Proceedings of the 35th European Safety and Reliability & the 33rd Society for Risk Analysis Europe Conference
Edited by Eirik Bjorheim Abrahamsen, Terje Aven, Frederic Bouder, Roger Flage, Marja Ylönen

doi: 10.3850/978-981-94-3281-3_ESREL-SRA-E2025-P9407-cd

Physiological Impact Assessment of Decision Support Systems on Control Room Operators: An ANCOVA Analysis

Joseph Mietkiewicz

School of Food Science and Environmental Health, Technological University Dublin, Ireland.
E-mail: joseph.mietkiewicz@tudublin.ie

Ammar N. Abbas

Computer Science, Technological University Dublin, Ireland E-mail: ammar.abbas@tudublin.ie

Chidera W. Amazu

Politecnico di Torino, Italy, E-mail: chidera.amazu@polito.it

Anders L. Madsen

Department of Computer Science, Aalborg University, Denmark. E-mail: anders@hugin.com

Maria Chiara Leva

School of Food Science and Environmental Health, Technological University Dublin, Ireland.
E-mail: mariachiara.leva@TUDublin.ie

This study investigates the physiological effects of AI based Decision Support Systems (DSS) on control room operators through a comprehensive analysis of multiple physiological indicators. Using data from 41 participants divided into control and experimental groups, we analyzed heart rate, temperature, electrodermal activity (EDA), and pupil diameter across three scenarios of increasing complexity. Analysis of Covariance (ANCOVA) was employed to control for baseline differences, revealing significant reductions in pupil diameter ($p = 0.0029$) for the DSS group, indicating lower cognitive load. While other physiological measures showed consistent trends suggesting reduced stress with DSS use, these differences were not statistically significant. The findings provide empirical evidence for DSS's positive impact on operator cognitive load, particularly during complex scenarios, while highlighting the need for comprehensive physiological monitoring in assessing human-system interaction.

Keywords: Decision Support Systems, Control Room Operations, Physiological Measurements, ANCOVA, Cognitive Load, Eye Tracking, Heart Rate Variability, Safety-Critical Systems

1. Introduction

The implementation of Decision Support Systems (DSS) in control room operations represents a critical advancement in industrial safety and efficiency. While the operational benefits of DSS have been well-documented, a comprehensive understanding of their physiological impact on operators is essential. Such an understanding is crucial for optimizing human-system interaction, ensuring that technological support enhances rather than hinders operator performance.

Previous research has primarily focused on performance metrics and subjective assessments of DSS effectiveness. However, physiological measurements provide objective insights into operators' cognitive and emotional states during task performance. Studies have employed individual physiological measurements to assess operators' responses to DSS, such as eye tracking to identify periods of high cognitive load and potential operator errors (Ikuma et al., 2014), and heart rate monitoring to observe the effects of new DSS implementations on operator stress (Fallahi et al., 2016).

This study extends existing research by combining multiple physiological measures to provide a more complete picture of the impact of DSS on control room operators. By integrating these measures, we aim to offer a more nuanced understanding of how DSS influences operators' cognitive workload, stress, and overall performance.

This paper presents experimental results focusing on the physiological data collected during DSS use or not in control room environments. The development of the DSS is detailed in Mietkiewicz et al. (2024), while the full experimental design is described in Amazu et al. (2024) and data in Amazu et al. (2024).

2. Methods

2.1. *Experimental Design and Objectives*

This study employed a controlled experimental design to evaluate the physiological impact of DSS on control room operators. The primary objective was to quantify changes in operator cognitive load and stress levels through comprehensive physiological monitoring while performing complex control room tasks. The experimental design incorporated three scenarios of increasing complexity, allowing for systematic assessment of DSS effectiveness across varying operational demands.

2.2. *Simulation Environment*

The experimental platform utilized a high-fidelity simulation of a formaldehyde production plant Demichela et al. (2017) use to create an operational control room environment, developed in collaboration with industry experts to ensure operational realism. The main interface of the simulator can be seen in fig. 1.

The simulation encompassed a comprehensive process flow including the Tank Section with methanol storage and alarm systems, which fed into the Methanol Section for liquid-to-gas conversion. The Compressor Section mixed compressed gas with vaporized methanol, heating the mixture to 200°C in a heat exchanger (REC2). The central Reactor Section contained the primary reaction vessel, surrounded by a Heat Exchanger network (REC1, REC2, REC3) managing thermal conditions. REC1 and REC3 provided product cooling through heat recovery systems, while the Absorber Section employed a countercurrent flow design for precise product absorption.

This comprehensive simulation environment provided a realistic platform for evaluating operator performance under varying conditions while maintaining experimental control and repeatability.

2.3. *Participant Characteristics and Group Assignment*

The study involved 41 participants (mean age = 25 years, SD = 5.4), primarily comprising chemical engineering master's students with foundational knowledge of process control principles. The Control Group (G1, n=20) operated the simulator without DSS support and the experimental Group (G2, n=21) utilized the integrated DSS throughout operations. Prior to experimental trials, all participants underwent standardized training to ensure baseline competency with the simulator interface and basic operational procedures. The study protocol was approved by both the Technological University Ethics Review Committee (Ref: REC-20-52A), with informed consent obtained from all participants.

2.4. *Experimental Scenarios*

Three distinct scenarios were developed to evaluate operator performance across a spectrum of operational complexity:

(1) **Pressure Indicator Control Failure:** Automatic pressure management system malfunction requires manual nitrogen inflow adjustment, causing pressure drop due to continuous pump operation.
(2) **Nitrogen Valve Primary Source Failure:** Primary nitrogen source failure necessitates backup system activation, demanding also operator intervention to control pump power and mitigate pressure decline.
(3) **Temperature Indicator Control Failure:** Control room operator faces challenges managing reactor temperature after cooling system malfunction, requiring coordination with field personnel to prevent overheating and ensure safe operational parameters.

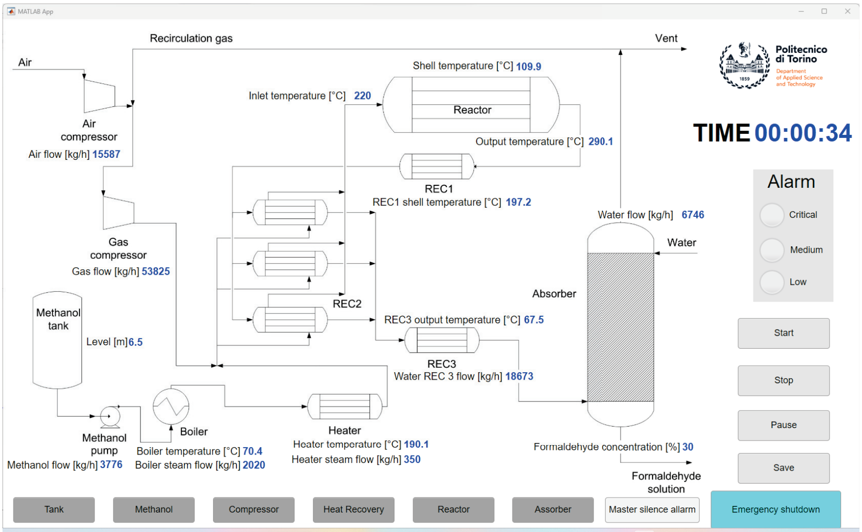

Fig. 1.: Process Flow Diagram of Production: Formaldehyde is synthesized by mixing methanol with compressed air, and then heating the mixture. This initiates a chemical reaction in the Reactor, followed by dilution in the Absorber to achieve the desired concentration. At the bottom, there are various mimics that the operator can display on another screen for a detailed process flow diagram of specific plant sections. First published in Lecture Notes in Computer Science vol 14294, pp 15–26 by Springer Nature

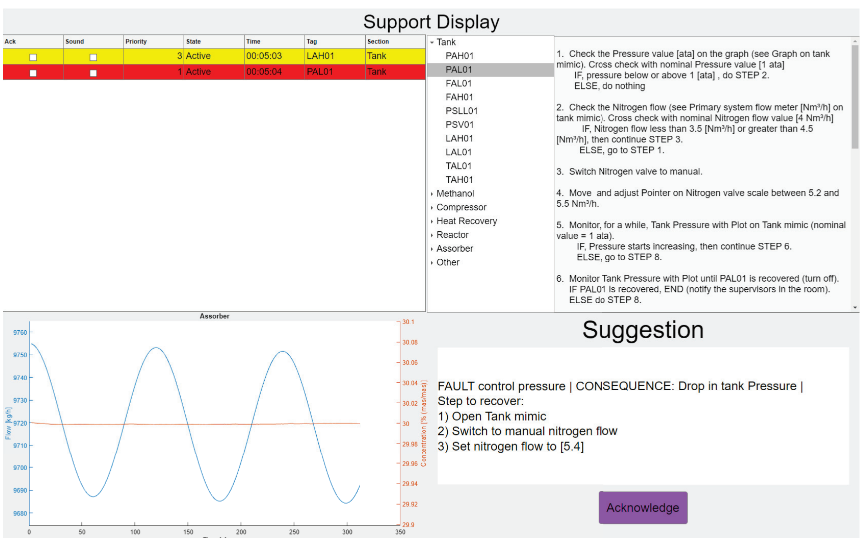

Fig. 2.: Support Panel Layout: In the upper left quadrant, the alarm list alerts the operator to current issues. The upper right quadrant houses the traditional procedure system. The lower left quadrant features a critical graph displaying production metrics. In the lower right quadrant, the DDS quadrant provides situation-specific recommendations to the operator, this quadrant is present or not depending of the experimental group. First published in Lecture Notes in Computer Science vol 14294, pp 15–26 by Springer Nature

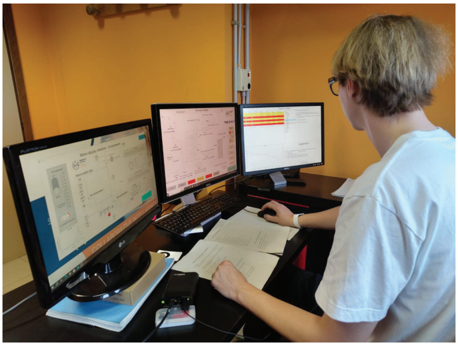

Fig. 3.: Participant with the AI configuration (G2) wearing the eye tracker and the smart watch.

2.5. *Data Collection and Instrumentation*

Using the EmbracePlus smartwatch Empatica (2024), physiological measures collected included:

- **Pulse Rate:** Measured as the number of pulses per minute, this metric serves as an indicator of arousal, stress levels, and emotional responses. Variations in pulse rate can provide insights into the operator's physiological reactions to different scenarios and decision-making processes Fallahi et al. (2016).
- **Temperature:** Body temperature measurements can indicate stress responses, changes in mental workload, or shifts in emotional states. This metric adds another dimension to our understanding of the operator's physiological state during task performance.
- **Electrodermal Activity (EDA):** This measure of the skin's electrical conductance is a sensitive indicator of emotional arousal and stress. Changes in EDA can reveal subtle emotional responses to different aspects of the control room tasks, providing a window into the operator's affective state Boucsein (1993).

The study used Tobii Pro Glasses 3 Tobii Pro Glasses 3 (2024) for eye tracking. This study focuses on the Pupile Diameters metric. Pupil diameter variations reflect changes in cognitive load and emotional arousal Ikuma et al. (2014). Pupil diameter was assessed at two times: during the baseline overview (pre-alarm) from the start of the experiment to the first critical alarm, and during the critical alarm, from its onset to its conclusion. The integration of these physiological metrics allows us to provide objective measurements to assess the operator workload and test the impact of the DSS on control room operators.

2.6. *Statistical Analysis Framework*

A robust statistical approach was implemented using Analysis of Covariance (ANCOVA) Huitema (2011) to evaluate physiological responses while controlling for individual baseline variations. The primary model was specified as:

$$\begin{aligned} Y = \beta_0 &+ \beta_1 Group + \beta_2 Scenario \\ &+ \beta_3 (GroupScenario) \\ &+ \beta_4 Baseline + \epsilon \end{aligned} \tag{1}$$

Where Y is the physiological measure of interest, β_0 is the intercept term, β_1 is the group effect coefficient, β_2 is the scenario effect coefficient, β_3 is the interaction effect coefficient, β_4 is the baseline adjustment coefficient, and ϵ is the error term.

Model assumptions were verified through Levene's test Levene (1960) for homogeneity of variance and visual inspection of residual plots. The Wilcoxon Rank-Sum tests Mann and Whitney (1947) were employed when assumption violations occurred, ensuring robust statistical inference across all analyses.

3. Results

3.1. *Pulse Rate*

The ANCOVA analysis for pulse rate, considering the three different scenarios, is summarized in Table 1. This analysis aims to understand the impact of the DSS on the pulse rates of control room operators. The ANCOVA results show no statistically significant difference between the groups overall ($p = 0.148$). However, the scenario factor is significant ($p = 0.021$), indicating that pulse rates differ significantly across the scenarios. The interaction between group and scenario is not significant ($p = 0.837$), suggesting that the effect of the DSS is consistent across scenarios.

Table 1.: ANCOVA Table for Pulse Rate

Source	SS	df	F	p
Group	56.63	1.0	2.13	0.148
Scenario	213.66	2.0	4.03	0.021
G×S	9.44	2.0	0.18	0.837
Baseline	17304.30	1.0	652.28	≥ 0.001
Residual	2281.49	86.0		

SS: Sum of Squares, df: Degrees of Freedom, F: F-statistic , p: p-value

The statistical check of the use of the ANCOVA mode can be seen in fig. 4. The linearity check plot shows a strong positive linear relationship between the observed and predicted pulse rate values, confirming that the linearity assumption is met. The independence of residuals plot displays residuals scattered randomly around zero, without any discernible pattern, indicating that the assumption of independence is satisfied. Lastly, the normality plot (Q-Q plot) shows the residuals closely following the diagonal line, suggesting that the normality assumption is also met. These plots collectively demonstrate that the data is suitable for ANCOVA analysis.

The adjusted means for pulse rate show that G2 consistently has lower pulse rates compared to G1 across all scenarios:

- **Scenario 1 (S1):** G1 = 82.67, G2 = 74.17
- **Scenario 2 (S2):** G1 = 79.65, G2 = 72.22
- **Scenario 3 (S3):** G1 = 80.52, G2 = 71.51

The results of the Levene's test are as follows:

- **S1:** Stat = 5.115, p = 0.031
- **S2:** Stat = 3.787, p = 0.061
- **S3:** Stat = 6.799, p = 0.014

Levene's test for equal variances reveals that the assumption of homogeneity of variances is violated for Scenarios 1 and 3 ($p \leq 0.05$), but holds for Scenario 2 (p = 0.061). To account for heteroscedasticity, the ANCOVA was conducted with robust standard errors (HC3). The results suggest that the DSS may have a positive impact on reducing control room operators' pulse rates, as evidenced by the consistently lower adjusted means for G2 compared to G1 across all scenarios. However, the group effect was not statistically significant, possibly due to the limited sample size or variability within groups. The significant scenario effect indicates that the complexity of the scenario influences pulse rates, with higher complexity leading to increased pulse rates. The lack of a significant interaction between group and scenario suggests that the DSS's effect on pulse rate is not dependent on the scenario complexity.

3.2. *Temperature*

The ANCOVA results for temperature across the three scenarios are summarized in Table 2. The statistical check of the ANCOVA for the Temperature can be seen in fig. 5. The model adjusts for baseline differences and examines the main effects of group and scenario, as well as their interaction effect on temperature. The table below presents the sum of squares, degrees of freedom, F-value, and p-value for each effect in the model.

Table 2.: ANCOVA Table for Temperature

Source	SS	df	F	p
Group	0.049	1.0	0.229	0.633
Scenario	4.179	2.0	9.811	≥ 0.001
G×S	0.022	2.0	0.051	0.950
Baseline	172.81	1.0	811.46	≥ 0.001
Residual	18.314	86.0		

SS: Sum of Squares, df: Degrees of Freedom, F: F-statistic , p: p-value

The statistical check plots fig. 5 confirm that the ANCOVA assumptions of linearity, independence of residuals, and normality are satisfied, validating the use of this analysis for the temperature data.

The results indicate that the main effect of the group (C(Group)) on temperature was not statistically significant ($F(1, 86) = 0.229$, $p = 0.633$). This suggests that there was no significant difference in adjusted temperature between the groups that used the DSS and those that did not, when averaging across all scenarios. However, the main effect of the scenario (C(Scenario)) was significant ($F(2, 86) = 9.811$, $p \leq 0.001$). This indicates that the temperature varied significantly across the different scenarios. The interaction effect between group and scenario (C(Group):C(Scenario)) was not significant ($F(2, 86) = 0.051$, $p = 0.950$), suggesting that the effect of the group on temperature

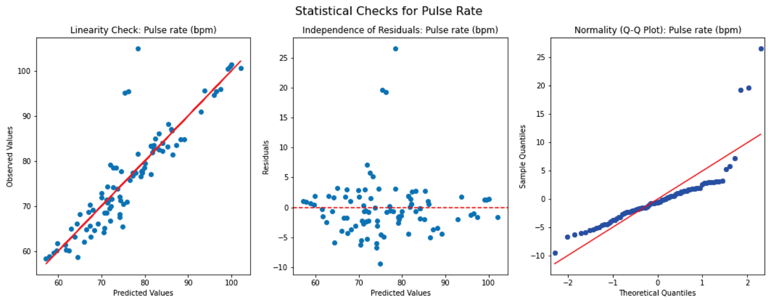

Fig. 4.: Statistical checks for Pulse Rate: Linearity, Independence of Residuals, and Normality.

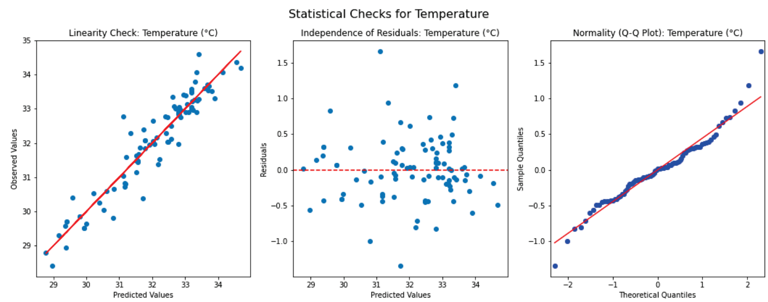

Fig. 5.: Statistical checks for Temperature: Linearity, Independence of Residuals, and Normality.

did not differ significantly across the scenarios. The adjusted means for temperature across the scenarios are presented below:

- **S1:** G1 = 31.28°C, G2 = 32.08°C
- **S2:** G1 = 31.68°C, G2 = 32.48°C
- **S3:** G1 = 31.89°C, G2 = 32.62°C

The adjusted means show that, in all scenarios, Group G2 (which used the DSS) had slightly higher temperatures than Group G1. This difference, although consistent, was not statistically significant as indicated by the non-significant main effect of the group. Levene's test for equal variances indicates that the assumption of homogeneity of variances is met for all scenarios ($p \geq 0.05$), supporting the validity of the ANCOVA results:

- **S1:** Stat = 0.042, p = 0.838
- **S2:** Stat = 0.044, p = 0.836
- **S3:** Stat = 0.027, p = 0.871

These p-values indicate that the assumption of equal variances was met for temperature across all scenarios, supporting the validity of the ANCOVA results. In summary, the ANCOVA analysis for temperature revealed significant differences across scenarios but not between the groups or their interaction. The consistent trend of higher temperatures in Group G2, although not statistically significant, suggests potential physiological differences associated with the use of the DSS, warranting further investigation in future studies.

3.3. *EDA*

The EDA doesn't pass the statistical check for ANCOVA. We simply take the mean of the EDA for participants during each scenario and divide it by the mean value of EDA during the breaks. In this way, we normalize the value and we can compare them without baseline effect. The Mann-Whitney U test is used due to the non-normality of the data as shown by the Shapiro test. There are no statistical differences between the two groups.

In scenarios 1 and 3 we can observe a lower EDA for group 2 at the opposite of scenario 2

Table 3.: EDA Statistical Test Results

Scenario	Test	p	SW1	SW2	Levene
S1	WRS	0.288	2.1e-6	0.099	0.34
S2	WRS	0.468	9.3e-6	0.650	0.13
S3	WRS	0.375	0.980	3.8e-5	0.46

WRS: Wilcoxon Rank-Sum test
SW1, SW2: Shapiro-Wilk test p-values

where group 2 has a higher EDA.

- **S1:** G1 = 1.4, G2 = 0.85
- **S2:** G1 = 1.4, G2 = 1.0
- **S3:** G1 = 1.1, G2 = 1.35

Although the differences in normalized EDA between the groups were not statistically significant, the observed patterns suggest that the DSS may have a subtle influence on operators' emotional arousal during the scenarios. The lower adjusted EDA for G2 in Scenarios 1 and 2 could indicate that the DSS helps reduce emotional arousal or stress in these situations. However, the higher adjusted EDA for G2 in Scenario 3 suggests that the DSS might not have the same effect in more complex scenarios. The lack of statistical significance could be attributed to several factors, such as the limited sample size, individual variability in EDA responses, or the need for more sensitive measures of emotional arousal.

3.4. *Pupil Diameter*

The pupil diameter was assessed at two different times during the experiment. The first time period, referred to as the Baseline Overview, occurred from the start of the experiment until the onset of the first critical alarm. The second time period, known as the Critical Alarm, began with the start of the first critical alarm and continued until the alarm had ended. The ANCOVA results for pupil diameter across the three scenarios are summarized in Table 4. The statistical check of the ANCOVA for the Pupil Diameter can be seen in fig. 5. The ANCOVA results (table 4) reveal a statistically significant main effect of Group (p = 0.0029) on pupil diameter, indicating a significant difference between G1 and G2, independent of the scenario and baseline measurements. The main effect of Scenario (p = 0.768) and the interaction between Group and Scenario (p = 0.849) were not statistically significant, suggesting that the differences in pupil diameter between the groups were consistent across scenarios.

Table 4.: ANCOVA Table for Pupil Diameter

Source	SS	df	F	p
Group	0.221	1.0	9.411	**0.003**
Scenario	0.012	2.0	0.264	0.768
G×S	0.008	2.0	0.164	0.849
Baseline	4.972	1.0	211.89	$\geq$**0.001**
Residual	1.971	84.0		

SS: Sum of Squares, df: Degrees of Freedom, F: F-statistic , p: p-value

The adjusted means for pupil diameter for each scenario and each group are shown below:

- **S1:** G1 = 3.626, G2 = 3.299
- **S2:** G1 = 3.498, G2 = 3.311
- **S3:** G1 = 3.618, G2 = 3.314

The significantly lower adjusted pupil diameter for G2 compared to G1 suggests that the DSS may help reduce cognitive workload and emotional arousal during critical alarm phases. This finding aligns with the expectation that the DSS would be most beneficial in supporting operators during complex, high-stress situations. The lack of significant effects for Scenario and the interaction between Group and Scenario indicates that the impact of the DSS on pupil diameter was consistent across the different levels of scenario complexity. This consistency suggests that the DSS may be effective in reducing cognitive workload and emotional arousal regardless of the specific scenario encountered by the operators.

4. Discussion

The significant reduction in pupil diameter among DSS users provides strong evidence for decreased cognitive load when using decision support tools. This finding aligns with previous research suggesting that well-designed technological support can reduce mental workload in complex operational environments. The consistent but non-significant trends in heart rate and EDA indicate potential stress reduction benefits that warrant further investigation. The lack of statistical signif-

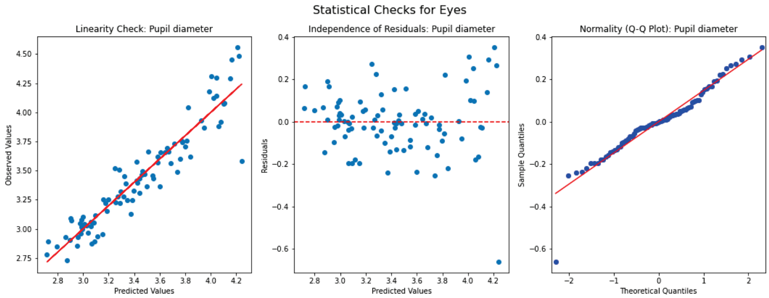

Fig. 6.: Statistical checks for Pupil Diameter: Linearity, Independence of Residuals, and Normality.

icance in these measures might be attributed to individual variability and the limited sample size.

5. Conclusion

This study provides empirical evidence that DSS implementation positively affects operator cognitive load, particularly as measured through pupil diameter. The findings support the implementation of DSS in control room environments while highlighting the need for comprehensive physiological monitoring in assessing human-system interaction. Future research should focus on larger sample sizes and longer-term physiological effects of DSS use.

Acknowledgement

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 Program under the Marie Skłodowska-Curie grant agreement no. 955901.

References

Amazu, C. W., J. Mietkiewicz, A. N. Abbas, H. Briwa, A. Alonso-Perez, G. Baldissone, D. Fissore, M. Demichela, and M. C. Leva (2024). Exploring the influence of human system interfaces: Introducing support tools and an experimental study. *International Journal of Human–Computer Interaction*, 1–18.

Amazu, C. W., J. Mietkiewicz, A. N. Abbas, H. Briwa, A. A. Perez, G. Baldissone, M. Demichela, D. Fissore, A. L. Madsen, and M. C. Leva (2024). Experiment data: Human-in-the-loop decision support in process control rooms. *Data in Brief 53*, 110170.

Boucsein, W. (1993). Methodological issues in electrodermal measurement. *Progress in electrodermal research*, 31–41.

Demichela, M., G. Baldissone, and G. Camuncoli (2017). Risk-based decision making for the management of change in process plants: benefits of integrating probabilistic and phenomenological analysis. *Industrial & Engineering Chemistry Research 56*(50), 14873–14887.

Empatica (2024). Embraceplus smartwatch. `https://www.empatica.com/embraceplus`. Accessed: 2025-01-29.

Fallahi, M., M. Motamedzade, R. Heidarimoghadam, A. R. Soltanian, and S. Miyake (2016). Effects of mental workload on physiological and subjective responses during traffic density monitoring: A field study. *Applied ergonomics 52*, 95–103.

Huitema, B. (2011). *The analysis of covariance and alternatives: Statistical methods for experiments, quasi-experiments, and single-case studies*. John Wiley & Sons.

Ikuma, L. H., C. Harvey, C. F. Taylor, and C. Handal (2014). A guide for assessing control room operator performance using speed and accuracy, perceived workload, situation awareness, and eye tracking. *Journal of loss prevention in the process industries 32*, 454–465.

Levene, H. (1960). Robust tests for equality of variances. In *Contributions to Probability and Statistics*, pp. 278–292.

Mann, H. and D. Whitney (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 50–60.

Mietkiewicz, J., A. N. Abbas, C. W. Amazu, G. Baldissone, A. L. Madsen, M. Demichela, and M. C. Leva (2024). Enhancing control room operator decision making. *Processes 12*(2), 328.

Tobii Pro Glasses 3 (2024). Tobii technology. `https://www.tobii.com/products/eye-trackers/wearables/tobii-pro-glasses-3`. Accessed: 2025-01.