, 2021). Looking to advance in this sense, this paper reviews some of the existing HRA methods' suitability for input into SCCs design and processes. The goal is to shed light on important issues and how the traditional methods available in the literature could provide inspiration and information for developing domain-specific methods.

The paper is organized as follows. Section 2 presents an overview of suitable methods analyzed in the scope of this paper. Section 3 proposes a set of desirable requirements for an HRA method to be applied in the SCC contexts. Section 4 presents the results of the HRA methods evaluation in the context of SCC. Finally, Sections 5 discusses the findings, and 6 concludes the paper.

## 2. Overview of suitable HRA methods

This section presents an overview of the HRA methods investigated in this paper. The methods were selected based on the following criteria: a) the existence of an underlying quantitative framework; b) the availability of the original publication to the authors; and c) adequacy to deal with uncertain human-related safety aspects in SCC design.

Despite the initial filtering, dozens of methods still comply with the above requirements and discussing all of them would not be feasible. Therefore, this paper limits the discussion to a sample of five traditional methods, based on how HRA is typically addressed in the industry. These methods are: *Tecnica Empirica Stima Errori Operatori* (TESEO), THERP, Human Error Assessment and Reduction Technique (HEART), Cognitive Reliability and Error Analysis Method (CREAM), and Standardized Plant Analysis Risk-Human Reliability Analysis (SPAR-H).

### 2.1. *TESEO*

TESEO was proposed in the 1980s using heterogeneous human reliability data from the literature (Bello and Colombari, 1980). Its modelling approach presupposes that the human error probability (HEP) is a multiplicative function of five parameters, accounting for 1) the type of activity to be carried out; 2) the time available to carry out this activity; 3) the human operator's characteristics; 4) the operator's emotional state; and 5) the environmental ergonomics characteristics.

The original publication provides the values to be adopted for the model parameters depending on the state of each factor. The type of activity, for instance, can be classified as "simple, routine", "requiring attention, routine", or "non-routine". In extreme cases, the multiplicative function can lead to HEP values greater than one. In this case, it is assumed that HEP = 1, i.e., the probability of success is null.

### 2.2. *THERP*

The THERP was developed in the Sandia National Laboratories, focusing on the Nuclear Power Plant context (Swain and Guttmann, 1983). The basic tools adopted for modelling are special event trees called "probability tree diagrams", which are a graphical task analysis method used to consider sequences of human actions according to their outcomes (correct or incorrect). The probabilities of each tree branch can be derived from basic and conditional human error probabilities (HEP) estimates provided by the technique. The outputs from the model are expected to be included in traditional probabilistic risk assessment models such as event trees and fault trees.

In addition to the basic modelling framework using single-point estimates, the THERP also includes guidelines for uncertainty treatment. This is done at the level of distributions for the HEPs. The authors propose using the lognormal distribution for this purpose, using the single-point estimate as a median and adopting a range ratio as the difference between the 5th and 95th percentiles.

In order to account for factors that influence human performance, the THERP provides a set of Performance Shaping Factors (PSFs) and discusses how they may modify the HEPs. For some PSFs, such as stress and experience level, multiplier factors can be used to change the nominal HEPs based on the PSFs states. In other cases, it is up to the analyst to assess which PSFs are applicable and infer to which extent they impact human performance.

### 2.3. *HEART*

The HEART was proposed with the objective of providing quantitative guidance for human factors engineers. It attempts to facilitate the improvement of high reliability in human-machine systems (Williams, 2015).

When applying the HEART, the HEP is a result of the product of three factors: a nominal HEP, the effect of error-producing conditions (EPC), and the proportion of the effect of each EPC. The nominal HEP can be obtained by matching the description of the task under analysis and one of the descriptions provided by HEART. The EPCs can be obtained by identifying which items are applicable to the scenario under analysis and extracting the corresponding multipliers from the HEART tables. Finally, the proportions are obtained through expert judgment and are a number between zero and one that represents to which extent a given EPC influences the HEP.

HEART also supports the consideration of uncertainty bounds for the nominal HEPs (NHEPs). For each generic type of task, the 5th and 95th percentile bounds are provided, although no specific probability distribution is reported for modelling the parameter uncertainty.

### 2.4. *CREAM*

The CREAM is a second-generation method that was developed in a context where the available methods focused solely on practical aspects of HRA and lacked solid conceptual basis (Hollnagel, 1998). Therefore, the CREAM provides a more sophisticated modeling process that avoids viewing PSFs simply as factors that multiply a HEP. Instead, it scrutinizes the dependencies among the PSFs for a more realistic approach.

In terms of quantitative analysis, CREAM proposes a basic approach for preliminary screening of human interactions, and the extended approach, which uses the results from the basic approach and refines them where more precision and details are needed. The extended approach supports the quantification of PSF influences (named "Common Performance Condition" or CPC in CREAM).

The HEPs are provided at the level of generic failure types for four cognitive functions: observation, interpretation, planning, and execution. For each generic failure type, CREAM informs the basic value and the lower and upper uncertainty bounds (5th and 95 percentiles, respectively).

### 2.5. *SPAR-H*

Developed to be used in Standardized Plant Analysis Risk (SPAR) assessment, SPAR-H method was published in 2004 (Gertman et al., 2004) as a refinement of the common HRA practices of that time. When compared to its predecessors, it incorporates significant advances from cognitive modelling and stimulus-response fields.

SPAR-H proposes a distinction between diagnosis and action, which generates different nominal NHEPs. The NHEPs are then modified based on the states of eight PSFs, by the use of multiplicative values, generating the HEP value to be adopted. The technique considers eight PSFs in total: available time, stress/stressors, complexity, experience/training, procedures, ergonomics/human-machine interface, fitness for duty, and work processes. The SPAR-H authors acknowledge that previous multiplicative HRA models were prone to inconsistencies since the resulting HEP value could be greater than one. As an alternative, they provide an adjustment formula, which works around the problem.

Regarding the uncertainty analysis, SPAR-H presents a treatment for the parameter uncertainty associated with the HEP estimates. As an alternative to the lognormal distribution suggested by other techniques such as THERP, whose support is not limited to $[0, 1]$, the authors propose the use of a beta distribution. In addition to the fact that its support is limited to the $[0, 1]$ interval, it can mimic the lognormal distribution and allows preserving the overall mean value after the multiplication of PSF values on the NHEP.

## 3. Desirable requirements

This section lists three desirable requirements for a HRA method suitable to analyze SCC operations and provide input into SCC design. The listed requirements correspond to an initial screening of the methods, and additional conditions can be added for further assessment, e.g.,. how well the methods address known issues of HRA.

### 3.1. *Similarities to the autonomous shipping context*

Most HRA methods were developed to satisfy the needs of specific application domains, such as nuclear power plants, aviation, and the process industry. Consequently, important elements of analysis - e.g., human actions taxonomies, human error probabilities, performance shaping factors (PSFs) - reflect the specific aspects of that domain, and their usage in other contexts should be done carefully.

Despite the limitations, it is frequently possible to draw analogies between different fields of application, conditioned to the fact that they have a reasonable level of similarity. For instance, we may consider the following task description extracted from THERP (Swain and Guttmann, 1983): “detecting stuck locally operated valves”. The description was provided for applications in the nuclear power plant context, but locally operated valves are not a component exclusive to this field. It is plausible to use the information provided by THERP in similar scenarios where the task is also applicable. However, even in similar application domains, the analysts should keep in mind that the contents of some technique are conditional to the circumstances in which it was developed. Therefore, they serve as a good initial source of information, but advanced analysis needs to account for the specific context of application and relevant evidences.

In short, what do we mean by “similarity to the autonomous shipping context” of an HRA technique? It refers to the existence of elements in terms of context, human activities, and PSFs that are akin to the autonomous shipping reality. Techniques concerned with human activities in control rooms are particularly interesting, given similarities with the SCC concept.

### 3.2. *Support to the consideration of epistemic uncertainty*

Any probabilistic risk assessment is subject to two types of uncertainty: aleatory and epistemic. The former refers to the inherent randomness of observable quantities, such as the number of errors in performing a task after a given number of demands. In its turn, the epistemic uncertainty is related to the imprecision due to incomplete knowledge and may be reduced as additional information becomes available.

Most autonomous shipping projects are in their initial stages of development. Consequently, it is prudent to assume that several design aspects conceived today will change in the near future, impacting the human reliability aspects significantly. The current lack of knowledge about the real operational and potential emergent conditions means that the epistemic uncertainty is relevant and should be treated properly.

The epistemic uncertainty can be further classified into three categories (Drouin et al., 2017): completeness, model, and parameter. The completeness uncertainty refers to known and unknown elements that contribute to the risk but are not considered in the analysis. The model uncertainty relates to aspects of the risk assessment that can be treated by different modeling approaches, but none of them is known to be dominantly the best. Finally, parameter uncertainty refers to the uncertainty regarding the values of input values adopted in the models, including probability distribution parameters.

The HRA methods mainly deal with parameter uncertainty by adopting uncertainty bounds to the human error probability values. This is a desirable feature for application in a relatively new domain such as autonomous shipping, due to the expected variability among contexts. The other types of epistemic uncertainty (completeness and model) are generally not directly treated but could be inferred by inspecting premises and confronting modeling approaches.

### 3.3. *Consideration of SCC-specific issues*

The SCC operators are expected to face specific issues due to some particularities of the work conditions. These factors are generally modeled in an HRA framework as PSFs, accounting for how much they influence the HEPs. It is necessary that a method used in maritime autonomous shipping adequately models the factors specific to this operation and their impact on operators' performance.

Ramos et al. (2018) present four specific factors that may influence the operators on SCCs significantly. Their relevance will depend on the SCC and the operational design, including how many vessels the operators are expected to monitor, the crew composition, etc. These factors include: a) information overload: an excessive amount of information the operator receives (information availability is important, but after a certain point it impairs the individual's processing ability). This factor can be particularly relevant in case an operator is monitoring multiple vessels; b) situation awareness: the ability of being aware of what is happening in the surroundings and interpreting the meaning of the available information; c) skill degradation: loss of skills due to disuse, mainly observed in high-reliability industries (where accidents are rare) and also as a consequence of automation; and d) boredom: a state of weariness due to lack of stimulation, which may impair work quality.

This is a preliminary list only, and other factors can be added. Examples include automation complacency, when an operator over-trusts the automation, impacting their ability to recognize a failure and react when needed, and automation under-reliance, when operators do not trust the automation and intervene when not necessary by the system (Presley et al., 2022).

## 4. Evaluation of HRA methods

The original publications of each HRA method considered in this paper were analyzed focusing on identifying how and to which extent they meet the desirable requirements listed in Section 3. The results are summarized in Table 1 and discussed in this section.

In terms of similarities to the autonomous shipping context, the original domain of application of each method was identified to check whether they were similar in some sense or not. Generically speaking, the SCCs resemble ordinary control rooms and, therefore, it is expected that several human tasks developed in this context are similar to other industries. From the five techniques, two of them were developed focusing on the nuclear power plant (NPP) domain, which resembles the SCC context: THERP and SPAR-H. The other three were not developed focusing on any specific context. However, the original CREAM publication (Hollnagel, 1998) mentions influences from the nuclear industry on the development of the method. Additionally, HEART is commonly adopted in the process industry.

Regarding uncertainty treatment, four of the five techniques provide guidance on the treatment of parameter uncertainty: THERP, HEART, CREAM, and SPAR-H. The treatment is mostly related to the probability distribution of the HEP value. While THERP and SPAR-H suggest parametric distributions for this purpose (lognormal and beta, respectively), the other methods are limited to providing values for the 90% uncertainty bounds. Other sources of epistemic uncertainty such as the model and completeness uncertainties are barely treated or mentioned.

The direct or indirect treatment of "information overload", "situation awareness", and "boredom" by TESEO was not identified. It could be interpreted that the "skill degradation" is included in the TESEO's operator's typologic factors, which account for personnel knowledge and training level. However, it is not clear if this typologic factor accounts for skill degradation issues, hence the "indirect" label in Table 1.

Despite the quantitative treatment given to specific PSFs, the THERP only discusses qualitatively the factors of interest for this paper. The "information overload" is discussed as part of the complexity PSF, acknowledging that the information load can reach a level at which the operator can no longer process, thus resulting in frequent errors. Furthermore, THERP addresses two components of "situation awareness" as PSFs: perceptual requirements and interpretation. The former is

Table 1. Summary of the HRA methods review

| Method | Domain | Epistemic uncertainty treatment | Treatment to SCC design challenges |
|---|---|---|---|
| TESEO | Generic | Not informed | Information overload: not identified.<br>Situation awareness: not identified.<br>Skill degradation: indirect.<br>Boredom: not identified. |
| THERP | NPP | Parameter uncertainty | Information overload: qualitative.<br>Situation awareness: qualitative.<br>Skill degradation: qualitative.<br>Boredom: qualitative. |
| HEART | Generic | Parameter uncertainty | Information overload: quantitative.<br>Situation awareness: partial, quantitative.<br>Skill degradation: quantitative.<br>Boredom: quantitative. |
| CREAM | Generic | Parameter uncertainty | Information overload: quantitative.<br>Situation awareness: not identified.<br>Skill degradation: indirect, quantitative.<br>Boredom: not identified. |
| SPAR-H | NPP | Parameter uncertainty | Information overload: indirect, quantitative.<br>Situation awareness: not identified.<br>Skill degradation: indirect, quantitative.<br>Boredom: not identified. |

treated from the point of view of how the human-machine interface delivers information to the personnel, while the latter is discussed in a straightforward manner, accounting for the mental processing skills. The “skill degradation” is treated under the “state of current practice or skill” PSF, which accounts for the effect of absence of practice to deal with emergency situations. The authors provide an illustration of the general shapes of curves that represent the effective coping with emergencies of personnel after initial training with and without practice of simulated emergencies. Finally, the “boredom” factor is also discussed in the “monotonous, degrading, or meaningless work” PSF, which accounts for negative effects of very low stress levels.

To a greater or lesser degree, HEART covers quantitatively all the selected SCC design concerns. The “information overload” is treated by the EPC 8, “a channel capacity overload, particularly one caused by simultaneous presentation of non-redundant information”. Regarding the “situation awareness”, only the aspects related to risk perception are addressed. This is done by the EPC 12, “a mismatch between perceived and real risk”. The “skill degradation” is treated by EPC 1, “unfamiliarity with a situation which is potentially important but which only occurs infrequently or which is novel”, which is associated to the highest HEP multiplier factor. Finally, the “boredom” aspect is treated by two distinct EPCs: 1) EPC 28, “little or no intrinsic meaning in a task”; and 2) EPC 34, “prolonged inactivity or highly repetitious cycling of low mental workload tasks”.

The CREAM provides direct quantitative treatment for the “information overload” aspect through a broad CPC, named “number of simultaneous goals”. The specific weight of the CPCs depends on the type of task being developed. The “skill degradation” is not treated directly; however, the “adequacy of training and preparation” CPC considers the impact of refreshing old skills. The original publication did not identify the treatment for “situation assessment” and “boredom” aspects.

In the SPAR-H, the effects of “information overload" are treated indirectly as part of the "complexity" PSF, which includes the mental effort required for a task, among other elements. As in the CREAM, the "skill degradation" is treated indirectly by the "experience/training" PSF, which considers the importance of training in accident scenarios and the time passed since training. Direct or indirect treatment for the other two factors of concern, "situation assessment" and "boredom", was not identified.

## 5. Discussion

At first glance, the results presented in Table 1 suggest that the HEART method complies with most of the requirements desired for an HRA method to support SCC design. Its generic aspect encourages the application in different domains. Additionally, the HEART enables the treatment of parameter epistemic uncertainty and its set of EPCs cover some of the most important particular issues of SCC design. Therefore, although its development dates back to the decade of 1980, it could serve as a starting point for developing an HRA method for the autonomous shipping context.

However, contributions from other methods as well as recent developments can not be neglected. For instance, the THERP provides interesting discussions regarding different PSFs, despite not providing quantitative treatment for all of them. The second-generation methods, SPAR-H and CREAM, provide underlying cognitive modeling, which enhances the comprehension of the human cognitive process and types of errors.

This paper was limited to analyzing the methods in their original formats but most of the techniques were augmented and improved in recent years. By incorporating resources such as Fuzzy Logic and Bayesian Networks, it is possible to improve the uncertainty treatment and consideration of dependencies (Martins and Maturana, 2013; Zhou et al., 2018). There are also important advances in terms of collecting data to improve the empirical basis of the methods (Jung et al., 2020) and aggregating data from different sources in prospective models (Maturana et al., 2021).

Furthermore, the methods selected for the analysis were restricted to "traditional" methods. These methods present shortcomings that can lead to a lack of traceability and reproducibility (Mosleh and Chang, 2004). More recently proposed methods, such as IDHEAS (Chang and Xing, 2016) and Phoenix (Ekanem et al., 2016), and third-generation approaches (Groth et al., 2019), should be assessed for their suitability for SCC operations. The discussions on state-of-the-art HRA, including best practices, data collection initiatives, model-based HRA, dependency assessment, and HRA methods for digital control rooms, are led by the Nuclear industry community. These discussions must be leveraged in defining a technical roadmap to a credible HRA method for maritime remote control rooms.

Finally, HRA allows for identifying, modeling, and quantifying human errors and their causes. It can provide important input to the design of SCCs so that human errors can be prevented or mitigated. Yet, it must be combined with important studies from the human factors engineering field when providing input to SCC design. Moreover, it can be combined with other methods such as Systems-Theoretic Accident Model and Processes (STAMP) and System-Theoretic Process Analysis (STPA) (Leveson, 2011), Functional Resonance Analysis Method (FRAM) (Hollnagel, 2012) for a more holistic analysis.

## 6. Conclusion

The MASS are a disruptive shipping technology and understanding the associated risks is a requirement for safe operations. This paper reviewed traditional HRA methods with the objective of elucidating how they approach important questions concerning human performance in remote control of MASS. The analysis demonstrated that the five selected methods selected could contribute to a greater or lesser extent, and the HEART method may provide an adequate foundation. Future analyses include expanding the methods' selection to include recently developed methods (IDHEAS, Phoenix, and other 3rd generation methods); adding analysis criteria concerning methods' abilities to overcome known traditional HRA limitations; and expanding the list of factors specific to MASS operation.

An HRA methodology for MASS operations also needs validation against empirical data, since the data used in traditional HRA methods may not apply to the SCC context. While historical data will not be available in the near and medium future and expert judgment may be limited by insufficient operational experience, data can be collected through simulation centers.

**Acknowledgement**

Jakub Montewka appreciates financial support received from National Center for Research and Development in Poland project A tool for strategic risk assessment for the waterways based on the principles-driven method (ASTRA), number MarTERA4/1/10/ASTRA/2023, under MarTERA programme.

## References

Bello, G. and V. Colombari (1980). The human factors in risk analyses of process plants: The control room operator model 'TESEO'. *Reliability Engineering 1*(1), 3–14.

Chang, C.-H., C. Kontovas, Q. Yu, and Z. Yang (2021). Risk assessment of the operations of maritime autonomous surface ships. *Reliability Engineering & System Safety 207*, 107324.

Chang, J. and J. Xing (2016). The general methodology of an integrated human event analysis system (idheas) for human reliability analysis method development. *PSAM 13*.

Drouin, M., A. Gilbertson, G. Parry, J. Lehner, G. Martinez-Guridi, J. LaChance, and T. Wheeler (2017). Guidance on the treatment of uncertainties associated with pras in risk-informed decisionmaking. Technical report, U.S. NRC.

Ekanem, N. J., A. Mosleh, and S.-H. Shen (2016). Phoenix – a model-based human reliability analysis methodology: Qualitative analysis procedure. *Reliability Engineering & System Safety 145*, 301–315.

Fan, C., J. Montewka, and D. Zhang (2022). A risk comparison framework for autonomous ships navigation. *Reliability Engineering and System Safety 226*, 108709.

Gertman, D., H. Blackman, J. Marble, J. Byers, C. Smith, et al. (2004). The SPAR-H human reliability analysis method. Technical report, US Nuclear Regulatory Commission.

Groth, K. M., R. Smith, and R. Moradi (2019). A hybrid algorithm for developing third generation hra methods using simulator data, causal models, and cognitive science. *Reliability Engineering & System Safety 191*, 106507.

Hollnagel, E. (1998). *Cognitive reliability and error analysis method (CREAM)*. Elsevier.

Hollnagel, E. (2012). *FRAM, the functional resonance analysis method: modelling complex socio-technical systems*. Ashgate Publishing, Ltd.

Jung, W., J. Park, Y. Kim, S. Y. Choi, and S. Kim (2020). HuREX–A framework of HRA data collection from simulators in nuclear power plants. *Reliability Engineering & System Safety 194*, 106235.

Leveson, N. G. (2011). *Engineering a safer world: Systems thinking applied to safety*. The MIT Press.

Martins, M. R. and M. C. Maturana (2013). Application of Bayesian Belief networks to the human reliability analysis of an oil tanker operation focusing on collision accidents. *Reliability Engineering & System Safety 110*, 89–109.

Maturana, M. C., M. R. Martins, and P. F. F. F. e Melo (2021). Application of a quantitative human performance model to the operational procedure design of a fuel storage pool cooling system. *Reliability Engineering & System Safety 216*, 107989.

Mosleh, A. and Y. Chang (2004). Model-based human reliability analysis: prospects and requirements. *Reliability Engineering & System Safety 83*(2), 241–253.

Presley, M., J. Julius, A. Wright, K. Gunter, and E. Collins (2022). Updating hra for digital environments. In *Proceedings of the 32nd European Safety and Reliability Conference (ESREL 2022)*, pp. 689–697.

Ramos, M. A. and A. Mosleh (2021). Human role in failure of autonomous systems: A human reliability perspective. In *2021 Annual Reliability and Maintainability Symposium (RAMS)*, pp. 1–6. IEEE.

Ramos, M. A., I. B. Utne, and A. Mosleh (2018). On factors affecting autonomous ships operators performance in a shore control center. *Proceedings of the 14th Probabilistic Safety Assessment and Management, Los Angeles, CA, USA*, 16–21.

Swain, A. D. and H. E. Guttmann (1983). Handbook of human-reliability analysis with emphasis on nuclear power plant applications final report. Technical report, Sandia National Laboratories.

Williams, J. C. (2015). HEART — a proposed method for achieving high reliability in process operation by means of human factors engineering technology. *Safety and Reliability 35*(3), 5–25.

Zhang, M., D. Zhang, H. Yao, and K. Zhang (2020). A probabilistic model of human error assessment for autonomous cargo ships focusing on human–autonomy collaboration. *Safety science 130*, 104838.

Zhou, Q., Y. D. Wong, H. S. Loh, and K. F. Yuen (2018). A fuzzy and Bayesian network CREAM model for human reliability analysis–The case of tanker shipping. *Safety science 105*, 149–157.