Advances in Reliability, Safety and Security, Part 5 – Kołowrocki, Kosmowski (eds.)
© 2024 Polish Safety and Reliability Association, Gdynia, ISBN 978-83-68136-17-3 (printed), ISBN 978-83-68136-04-3 (electronic)

Advances in Reliability, Safety and Security
ESREL 2024
Monograph Book Series

Lessons Learnt From Human-In-Loop Experimental Scenarios For Advanced Automation In Process Safety

Chidera Winifred Amazu[a,b], Joseph Mietkiewicz[b,c], Ammar N. Abbas[b,d], Gabriele Baldissone[a], Micaela Demichela[a], Davide Fissore[a], and Maria Chiara Leva[b]

[a]*Politecnico di Torino, Turin, Italy*
[b]*Technological University Dublin, Dublin, Ireland*
[c]*Hugin Expert A/S, Aalborg, Denmark*
[d]*Software Competence Center Hagenberg (SCCH), Hagenberg im Mühlkreis, Austria*

Abstract

Operators in process control rooms play an essential role in monitoring and controlling the processes of safety-critical plants, such as chemical plants, power plants, and oil refineries. These operators must be able to interact effectively with the systems to ensure safe operations. However, organisational, technical, and personal factors are essential in operators' ability to ensure plant stability. This paper builds upon the outcomes and observations of an experimental study to address the critical evaluation gap in human system interfaces (HSIs), which serve as decision-support tools within control room environments. The study compared three different setups of human system interfaces in four human-in-the-loop (HITL) configurations, incorporating two alarm design formats (Prioritised vs non-prioritised) and three procedural guidance setups (e.g. one presenting paper procedures, one offering digitised screen-based procedures, and lastly an AI-based procedural guidance system). In this follow-up work, the authors present the lessons learned from the experiment, emphasising the aspects that lead to a deeper understanding of the dynamic interactions between the human operator and the features of the Human Machine Interface (HMI), also considering the criticality of the situation, the time available and the support provided for decision making and response actions. This paper aims to provide actionable knowledge related to design features for supporting decision-making in process control rooms, considering the possibilities also offered by enhanced data analysis and AI tools.

Keywords: design of experiment, decision support, AI, human system interfaces, human error

1. Introduction

To ensure process safety in Seveso Sites, operators in control rooms interact with several system interfaces, including mimic diagrams, comprising static and dynamic information, the alarm list, the charts and trends of selected variables over time and much more. In both regular and abnormal situations, they rely on their expertise, training, procedural guidance, knowledge of their teammates, and other support tools. This emphasises the importance of human, organisational and technical factors in supporting the operators during everyday scenarios and critical situations to ensure optimal situational awareness, manageable mental workload, and good baseline conditions for actions.

In this context, the research underpinning decision support systems and their different formats, such as AI-based support systems, procedure formats, and interface displays, is often within specific industries aimed at the design of Distributed control system interfaces and the one published mainly focused on comparing one format to another but not investigating the impact on operators or the contribution of individual factors to the overall picture (Xu et al., 2008; Shi et al., 2021). Different tools have been explored to make these comparisons, including questionnaire-based assessments like NASA-TLX, biometric-based methods like Eye tracking, and neurometric methods like electroencephalogram (EEG). So far, there have not been many publications related to setups that simulate a combination of these support tools in an ecologically valid scenario able to mimic safety-critical situations, such as the ones related to possible alarm flood conditions, etc.

In a recent experimental study by the authors, three distinct human system interfaces were considered in four human-in-the-loop (HITL) configurations, incorporating two alarm design formats (Prioritised vs non-prioritised) and three procedural guidance setups (e.g. one presenting paper procedures, one offering digitised screen-based procedures, and lastly an AI-based procedural guidance system). The goal was to observe how those different configurations would impact the workload, situational awareness and performance of the participants. This paper builds upon the outcomes and observations of the experimental study to address some possible gaps in human system interfaces (HSIs) and decision support tools within control room environments. The focus of this paper is not on the comparative analysis of these groups. Instead, we delve into the nuanced learnings and findings that emerged during the experiment. This paper seeks to contribute practical knowledge that can inform decision-making for process control room optimisation and management. The insights gained are considered helpful for safety and efficiency in process control environments. By sharing these lessons, we hope to provide actionable recommendations that can be implemented in similar settings to improve operator performance and overall system safety.

2. Method

The experimental study used four human-in-the-loop configurations with different alarm and intervention procedure support formats (Amazu et al., 2023). The first group operated under baseline conditions, the alarms were not arranged according to their criticality and procedural guidance was only provided on a paper document. In contrast to the first group, the second group had the alarms prioritised but with procedures similar to those of Group 1. Group 3 had access to procedures on the interface display. The goal was to determine if this digital access to procedures could enhance operational efficiency and faster response times compared to traditional paper methods. The last group, Group 4, represented the most advanced support level, incorporating an AI decision support system (DSS). The focus was to measure AI's additional benefits in improving operator performance, reducing errors, and enhancing overall system safety and efficiency, especially compared to Group 3's digital procedures. The description of the AI-based DSS is detailed in (Mietkiewicz et al., 2023a, 2023b; Mietkiewicz and Madsen, 2023)

2.1. Simulator

Our experiment utilised a simulator replicating a formaldehyde production facility, focusing on evaluating a DSS in a control room-like environment. The simulator was designed to mirror the complexities of a real-world plant, producing a 30% formaldehyde solution at 10,000 kg/hr through methanol partial oxidation. It featured several key sections forming the mimics the operators had to perform intervention actions on Tanks, Methanol, Compression, Heat Recovery, Reactor, and Absorption units, each integral to the production process (Figure 1).

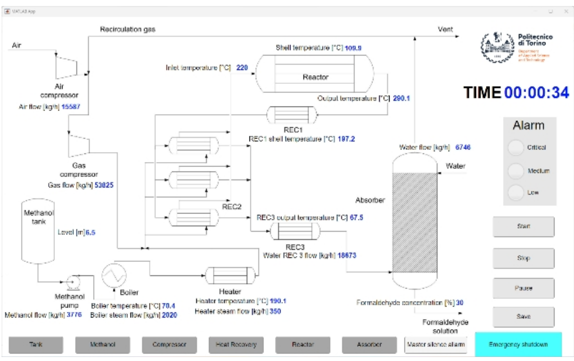

Fig. 1. Main screen of the simulator.

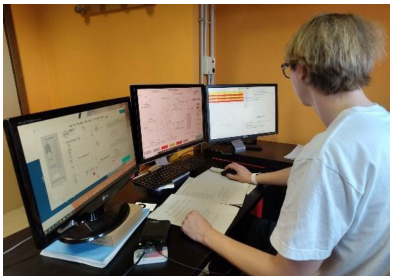

Fig. 2. Participant running one scenario.

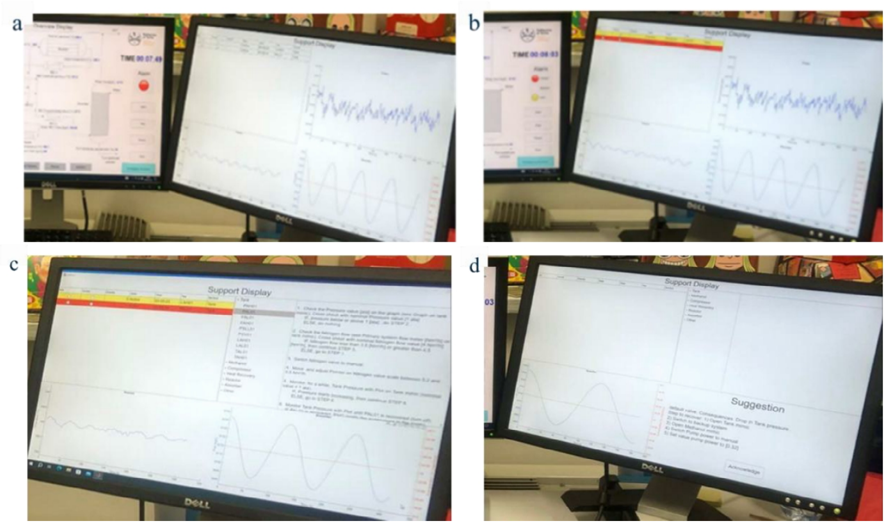

Fig. 3. (a) Right monitor (support display Group 1); (b) Right monitor (support display Group 2);
(c) Support display Group 3; (d) Support display Group 4.

A central aspect of our experimental setup was the "Support interface display," a unique addition that contained the alarm lists and prioritisation in the required cases, the digital procedure for group 3 participants and the extra AI support for the set-up of group 4 participants (Figure 3). This panel included tools presenting the alarms that were activated out of the 80 different alarms in the plant, the procedures, and the trend graphs of critical systems. The design and arrangements presented on this interface were crucial in creating a realistic control room environment for the participants.

To rigorously test the DSS's effectiveness, we selected three safety-critical initiating events that can be considered to be commonly associated with this type of plant. The scenarios were ranked according to an increasing level of complexity:

- Scenario 1: Failure of the methanol storage tank pressure indicator controller, resulting in the operator having to control the nitrogen valve manually.
- Scenario 2: the sequence of events is prompted by the failure of the nitrogen valve itself, which requires first troubleshooting if the valve works and then alternatively switching to a backup system.
- Scenario 3: here, the sequence of events is prompted by the failure of the temperature indicator controller between the heat recovery system and the absorber. This scenario also included the failure of the manual control slider for the cooling water system, which alternatively meant that the operator would manually monitor the reactor temperature and ensure that the reactor was overheating until the other systems could be fixed.

These scenarios were designed to be ecologically valid for testing the various DSS configurations across various situations, from direct manual interventions to complex problem-solving and adaptive decision-making under stress. The aim was to assess the system's utility in enhancing operator performance and the overall resulting system performance in terms of process safety.

2.2. Participants

The experimental study involved 92 individuals, 36 females and 56 males, who voluntarily participated (Figure 2). The participants' pool comprised junior process engineers drawn from the chemical engineering master's students and a few staff members of Politecnico di Torino. The age range of the participants was quite broad, spanning from 21 to 61 years, with a standard deviation of SD = 5.4 years.

The varying levels of participants' experience allowed us to compare Chemical Engineering final-year master students with professionals with backgrounds in process safety and automation. The insights collected from the participants were also valuable in understanding how individuals with different levels of familiarity with control room environments and chemical engineering principles interacted with and benefited from the various DSS configurations.

3. Human error (collated from recorded observations)

This section delves into a range of observations noted during the experiment. Spanning a total duration of two hours, the experiment was structured into two distinct phases: the first hour dedicated to training and the second hour focused on navigating through the three scenarios. Throughout this period, various noteworthy events and atypical behaviours were observed, offering valuable insights into the dynamics of control room operations and decision-making processes.

First, we detail the human errors and violations observed during the task. We also decided to discuss the errors reported from our design of experiment in the initial testing phase, and how we observed them to have affected the participants. We discuss the consequences of the most relevant failures. The errors and violations during the task are presented based on the definition of human error (slips, lapses, mistakes) and Violations given by James Reason (Reason, 2011).

3.1. During task

Here, we outline some of the slips, lapses, mistakes and violations observed by the participants during the test.

- Slips: The operator opened, used or followed the wrong suggestion procedure, valve, etc, or identified the wrong alarm.
  - Some participants in Group 1 used the wrong procedure to intervene in the scenario because they did not know precisely which alarm was more critical. However, they revisited the correct procedure after having tried and failed to resolve the issue.
  - It was observed that participants sometimes accessed incorrect procedures due to inaccurately identifying the names of the alarms.
  - Several participants also did not immediately notice the AI-based DSS suggestion.
- Lapses: in this case, the operator forgets to act.
  - There were cases where the participants forgot to acknowledge the critical alarm after training and practice on the proper operating procedure.
  - While using the AI DSS, some participants inadvertently skipped crucial steps, failing to solve the issues.
  - Participants frequently overlooked the acknowledgement step required for the AI system.
- Rule-Based Mistakes and Violations: When participants decide not to follow the suggestion, use the procedure or do the correct thing.
  - Communication was especially a part of the third scenario and key to succeeding in the task. However, even though the task step 'Call Supervisor' was mentioned in the procedure or AI support, the operator sometimes failed to call the supervisors.
  - Participants, uncertain of the consequences of their actions, often perceived their efforts as ineffective. This uncertainty led to their decision to deviate from the recommended actions, engaging in alternative, sometimes incorrect, procedures.
  - Participants were instructed to focus solely on critical alarms during the experiment. However, it was observed that they often addressed other alarms as well, mainly when no critical alarms were present.

- Knowledge-based mistakes: When the participant is not able to interpret the situation correctly or plan the correct action.

Following this initial classification of participants' errors, we also decided to discuss the pitfalls in the design of the experiment that might have directly or indirectly influenced the participants' behaviour during the testing phase.

3.2. Issues in the design of the experiment during the testing phase

The procedures, interface displays, and alarm displays were optimised following a pre-test usability study. The original issues from the design were addressed, and the designs were updated before the main study phase. Recommendations are given later based on the learnings from the usability study and during the test. The outcomes presented in this section are observations collected during the test connected with some of the key design issues, and how they influenced the participants' behaviour.

3.2.1. Alarm display design

The lack of alarm prioritisation impacted primarily those in group 1. This resulted in trial and error in choosing the critical alarm to intervene on. The time-based and behavioural performances also showed that this group had more prolonged reaction and response times than others. They also were prone to opening more mimics than required and acknowledging more alarms than the others because they lacked an awareness of what was critical.

Thus, having alarms prioritised with the correct colour schemes familiar to mental models, like red indicating critical, orange indicating medium, and yellow indicating low, made a big difference for the other groups.

3.2.2. Display: colour coding

In an earlier design, the emergency button and an 'all alarm silence button) were placed side by side with the same colour schemes for both. This resulted in a participant inadvertently clicking the emergency button instead of the 'All alarms silence button'.

Also, after changing the emergency button to a blue colour, an incident observed during the experiment involved a unique case with one participant, who inadvertently clicked the emergency shutdown button when they intended to close the mimic interface. Notably, the 'Close' and 'Emergency Shutdown' buttons were designed in blue (see Fig. 4.), leading to this confusion. This incident highlights the importance of colour coding and distinct design elements in control room interfaces.

The similarity in colour and possibly the proximity of these two crucial buttons could easily lead to such errors, especially under high-pressure situations where operators are required to make quick decisions. This observation suggests that interface design in control rooms should focus on functionality and preventing potential human errors. Colour schemes and their use play a vital role in helping operators quickly distinguish between different functions and commands, as pointed out in the paper by Mietkiewicz et al. (2023).

3.2.3. Display: feedback mechanisms

Below we propose a short list of the key issues that needed to be addressed during testing regarding the interface design and the feedback mechanism between the simulator and the participants.

Suggestion acknowledgement button: To follow the progressive use of AI-based support, an acknowledgement button was introduced, which had to be clicked by the participant as a way of acceptance. An active suggestion came with a purple-coloured acknowledgement button; the button became white upon acknowledgement. This button was unavailable from the onset, which did not help the operators and researchers. Adding such feedback helped reduce agitation from the participants, who mostly resolved to follow the digitised procedure out of frustration in some instances.

End of test pop-up button: the participants also complained that there was no pop-up button to notify them of a success or failure of their actions. This was further implemented, and when tested, the participants perceived it to help let them know if their task was done. Such features can be added throughout the task after key actions are completed to give the operators feedback. Other cues, such as alarms, can be helpful, but alarms are already overwhelming for operators. Hence, more natural language integrated with pop-up buttons can serve as a better cue.

Using distinct colours, shapes, and spatial arrangements can significantly reduce the likelihood of accidental presses or misinterpretations. Additionally, these incidents pointed out the need to have regular usability testing and feedback loops with actual operators to identify and rectify such design flaws.

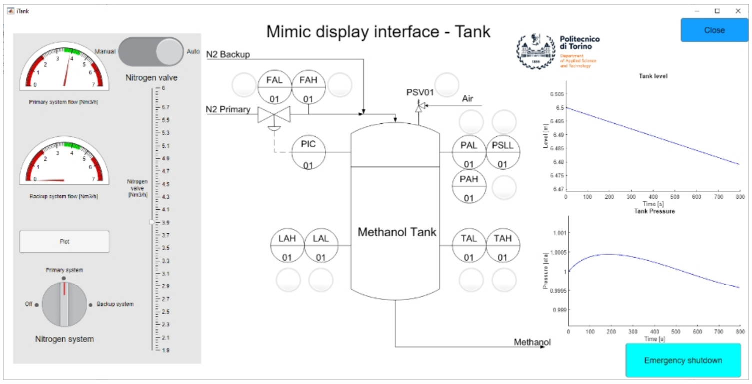

Fig. 4. Tank mimic (top left the close button; bottom down, the emergency button).

3.2.4. Procedure

The procedure failed to give, in some cases, an exact waiting time duration after a specific control action was suggested to be performed. In some cases where waiting time could not be precisely indicated, the participants complained that the word ‘monitor for a while’ was not helpful as they were tempted to react immediately, given time constraints. There were cases where participants were referred to a second procedure while on one; this required extra searching and scanning activities for new tasks, and although the participants could keep track, it was considered distracting.

3.2.5. Communication

An additional observation, particularly pertinent to Scenario 3, revolved around the participants' communication with the supervisor. The scenario was designed with the expectation that operators would call the supervisor for instructions and strictly adhere to them. However, the actual responses of the operators diverged from this expectation. In this scenario, the supervisor advised the operators to concentrate solely on the reactor alarm despite the occurrence of multiple other alarms. Interestingly, none of the participants focused exclusively on the reactor alarm.

The prevalence of numerous alarms seemed to overwhelm the operators, preventing them from adhering strictly to the supervisor's instructions or the context of the procedures. Instead, they found themselves compelled to address other alarms as well. This deviation from the expected course of action highlights a critical aspect of control room operations: the challenge of prioritising tasks in a high-alarm environment. It suggests that when faced with many alarms, operators may find it challenging to focus on a single aspect of the system, even when directed to do so by a supervisor.

This observation underscores the complexity of decision-making in control room environments, particularly when operators are inundated with multiple alarms. It also points to better alarm management systems to help operators prioritise alarms more effectively. Furthermore, this scenario illustrates the importance of flexibility in operational protocols and the need for systems that can adapt to the dynamic nature of real-world control room environments.

In summary, this observation from Scenario 3 provides valuable insights into how operators manage competing demands and highlights the need for more intuitive and adaptive support systems in control rooms. It also emphasises the importance of realistic training scenarios that prepare operators for the complexities of managing multiple simultaneous alarms.

4. General observations

A notable observation emerged in our study regarding using the AI-based DSS in Group 4. It was evident that participants' engagement with the decision support varied significantly, with trust in the system playing a central role in its utilisation. Many participants opted to rely on the digitised procedures instead of the AI-provided suggestions, indicating a preference for familiar methods over the new system. Interestingly, some participants chose to use AI support only in specific circumstances, mainly when they realised that time constraints made it impractical to follow traditional procedures thoroughly. This selective usage suggests that the optimal application of the AI support system might be employed as a supplementary tool when conventional methods are insufficient or too time-consuming.

However, a key challenge identified was the redundancy of information presented by the AI support system compared to the digitised traditional procedures. When participants attempted to follow both simultaneously, it occasionally led to confusion and perceived conflicts between the two sources of guidance. This misunderstanding impacted their performance negatively and diminished their trust in AI support. When such conflicts arose, participants were observed to discontinue using the AI support, perceiving it as unhelpful and detrimental to their operational efficiency.

This observation highlights the benefit of integrating DSS into control room operations. However, this must be done with careful attention to the design to avoid information overload and potential conflicts with existing procedures. Ensuring clarity and compatibility between traditional methods and new decision support tools is essential to enhance operator performance and establish trust in these advanced systems.

5. Data collection methods

5.1. Situational awareness and workload assessment

The situation presence assessment method (SPAM) was used for the data collection on situational awareness alongside the situational awareness rating technique (SART) (Amazu et al., 2023). The SART questions were streamlined into a single online questionnaire incorporating the NASA-Task Load Index questionnaire. For the SPAM (following a concurrent think-aloud approach), the participants were asked context-specific questions at three different points in time during each scenario. The questions were aimed at inquiring about the participant's understanding of the situation at the perception of the onset of an alarm, then the interpretation and planning of what they needed to do, and lastly, to provide a brief overview of the expected outcome of the actual actions they decided to carry out. Specifically, the questions asked were:

1) Perception (Question 1):
   Which of these alarms, in your opinion, must be verified first? and why?
   (AI system: What is the AI decision support system about?)

2) Understanding (Question 2):
   Why do you think the PAL01 alarm is activated? and what do you intend to do?
   (AI system: What was the suggestion on? Was it clear what you were expected to do and why?)

3) Projection (Question 3):
   Now that you have done this, what do you think will change in the system? Why?

During the estimated test time, many participants were asked the questions around the 6th, 8th and 12th minute. However, those in group 4, got earlier suggestions from the AI decision support system; hence, the timing for the first question was as early as the 2nd minute.

The participants perceived completing the SART and NASA-TLX questions after each scenario as a relaxing downtime. However, the SPAM questions asked during the test were considered intrusive by some participants, particularly while they were constrained with time or facing many alarms. The observers, though, were all instructed to pay attention to the participants' state and the expected timing before asking. This is understandable, given that the simulator is not paused for these questions and aligns with findings from the literature on the Intrusiveness of SPAM (Endsley, 2021). In our experience, however, SPAM provides better insight than SART or SAGAT regarding the actual situational awareness of participants in action at critical points during the running of the scenarios, even if it is sometimes considered intrusive by the participants.

5.2. Biometrics

Eye-tracking (using Tobii Glasses 3) and Heartbeat monitoring (using Empatica Care Lab) methods were used to collect objective data on the state of operators. These are standard techniques used in literature to understand operators' cognitive and physiological states, including but not limited to workload, situational awareness, and stress (Hinss, 2022).

There were generally no issues for the participants regarding using the watch. However, the pre-and post-setup processes had to be done correctly to avoid possible data losses. Likewise, the eye-tracker was not reported as disturbing for the participants. Similar attention had to be paid to the calibration and the validation of the gaze before the start of the test. Their benefit in understanding the 'Why' underpinning human performances and possible human error still needs to be explored in future work.

6. Discussion

6.1. Preliminary lessons learnt from the experiment

Considering the observations from our experimental study, we have developed a set of guidelines for designing effective and efficient human-machine interaction given different combinations of decision support systems in control room operations. These guidelines address the challenges and opportunities identified in our research regarding the process industry. The aim is to optimise the benefits of these systems while mitigating potential drawbacks, ensuring that operators can effectively and safely manage complex industrial processes.

6.1.1. Lessons learnt regarding displays and procedures (general)

The intricacies in design consideration for the display and procedures go hand in hand. This is because the wording used to describe specific systems, functions, interfaces, etc., as described in the procedure, must match what is used on the display. Some of these suggestions based on our study can already be picked up from ISO 11064 – Ergonomic Design of Control Centres and the HPOG guideline for procedure design (ISO, 2006; HPOG, 2021).

- Use the exact wording. We had to emphasise the similarities in the words 'mimic' and 'section' used in the display and procedures. Therefore, the terminology most widely used during communication must also be deployed in procedural writing or interface design.
- The numbering of process values should be the same in decimal places on both display and procedure.
- Provide time-based indication when providing instruction related to time-related task steps (e.g. do not use vague sentences as "continue to monitor for a while" but rather be specific: "continue to monitor the situation for the next 2 to 5 minutes to ensure the system is still stable").
- In cases where procedures point or redirect the operator to another procedure, ensure that there is a digital hyperlink or that exact details such as the name of the procedure, plant section, and page number are provided to access the navigation in paper formats.
- Provide as much feedback to operators as possible using acceptable colours and less intrusive pop-ups following each HMI-mediated task.
- Cluster command button also considers the frequency of usage and functional allocation (e.g. a navigation command should not be placed close to an emergency shutdown).
- When defining nominal values, include threshold and limit values +/- to support operators' understanding of risk.
- Consider revising alarm rationalisation and arrangement for mimics and overview display after a few months of operational usage to verify the actual frequency of correlation and clustering for alarms and frequency of usage for the mimics.

6.1.2. Lessons learnt regarding paper-procedure (groups 1 & 2 configurations)

- Include an organised table of contents with a consistent order, whether numerically or alphabetically.
- Consider using colour stickers to differentiate plant sections, if possible, for the procedures on paper.

6.1.3. Lessons learnt on digitised procedures (group 3 configuration)

The following guidelines are given for the configuration with procedures on screen.

- The way the procedures are organised influences the keenness of operators to access and follow them. Organising the procedures by plant section on the display and subsequently numerically for each tag was very helpful for navigation. However, in some cases where the numbers varied, participants started to find it confusing but adjusted after a while. Therefore, ensure a consistent order, whether numerically or alphabetically.
- Automatic opening of the appropriate procedure upon double-clicking on the alarm in the alarm list. This feature would significantly reduce the time spent searching for the correct procedure and optimise the use of the on-screen procedure.

Optionally, further research can be done to compare the procedure representation format used in this study with flow-chart-based procedures.

6.1.4. Lessons learnt for the AI-based recommendation (configuration)

Guidelines for the effective use of AI-based DSS in control room operations:

- Implement a system where operators are not immediately shown the suggestions from the AI DSS
- Provide operators with the option to choose whether to view and follow the suggestions actively. This choice should be made based on their assessment of the situation. In this way, the operator should not depend entirely on the AI-DSS as advised in (Hidekazu, 2005).
- Encourage operators to rely on their expertise and judgment first. The AI DSS suggestions should be utilised only when operators feel they cannot effectively resolve the situation independently. This approach ensures that the system is used as a supplementary tool, enhancing decision-making in more complex or uncertain scenarios.
- Ensure that operators are thoroughly familiar with the decision support system. They should be well-versed in its functionalities, potential benefits, and possible drawbacks, enabling them to make informed decisions about when and how to use the system effectively. This approach also aligns with the recommendations presented in (Hidekazu, 2005).
- Enhance the decision support system with real-time feedback mechanisms. When following the system's suggestions, participants often faced uncertainty about the effectiveness of their actions. Incorporating feedback that confirms or guides their actions can significantly improve the system's usability and effectiveness. This approach also aligns with the recommendations presented in (Power, 2002).

6.1.5. Lessons learnt regarding induction training (All)

- Training on the use of procedures and displays should go together. We observed that this aided the familiarity with both supports. In this sense, a walk-through training routine regarding some basic and critical procedures could be a useful way to provide information regarding how and why a certain task should be executed and let the operator familiarise with the way to implement it.
- Ensure operators are retrained after each update, significantly when the terminologies are changed and/or implemented in new designs. These can confuse operators with a different mental model and induce errors based on misunderstanding and confusion.

6.2. Lessons learnt on the data collection methods

- Questionnaire and situational awareness observation protocols such as SPAM and NASA TLX can only be used during plant drills or training. They should also be used after new decision support tool updates through experimental studies. They provided insight into subjects' workloads and situational awareness.
- The eye-tracking tools are also accepted only in testing and simulator conditions, and they can be used to collect metrics associated with mental workload (such as pupil diameter). They can be used to shed light concerning the focus of attention (gaze direction and related heat maps). With sufficient data, this can augment the decision-making for optimisation of parts of the interfaces and information provision. However, the current data set has only been used for preliminary exploratory analysis. The potential of the data collected during the test will be explored in future work.
- Heart rate and electrodermal activities can easily be collected, as in our case, using hand-worn, very comfortable watches. Other techniques or data sources, such as behavioural markers from operational logs, can be explored for real-time applications to support data collection for situational awareness or workload prediction. However, to gain useful insight, a derived measure of heart rate variability needs to be obtained, and this type of indicator requires significant data processing (Faust et al., 2022).

7. Conclusion

The experimental studies have been instrumental in elucidating the intricacies and potential of various human-in-the-loop (HITL) configurations in control room operations, encompassing both existing and prospective setups. The primary insight from these studies underscores the undeniable significance of human-centric design in these systems, irrespective of technological advancement. Participants exhibited tendencies for errors and procedural deviations in each configuration, including those incorporating AI-based suggestions for decision support. This finding accentuates a critical element of future system design: the lessons learned here are invaluable for both the design and training phases, especially in integrating new AI-based systems with established operational frameworks. A human-centric approach is imperative, not merely as a design principle but as a fundamental requirement for optimal decision support, operator performance and safety. As we progress into environments increasingly characterised by automation and AI, these results underline the necessity of a balance where technology complements, rather than supplants, human judgment and expertise. The evolution of industrial operations depends on developing systems that are not only technologically sophisticated but also intuitively resonate with the operators' needs and skills. Consequently, this study provides a strategic guide for future developments in the domain, promoting a synergistic fusion of human abilities and artificial intelligence in safety-critical environments.

Acknowledgements

This work has been done within the Collaborative Intelligence for Safety-Critical Systems project (CISC). The CISC project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement no. 955901. The authors thank Rob Turner and Adrian Kelly in Yokogawa and EPRI Europe, respectively, for contributing to the development of the human system interfaces. Finally, thanks to the participants for the time they dedicated to this study and their feedback.

References

Amazu, C. W., Briwa, H., Demichela, M., Fissore, D., Baldissone, G. & Leva, M. C. 2023. "Analysing 'Human-in-the-Loop' for Advances in Process Safety: A Design of Experiment in a Simulated Process Control Room." 33rd European Safety and Reliability Conference (ESREL 2023), no. November, pp. 2780–2787.

Endsley, M. R. 2021. A Systematic Review and Meta-Analysis of Direct Objective Measures of Situation Awareness: A Comparison of SAGAT and SPAM. Human Factors. 2021. Feb;63(1):124-150. doi: 10.1177/0018720819875376. Epub 2019 Sep 27. PMID: 31560575.

Faust, O., Hong, W., Loh, H. W., Xu, S., Tan, R.-S., Chakraborty, S., Barua, P.D. & Molinari, F. 2022. Heart rate variability for medical decision support systems: A review. Computational Biology and Medicine, 145 (2022), Article 105407.

Hidekazu, Y. 2005. Human-machine interaction in nuclear power plants. Nuclear Engineering and Technology, 37(2), pp. 151-158.

HPOG Human Performance Oil Gas. 2021. Best Practice in Procedure Formatting. Retrieved January 30, 2023, from https://www.hpog.org/resource-centre/best practice-in-procedure-formatting.

ISO 11064-7: 2006(en). Ergonomic design of control centres — Part 7: Principles for the evaluation of control centres.

Mietkiewicz, J., Madsen, A.L. 2023. Enhancing Control Room Operator Decision Making: An Application of Dynamic Influence Diagrams in Formaldehyde Manufacturing. In European Conference on Symbolic and Quantitative Approaches with Uncertainty. Cham: Springer Nature Switzerland, pp. 15-26. DOI: http://dx.doi.org/10.3850/978-981-18-8071-1_P531-cd.

Mietkiewicz, J., Abbas, A.N., Amazu, C.W., Baldissone, G., Madsen, A.L., Demichela, M. & Leva, M.C. 2023b. Enhancing Control Room Operator Decision Making. Preprints, 2023121983. https://doi.org/10.20944/preprints202312.1983.v1.

Mietkiewicz, J., Abbas, A.N., Amazu, C.W., Madsen, A.L. & Baldissone, G. 2023a. Dynamic Influence Diagram-Based Deep Reinforcement Learning Framework and Application for Decision Support for Operators in Control Rooms.

Power, D.J. 2002. Decision support systems: Concepts and resources for managers. Quorum Books.

Reason, J., Manstead, A., Stradling, S., Baxter, J. & Campbell, K. 2011. Errors and Violations on the Roads: a Real Distinction? Ergonomics. 33. 1315-32. 10.1080/00140139008925335.

Shi, C., Noah, B., Rothrock, L. & Kim, J. 2021. Potential benefits of eye tracking within process control monitoring tasks. Human Factors and Ergonomics in Manufacturing. 2021; 31:316–326. https://doi.org/10.1002/hfm.208922.

Xu, S., Song, F., Li, Z., Zhao, Q., Luo, W., He, X. & Salvendy, G. 2008. An ergonomics study of computerised emergency operating procedures: Presentation style, task complexity, and training level. Reliability Engineering and System Safety, vol. 93, no. 10, pp. 1500–1511.