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_P645-cd

On the Use of Simulators to Gather Human Performance Data of Remote Maritime Operations

Danilo Taverna Martins Pereira de Abreu

Analysis, Evaluation and Risk Management Laboratory (LabRisco), University of São Paulo, Brazil. E-mail: danilo.abreu@usp.br

Marcos Coelho Maturana

Analysis, Evaluation and Risk Management Laboratory (LabRisco), University of São Paulo, Brazil. E-mail: marcos@labrisco.usp.br

Marcelo Ramos Martins

Analysis, Evaluation and Risk Management Laboratory (LabRisco), University of São Paulo, Brazil. E-mail: mrmartin@usp.br

The maritime industry is witnessing profound changes in the way that their assets are operated. With the rapid advance of new generation telecommunication technologies, it is practicable to design systems that are deployed on the sea and operated remotely from a shore control center (SCC). This tendency has several advantages such as reducing the human exposure to harsh environments and reducing transportation costs of workers from/to the workplace. By doing so, the physical systems tend to have a higher level of automation, but the human operators are not removed from the loop at all; instead, they are moved to a different location and interact with the systems in a different manner. To achieve high standards of safety, recent research focused on understanding the operators' tasks on SCCs and assessing the potential errors and the corresponding risks. However, as in other industries, obtaining human performance data is a challenging task. One promising alternative is to use training simulators for this purpose. Aiming at contributing to the knowledge of these applications, this paper presents a discussion on the main challenges and interesting solutions regarding the use of simulators to collect human reliability data.

Keywords: Human Reliability Analysis, Simulators, Remote Operations, Risk Analysis.

1. Introduction

The exploration of coastal and oceanic waters can bring enormous economic benefits for the humanity but was always marked by significative difficulties. For instance, the production of oil and gas in deep and ultra-deep waters faces harsh environmental conditions, difficulty of access due to large distances from shore, and potential safety issues with significative consequences (Onugbolu et al. 2012). The maritime shipping is another example of this trade-off as it contributes to a huge part of the international commerce while imposing important personnel, environmental, and financial risks.

In response to these questions, the advent of new technologies related to robotics, data-signal transmission, and teleoperations offers interesting solutions to the stakeholders. Due to the complex challenges, the most popular solutions appear to be the semi-autonomous ones, where human operators are included in the loop to provide cognitive assistance and support to the decision-making process (Shukla and Karki 2016; Wróbel, Gil, and Chae 2021). In this way, these sociotechnical solutions try to benefit from the best of two worlds: on one hand, the flexibility of human cognitive processes; on the other hand, the endurance of physical systems that can surpass harsh environments.

However, when employing human workers in the loop, the obtained flexibility does not come at no cost. As is already known in several industries, the natural variability of human performance can be associated with spurious modes of operation and potential chains of errors. Therefore, when designing the semi-autonomous operations and even the layout of remote-control rooms, it is important to consider the impact of human reliability matters.

One of the main approaches to this issue is to apply the human reliability analysis (HRA). It comprises a vast set of techniques for the systematic identification and analysis of the causes, consequences, and contributions of human failures in sociotechnical systems (Mkrtchyan, Podofillini, and Dang 2015). One of the main areas of application relates to the development of probabilistic risk assessment to comply with regulatory needs. When normative requirements are not a question, HRA is still useful to provide insights regarding the uncertainties related to human performance and the main points of improvements in a given context.

The knowledge about human reliability evolved considerably in the past four decades. Efforts from high-risk sectors (with honorable mention to the nuclear industry) provided advances in terms of modelling the human error phenomena, understanding of the cognitive processes, quantification of human error probabilities (HEP), and evaluation of factors that influence the human performance. These advances are traditionally grouped into three generations (Norazahar 2020; Ade and Peres 2022). The exact definition and limits of each generation vary slightly among authors but are, nonetheless, useful for comprehending the main attainments along the time and what still needs to be done.

Groth, Smith, and Moradi (2019) presented a thorough listing of the current requirements of third generation HRA methods, including aspects related to comprehensiveness, research base, adaptability, flexibility, and support to multi-purpose applications. Particularly, regarding the research base, one of the factors that still impacts the quality of HRA usage is the lack of empirical human performance data. Most of the methods rely on the use of expert elicitation, which provides an interesting starting point for quantification methods but should be used with caution due to the inherent human biases and difficulties of dealing with probabilistic reasoning (Caverni, Fabre, and Gonzalez 1990; Kahneman and Tversky 2013).

One alternative to improve the empirical base for human performance data collection is related to the use of simulators. In several domains, the simulators are typically used to the training of human operators. Several aspects of the real world are reproduced with a high level of fidelity, serving as a fair basis for what the human operators should expect in their daily routines or emergency scenarios. There are already initiatives in the nuclear industry related to development of databases and data collection frameworks based on the observations of activities executed at control room simulators – e.g., the Scenario Authoring, Characterization, and Debriefing Application (SACADA) database (Chang et al. 2014), and the human reliability data extraction (HuREX) framework (Jung et al. 2020).

Acknowledging the importance of the data gathering questions discussed above, this paper presents a discussion towards the use of simulators to gather human performance data for the application of HRA on the domain of maritime remote operations. The main content is divided into three parts. Section 2 presents a discussion on how to choose the right type of simulator; section 3 discusses the costs of constructing and maintaining a simulator; finally, section 4 addresses the data validity questions.

2. Choosing the right type of simulator

A simulator is a piece of equipment that is designed to represent real conditions (Cambridge Dictionary 2022). The representation is limited and capable of reproducing only selected aspects of the real physical system. Consequently, when performing training activities or observing the performance tendency in each scenario, the users should be aware that some aspects of the reality may not be reproduced with fidelity.

Despite these limitations, the use of simulators has beneficial outcomes in several industries, such as offshore (Wilkinson 2016), healthcare (Kirkman et al. 2014), aviation (Hays et al. 1992), and nuclear (Joe and Kovesdi 2021). This is because the simulators are developed to represent only the useful aspects of the physical system or process of interest. Thus, the applicants can achieve the specific training objectives or observe the phenomena of interest.

A reasonable first step to successfully apply simulators to any field of application is the selection of an adequate type of equipment, including hardware and software features. The International Atomic Energy Agency (IAEA) provides an interesting classification of several simulator types for nuclear power plant control rooms, which is useful to understand the variance of fidelity and complexity among the most common options (IAEA 1998; 2004):

- Full scope simulator: a simulator that incorporates detailed modelling of the physical systems and a replica of control room operating consoles. Given their specificities, the full scope simulators can be used for personnel training of a specific plant.
- Part task simulator: a simulator developed to train operators on specific aspects of the system. The specific phenomena may be simulated more accurately than in a full scope simulator. Examples include the startup of diesel generators and operation of other plant components.
- Basic principle simulator: a simulator dedicated to demonstrating and illustrating the main principles behind the physical system. It is useful to support trainees on learning the fundamental operations. The human-machine interface can be provided by simple desktop computers or equivalent.
- Compact simulator: a simulator that provides means to train procedures on a simplified form. The modelling depth and fidelity are equivalent to a full scope simulator, but the human-system interface (HSI) is simplified (e.g., computer-driven displays with touchscreens, mouse-control of graphical equivalents of typical controllers).

Therefore, the types of simulators vary essentially in terms of costs and validity for a given purpose – which are further discussed in sections 3 and 4, respectively. A full scope simulator has an elevated degree of fidelity, which can be useful for training operators for a specific type of plant. However, it is expensive and may not always be the best option depending on other use objectives. If an analyst wants to study different HSIs, a generic compact simulator with reprogrammable displays can be a better option.

3. Costs of construction and operation

When selecting a simulator concept, the development of cost-benefit analysis is essential to ensure that the research or training objectives can be achieved without unnecessary expenses. To do this, it is important to analyse the expected costs carefully for each option available.

It is useful to divide the expected costs in two types: capital expenditures (CAPEX) and operational expenditures (OPEX). The CAPEX accounts for the resources used to acquire and adapt assets such as properties and equipment. The OPEX is related to the expenses to maintain the facility during its lifetime. Table 1 presents an overview of the main capital and operational expenditures related to the construction and operation of a simulation center. It is based on the comprehensive listing presented by Kurrek and Devitt (1997) and is expanded with other common costs inferred from typical applications.

Table 1. Overview of CAPEX and OPEX for simulator construction and operation

Capital expenditures
• Equipment (e.g., computers, basic software, server, keyboards, levers, joysticks, audio devices). • Property (i.e., rooms to deploy the equipment). • Property renovation and adaption. • Development of simulator software. • Personnel costs.
Operational expenditures
• Administration. • Rent. • Equipment maintenance. • Property maintenance. • Internet, energy, and water bills. • Maintenance of software developed in house. • Third-party software license. • Insurance.

The content of Table 1 mainly reflects the costs associated with full scope simulators, which contains important expenses related to property acquisition (or rental) and management, in addition to significant equipment costs. These costs may not be applicable to other concepts, such as the basic principle and compact simulators.

Furthermore, it is important to note that some simulation centres can generate revenues as they are able to provide courses and corporate trainings (Kurrek and Devitt 1997). Therefore, even if the center is not 100% sustainable as would be desirable in a non-profit organization, it is possible to reduce the net operating costs depending on how the simulator is used.

4. Validity of data obtained from simulator studies

The proposal of using simulators to collect human performance data aims at overcoming the challenge of quantifying HRA models. The data scarcity inherent to this field of study as well as concerns about the potential subjectivity of experts' opinions motivate the search for empirical and statistically relevant data sources. However, at this point, important questions emerge. How representative is the data collected in simulators? Furthermore, is it feasible to collect statistically significant samples?

4.1. *Data fidelity*

The simulators represent the elements of the real world in a limited way to a lower or higher degree – see Section 2. Despite the important contributions to achieve training objectives, not all aspects of the operational conditions to be faced by human workers are reproduceable with fidelity. For example, the simulation of offshore emergency situations can be performed using a desktop-based software capable of creating credible virtual emergency scenarios (Musharraf et al. 2019). In this way, the trainees are expected to gain knowledge about emergency procedures, therefore achieving the training objectives. It is also possible to collect data regarding human error probabilities and performance shaping factors (PSFs) using evidence from such experiments, as demonstrated by Musharraf, Khan, and Veitch (2019). PSFs are elements internal or external to the operators that may influence their performance (e.g., fatigue, training, human-machine interface, environmental conditions). Still, the human operators that take part in the simulations will obviously not be exposed to the dangers of real emergency conditions, which could change their actual responses. There are natural limits for what is in fact observable in simulators.

When viewing the simulators as a source of information to HRA, another important aspect to take into consideration is the variance of data collected from different installations. As an example, let's take hypothetical simulators for shore control centres (SCC) of maritime autonomous surface ships (MASS). Different organizations may develop several SCC simulators with the purpose of training their operators to operate specific classes of MASS. At a high level of abstraction, we could say that all of these simulators pertain to the same domain of application and, therefore, data from all of them could be grouped in a large database on a joint industry project. From another point of view, at a lower level of abstraction, each installation may have its own particularities – e.g., HSI, ship types – and it could be argued that the data points collected from multiple installations can be incompatible among them.

Therefore, the analysts developing HRA should always be careful when using data from databases to ensure that the degree of representativeness is adequate. It may be interesting to group data from different installations in some cases but only if they have similar features that make it reasonable to do so. The main objective behind the grouping of different databases is generally to increase the sample size, allowing an improvement of the statistical representativeness. There are interesting ways in which this can be achieved, as will be discussed in the next section.

4.2. *Statistically significant data*

The use of simulators to gather human performance data faces has an important problem: each simulation trial is generally expensive and time consuming. Consequently, it is difficult to collect statistically significant amounts of data to quantify the HEPs. Furthermore, when trying to observe the combinations among PSF states to understand their joint effect on the human performance, it is also hard to obtain significant samples. How to overcome this challenge?

Regarding the quantification of HEPs, it is useful to define a convenient level of abstraction to perform the data collection. The traditional HRA models typically divide the human tasks in several levels of abstraction that range from high-level tasks to more elementary cognitive activities, as illustrated in Figure 1 (the specific nomenclature may vary slightly among different groups of authors). At the lowest level, types of human errors are appended to each task, accounting for the different ways that the failure of a cognitive activity may occur.

When quantifying the HEP for a given task, Eq. 1 is commonly adopted, where HEP_i is the human error probability for the i-th task, n_i is the number of opportunities to perform the task, and m_i is the number of errors observed.

$$\mathrm{HEP}_i = \frac{m_i}{n_i} \tag{1}$$

As pointed out by Jung et al. (2020) the number of opportunities to observe a given task is relatively low for high-level tasks and is higher for low-level tasks and elementary cognitive activities. Again, using the tasks performed in a SCC for MASS, the "high-level task" of avoiding a collision is expected to be observed only a few times in a simulation campaign; yet the cognitive activities of information acquisition, decision-making, and execution of steering commands are observed several times. Therefore, it is easier to ensure significative quantity of data if the observations are made in a lower level of abstraction, since there are more observation opportunities. However, to do this, it is necessary to ensure meaningful definitions for these low-level human tasks, whose should also be replicable among different studies.

Another concern refers to the collection of data to inform the influence of PSFs on the human performance. When trying to account for the influence of PSFs on the HEPs, the number of potential scenarios (defined here as a combination of PSF states) grows exponentially with the number of PSFs and their states. Taking the simplified model of Figure 2 as an example, we have a human task that is influenced by three PSFs. Each PSF has two possible states. The total number of scenarios is $2^3 = 8$. If one more PSF with two states is added, the number of possible scenarios doubles, as we would have 16 combinations. If one of the three PSFs had three states, the number of scenarios would increase significantly too – from 8 to 12 possibilities.

If ensuring a significant number of observations for the HEP estimation alone is a challenging task, contextualizing the HEPs along with the PSFs is even more expensive and time consuming. Therefore, we need also to find a method of ensuring that the gathering of PSF data points is also consistent and feasible.

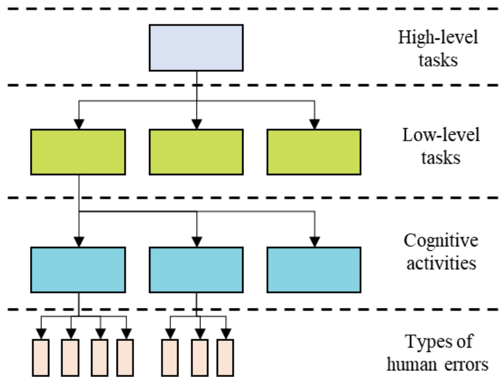


Figure 1. Levels of abstraction for HRA tasks and types of error

In this regard, Groth and Mosleh (2012) present a data-informed PSF hierarchy for model-based HRA, which addresses important quantitative issues. Firstly, the idea of standardizing a PSF set is important, since it facilitates the data gathering from different sources, using similar definitions. In the same sense, each PSF must have objective direct or indirect metrics, allowing for the removal of subjectivity in the data gathering process. Furthermore, since the PSF set is hierarchic, it could be expanded for detailed qualitative analysis and collapsed for quantitative analysis.

If the data gathering is guided by some type of modelling framework, it is also possible to reduce the effort by assuming plausible modelling hypotheses. For instance, if a Bayesian network model is used and it is reasonable to assume that the parent nodes of a child node are independent, then the noisy-or gate is an interesting option. In this case, only n values are sufficient to fulfil the conditional probability tables (CPTs), where n is the number of parent nodes (Chen and Huang 2014). Another alternative is to collect data values only for extreme combinations of PSFs (e.g., all

PSFs in their worst or best states) and use them as anchors to apply interpolation rules (Mkrtchyan, Podofillini, and Dang 2016) – see Figure 3. This approach presupposes that the influence of parent nodes states on the child nodes is strictly positive – i.e., an improvement on the PSF states only increases the success probability of a task, while a deterioration on the parent node states only reduces the probability of success.

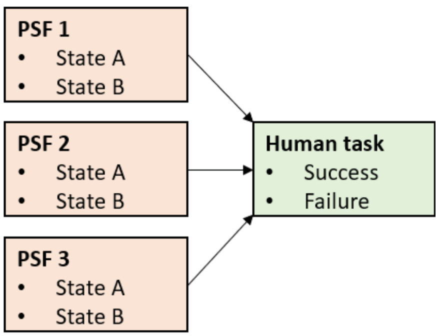
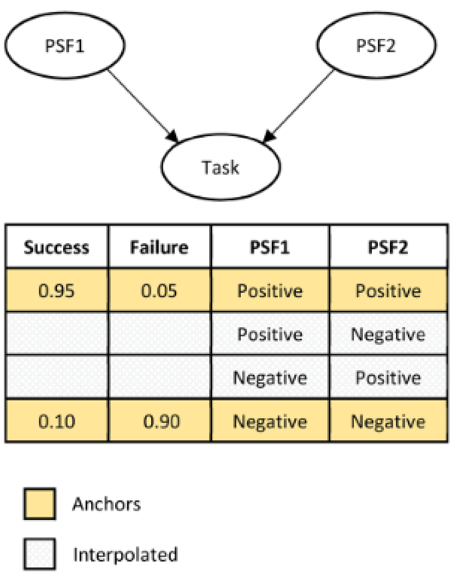


Figure 2. Example of HRA model with three PSFs and one human task



Success	Failure	PSF1	PSF2
0.95	0.05	Positive	Positive
		Positive	Negative
		Negative	Positive
0.10	0.90	Negative	Negative

Figure 3. Interpolation of Bayesian network CPTs using anchors (Abreu et al. 2022)

5. Conclusion

The use of simulators can be an important tool to develop safer remote maritime operations. In addition to their inherent capabilities of providing operator training, they can serve as a source of human performance data, allowing improving the basis for estimation of HEPs and the contributions of PSFs. Through these resources, we may achieve continuous improvement and design safer systems. However, there is no panacea when it comes to the quantification of HRA models. It is important to acknowledge the limitations and challenges involving the use of simulators and try to deliver ways to overcome them.

This paper presented a brief discussion of what are some of the main concerns that should be addressed: a) the selection of appropriate simulator concepts to support the HRA studies; b) the costs of constructing and maintaining simulation facilities; and c) the validity of data gathered from simulation trials. While recognizing that each problem is complex in its own ways, the authors hope that the discussion presented above serves as a first step to HRA practitioners who are challenged by the emergency of semi-autonomous systems.

References

Abreu, Danilo T.M.P., Marcos C. Maturana, Enrique Lopez Droguett, and Marcelo R. Martins. 2022. "Human Reliability Analysis of Conventional Maritime Pilotage Operations Supported by a Prospective Model." *Reliability Engineering and System Safety* 228 (December). https://doi.org/10.1016/j.ress.2022.108763.

Ade, Nilesh, and S. Camille Peres. 2022. "A Review of Human Reliability Assessment Methods for Proposed Application in Quantitative Risk Analysis of Offshore Industries." *International Journal of Industrial Ergonomics* 87 (January): 103238. https://doi.org/10.1016/j.ergon.2021.103238.

Cambridge Dictionary. 2022. "Simulator." 2022. https://dictionary.cambridge.org/dictionary/english/simulator.

Caverni, Jean-Paul, Jean-Marc Fabre, and Michel Gonzalez. 1990. "Cognitive Biases: Their Contribution for Understanding Human Cognitive Processes." *Advances in Psychology*

68: 7–12. https://doi.org/10.1016/S0166-4115(08)61311-4.

Chang, Y. J., Dennis Bley, Lawrence Criscione, Barry Kirwan, Ali Mosleh, Todd Madary, Rodney Nowell, et al. 2014. "The SACADA Database for Human Reliability and Human Performance." *Reliability Engineering and System Safety* 125: 117–33. https://doi.org/10.1016/j.ress.2013.07.014.

Chen, Wei, and Shuping Huang. 2014. "Human Reliability Analysis for Visual Inspection in Aviation Maintenance by a Bayesian Network Approach." *Transportation Research Record*. National Research Council. https://doi.org/10.3141/2449-12.

Groth, Katrina M., and Ali Mosleh. 2012. "A Data-Informed PIF Hierarchy for Model-Based Human Reliability Analysis." *Reliability Engineering and System Safety* 108 (December): 154–74. https://doi.org/10.1016/j.ress.2012.08.006.

Groth, Katrina M., Reuel Smith, and Ramin Moradi. 2019. "A Hybrid Algorithm for Developing Third Generation HRA Methods Using Simulator Data, Causal Models, and Cognitive Science." *Reliability Engineering and System Safety* 191 (November). https://doi.org/10.1016/j.ress.2019.106507.

Hays, Robert T., John W. Jacobs, Carolyn Prince, and Eduardo Salas. 1992. "Flight Simulator Training Effectiveness: A Meta-Analysis." *Military Psychology* 4 (2): 63–74. https://doi.org/10.1207/s15327876mp0402_1.

IAEA. 1998. "IAEA-TECDOC-995 Selection, Specification, Design and Use of Various Nuclear Power Plant Training Simulators."

———. 2004. "IAEA-TECDOC-1411 Use of Control Room Simulators for Training of Nuclear Power Plant Personnel." International Atomic Energy Agency.

Joe, Jeffrey C., and Casey Kovesdi. 2021. "Using Qualified On-Site Nuclear Power Plant Simulators in Human Factors Validations of Control Room Upgrades." In *Advances in Simulation and Digital Human Modeling*, edited by J.L. Wright, D. Barber, S. Scataglini, and S.L. Rajulu, 171–78. https://doi.org/10.1007/978-3-030-79763-8_20.

Jung, Wondea, Jinkyun Park, Yochan Kim, Sun Yeong Choi, and Seunghwan Kim. 2020. "HuREX – A Framework of HRA Data Collection from Simulators in Nuclear Power Plants." *Reliability Engineering and System Safety* 194 (February). https://doi.org/10.1016/j.ress.2018.07.036.

Kahneman, Daniel, and Amos Tversky. 2013. "Prospect Theory: An Analysis of Decision Under Risk." In *Handbook of the Fundamentals of Financial Decision Making*, edited by Leonard C. MacLean and William T. Ziemba, 99–127. World Scientific. https://doi.org/10.1142/9789814417358_0006.

Kirkman, Matthew A., Maria Ahmed, Angelique F. Albert, Mark H. Wilson, Dipankar Nandi, and Nick Sevdalis. 2014. "The Use of Simulation in Neurosurgical Education and Training." *Journal of Neurosurgery* 121 (2): 228–46. https://doi.org/10.3171/2014.5.JNS131766.

Kurrek, Matt M., and J. Hugh Devitt. 1997. "The Cost for Construction and Operation of a Simulation Centre." *Canadian Journal of Anaesthesia* 44 (11): 1191–95. https://doi.org/10.1007/BF03013344.

Mkrtchyan, L., L. Podofillini, and V. N. Dang. 2016. "Methods for Building Conditional Probability Tables of Bayesian Belief Networks from Limited Judgment: An Evaluation for Human Reliability Application." *Reliability Engineering and System Safety* 151 (July): 93–112. https://doi.org/10.1016/j.ress.2016.01.004.

Mkrtchyan, L., L. Podofillini, and V.N. Dang. 2015. "Bayesian Belief Networks for Human Reliability Analysis: A Review of Applications and Gaps." *Reliability Engineering & System Safety* 139 (July): 1–16. https://doi.org/10.1016/j.ress.2015.02.006.

Musharraf, Mashrura, Faisal Khan, and Brian Veitch. 2019. "Modeling and Simulation of Offshore Personnel during Emergency Situations." *Safety Science* 111 (January): 144–53. https://doi.org/10.1016/j.ssci.2018.07.005.

Musharraf, Mashrura, Allison Moyle, Faisal Khan, and Brian Veitch. 2019. "Using Simulator Data to Facilitate Human Reliability Analysis." *Journal of Offshore Mechanics and Arctic Engineering* 141 (2). https://doi.org/10.1115/1.4042538.

Norazahar, Norafneeza Binti. 2020. "Human Factors Risk Assessment." In *Methods in Chemical Process Safety*, 4:289–302. https://doi.org/10.1016/bs.mcps.2020.02.005.

Onugbolu, Olisaka, Carlos Rodrigues, Roberto Maueler, David Wayne Lewis, and Charles Loyola. 2012. “Deepwater Drilling Operations Challenges of Exploration Well in Alpha Block, Gulf of Guinea: Lessons Learned.” In *Nigeria Annual International Conference and Exhibition, Lagos, Nigeria, August 2012*. SPE. https://doi.org/10.2118/162959-MS.

Shukla, Amit, and Hamad Karki. 2016. “Application of Robotics in Offshore Oil and Gas Industry-A Review Part II.” *Robotics and Autonomous Systems* 75 (January): 508–24. https://doi.org/10.1016/j.robot.2015.09.013.

Wilkinson, John. 2016. “The Use of Simulators for on and Off-Shore Training for Control Room Operators and Drillers.” In *Hazards 26*.

Wróbel, Krzysztof, Mateusz Gil, and Chong Ju Chae. 2021. “On the Influence of Human Factors on Safety of Remotely-Controlled Merchant Vessels.” *Applied Sciences (Switzerland)* 11 (3): 1–17. https://doi.org/10.3390/app11031145.