Utilisation of multi-actor human-in-the-loop rail simulation to examine the impact of ATO-GoA2

Sarah A. Kusumastuti[1], Julia C. Lo[2], Tom H.J. Kolkman[1] & Simone Borsci[1]

[1]University of Twente, The Netherlands, [2]ProRail, The Netherlands

SUMMARY

This paper outlines the utilisation of a human-in-the-loop simulation of rail operations as part of Europe's Rail Joint Undertaking (ERJU) research on Automatic Train Operations-Traffic Management System (ATO-TMS). An experiment was designed to study performance and communication of train traffic controllers, network controllers, and train drivers under Automated Train Operations (ATO) Grade of Automation (GoA) level 2 systems compared to a system with no ATO implementation. Results indicate that workload appear to be higher for traffic control operators in GoA level 2. There are no differences in communications and collaboration among traffic control room operators nor between traffic control and train drivers compared to the current system. Operators' trust in automation was lower after experiencing the simulation. Qualitative feedback indicate that most traffic control operators perceive limited improvements offered by GoA 2, and so do most train drivers. Overall, the simulation managed to provide rail operators with an impression of how automated systems can impact their communication and way of working.

KEYWORDS

Human-in-the-loop simulation, railways, traffic controller, dispatcher, train driver, control room, communication, automatic train operations, grade of automation, multi-actor

Introduction

A human-in-the-loop (HITL) simulation environment offers the ability to test interactive technologies and designs in a safe and controlled environment for both exploratory and experimental research for human factors in rail workspace and operations. HITL simulations have been utilised in the railway domain to examine how train operators (including traffic controllers/dispatchers and/or network controllers) interact with new technologies (Balfe et al., 2018; Farrington-Darby, et al., 2006; Hillege et al., 2020; Lo, 2020; Sharples, et al., 2011).

The railway traffic control room operations in the Netherlands consist of traffic controllers (also known as dispatchers in other regions) responsible for among others route setting, conflict resolution, communication with drivers, etc. as well as network controllers, who work with traffic controllers to manage rail traffic and timetabling. The network controller and traffic controllers usually operate in the same railway traffic control room, and a network controller oversees several traffic controllers. Traffic controllers are responsible in maintaining communication with train drivers.

Traffic control room operations is a team task among train traffic control room operators as well as between control room operators and train drivers. Meanwhile, much of the HITL rail simulation research involving traffic control operators focuses on individual traffic controller/dispatcher performance (Kusumastuti et al., 2025).

As part of Europe's Rail Joint Undertaking (ERJU) FP1 MOTIONAL project, we developed and conducted tests on a HITL simulation for the Dutch railway system involving train traffic control room operators and train drivers. This multi-actor simulation is the first of its kind in the railway domain, as many previous HITL simulations in the railway domain focuses on individual performances of train drivers or traffic controllers.

The purpose of the simulation is to examine the performance and communication of train traffic controller or dispatcher/network controller/train driver under Automated Train Operations-Traffic Management System (ATO-TMS) Grade of Automation (GoA) level 2. In this setup, GoA 2 trains run automatically for driving tasks (starting, accelerating, cruising, braking, stopping). The main research question for this simulation study is: what is the impact of driving with ATO GoA2 on the way of working of operational actors and on the network system?

Methods

The simulation took place at the ProRail simulation facilities in Utrecht, Netherlands. An entire day is devoted for a complete experiment in the simulation, which consists of 4 simulation sessions/runs.

Simulation setup

The simulation consists of two separate but connected workspaces. The first is a room for traffic control room operations consisting of two train traffic control workstations and one network controller workstation. At another room there is a train driving simulation. Communications between rooms is available through a connected phone system, similar to what is available in current rail operations. Between the two simulation rooms is the control room for the simulation environment.

The simulation covers areas from Hoofddorp to Weesp, which includes the Schiphol Airport station. Figure 1 shows an approximation of the geographical area covered in the simulation in respect of the entire Dutch railway system.

Figure 1: Geographical area covered in the HITL simulation (Nederlandse Spoorwegen, 2024)

The full simulation study employs a 2 by 2 design which produces a total of 4 simulation runs for each day (Table 1). Each simulation is assigned an experimental condition and a scenario. There are two experimental conditions in the simulation:

1. Current situation: this condition simulated the Dutch railways without ATO trains with the driver operating a train with current specifications
2. ATO situation: Railway system with 100% ATO GoA 2 trains with the driving operating a train with ATO GoA 2 capabilities

There are two scenarios that happens in each situation

1. Delay/disturbance scenario: multiple trains (+6, +7 and +10 min) are introduced, which lead to secondary delays.
2. Disruption scenario: The simulation environment runs an undisrupted train operation experience for 10 minutes, followed by two disruptions on two areas, which are a police intervention on station Schiphol airport (train cannot depart) and a switch malfunction in Diemen Zuid. Traffic controllers can reroute trains and the network controller may need to cancel trains. No disruption mitigation procedure is implemented.

The simulation was conducted in the same 1-2-3-4 order during each experiment.

Table 1: Overview of experimental sessions

	Experimental scenario	
Condition	Delay	Disturbance
Current situation	Session 1	Session 2
ATO situation	Session 3	Session 4

Data collection

There are several parallel data collections during the simulation sessions. System logs were collected of each operators' actions on the simulation. The sessions are recorded in camera that points at the interfaces of traffic controller and train driver. Additionally, logs of the phone communication were kept in the system, which includes timestamp, duration, and direction of call during the simulation. In each of the four sessions, paper questionnaire with a list of questions was handed to be completed by the participants. The question list varies by session. Additionally, at the end of all sessions, a debriefing and discussion session was conducted. The data collected during the questionnaires include:

Task performance

The actions of participants in the simulation such as choices and timestamps are logged in the simulation program

Workload

Workload measures are collected after each simulation session. The instrument used is a modified DLR-WAT 7-point scale (Brandenburger et al., 2023) indicating workload intensity, translated to Dutch). An English version of the response item is shown in Table 2. This scale was included in the questionnaire sheet of all four sessions.

Table 2: Item for workload

1	2	3	4	5	6	7
Extreme underload	Very low workload (underload)	Low workload	Optimal workload	High workload	Very high workload (overload)	Extreme overload

Trust in automation

Five items from the Trust in Automation scale (Körber, 2018) relevant to the context and translated to Dutch was presented in the beginning at the end, the items were statements evaluated on a 7-point likert scale. An English version is shown in Table 3. This scale was included in the questionnaire sheet of the first and last session.

Table 3: Items for trust in automation

	Strongly disagree				Strongly agree	
1. One should be careful with unfamiliar automated systems	1	2	3	4	5	don't know
2. I rather trust a system than I mistrust it	1	2	3	4	5	don't know
3. Automated systems generally work well	1	2	3	4	5	don't know
4. I trust [automated] systems [to assist me in doing my tasks].	1	2	3	4	5	don't know
5. I can rely on [automated] systems [to assist me in doing my tasks]	1	2	3	4	5	don't know

Simulation validity

A HITL simulation validity scale from Lo, et al. (2016) was included in the question sheet at the second session. An English version of the scale is shown in table 4. The scale consists of 9 items on a 5-point Likert-scale from strongly disagree to strongly agree.

Table 4: Items for simulation validity scale

	Strongly agree				Strongly disagree	
1. The representation of the timetables is sufficient for the task I perform in the simulator	1	2	3	4	5	don't know
2. The simulation environment felt more or less like my own work environment	1	2	3	4	5	don't know
3. The infrastructure model is sufficiently realistic for the task in the simulator	1	2	3	4	5	don't know
4. (Inverted) I do not have all the necessary information needed in the simulator to perform my task	1	2	3	4	5	don't know
5. The behaviour of the simulator work is sufficiently comparable to my operational systems	1	2	3	4	5	don't know
6. The simulator contains the necessary functionalities to perform the task	1	2	3	4	5	don't know
7. The chosen scenario in the simulator is similar to a situation that appears in real life	1	2	3	4	5	don't know

8. I can apply the information from the information sources in the simulator in a similar way as in the real world	1	2	3	4	5	don't know
9. The processes (interactions, communication) in the simulator are the same as those found in a similar situation at my workplace	1	2	3	4	5	don't know

Communication and collaboration

At least one observer is required to be present in each simulation room. For communication between driver and traffic control room telephone calls were logged, which includes timestamp, duration, and direction of call during the simulation. For communication between traffic and network controllers, the in-room observer notes the frequency of communication of communication between operators. In the plenary debriefing with all participants questions were included about the experienced communication and collaboration.

ATO simulation experience

After conducting all sessions, all participants were included in a debriefing session that includes open ended questions on their perception of the ATO systems that were simulated in the experiment.

Results

Participants

In total seven simulations took place with 26 participants consisting of seven train drivers, thirteen traffic controllers and six network controllers. One simulation day was cancelled due to extreme weather conditions in the Netherlands. Two participants could not participate due to illness.

On average the work experience of all participants in the rail sector was 13.72 years (SD = 10.03). When split out, the average work experience per role was train drivers (M = 19.37 SD = 6.38), traffic controllers (M = 8.5, SD = 6.59) and network controllers (M = 12, SD = 11.02).

Learning effects regarding the scenario could be found for all participants: all operators indicated that they experienced the scenarios in the 'ATO situation' to be easier because they already knew them from the 'current situation' condition

Task performance

Individual task performance for each operational role was initially intended to be measured. However, due to technical issues simulation runs sometimes needed to be paused or interrupted. As such, system log files were not analysed to measure task performance.

Workload

Workload was analysed as a repeated measures linear mixed model with participant ID as a random effect to account for missing data collection across different sessions. The distribution of workload values across all sessions is shown in a boxplot in Figure 2. Overall, there are no statistically significant differences found in workload between roles (p-value=0.59). Nonparametric Kruskal-wallis testing indicate that the workload overall in the ATO situation is higher than the current situation (p-value=0.02).

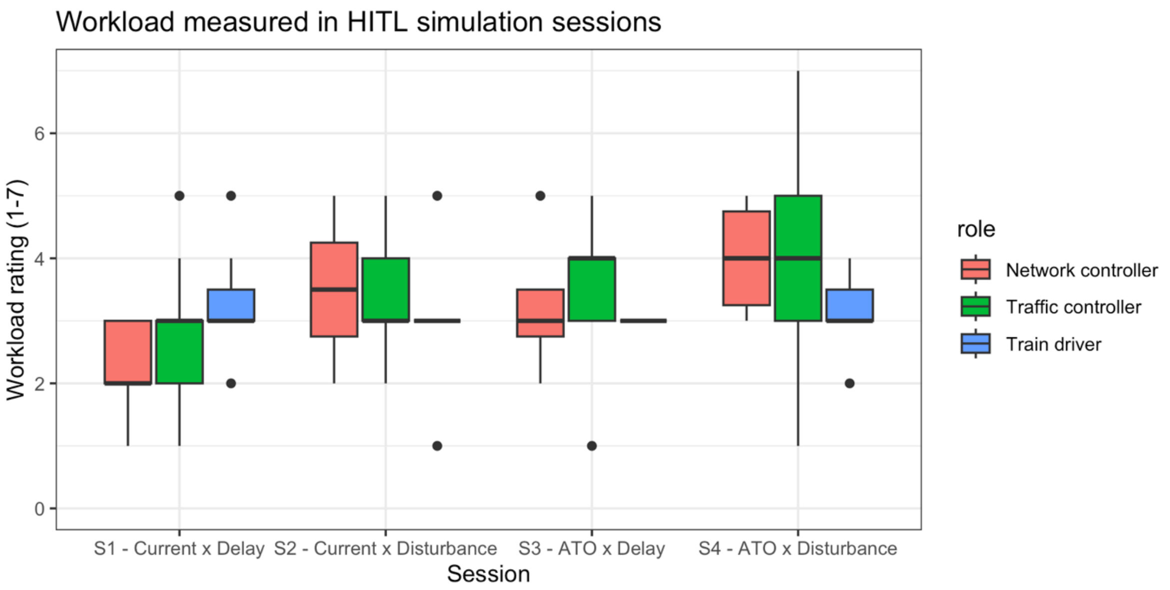

Figure 2: Workload of train operators in each simulation session for each operator type

Trust in Automation

Trust in automation was collected prior to the first session and after the last session. Figure 3 shows the distribution of trust in automation scores for each role in a boxplot. the maximum value for the scale is 25. Nonparametric testing with Wilcox test indicates that the score after the experiment is significantly lower than before the experiment (p-value = 0.045). There are no significant differences detected between the different roles (p-value = 0.79

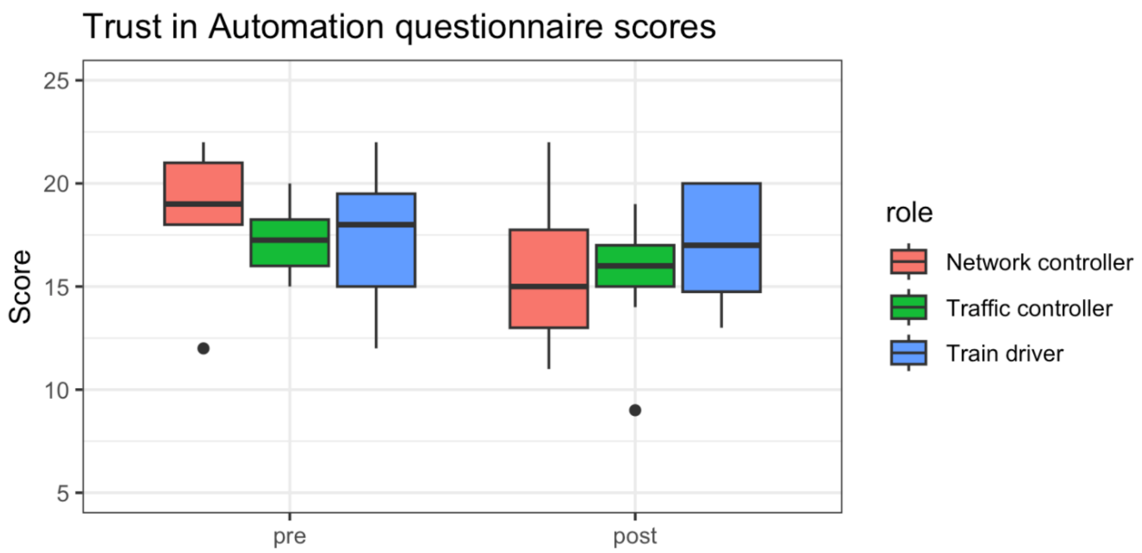

Figure 3: Pre- and post-experiment scores for Trust in Automation scale by each operator type

Simulation validity

The average simulation validity score across all participants is 31.7 with a standard deviation of 5.5 from a maximum score of 45. Figure 4 shows the distribution of simulation validity scores for each role in a boxplot. Pairwise nonparametric testing with the Wilcox test discovers that the score for traffic controllers (M = 34.6, SD = 4.7) are significantly higher than those of train drivers (M =

27.8, SD = 5.1)(p-value = 0.04) but not with network controller (M = 30.2, SD = 4.8)(p-value = 0.4). There is also no statistically significant difference between train driver and network controller (p-value = 1)

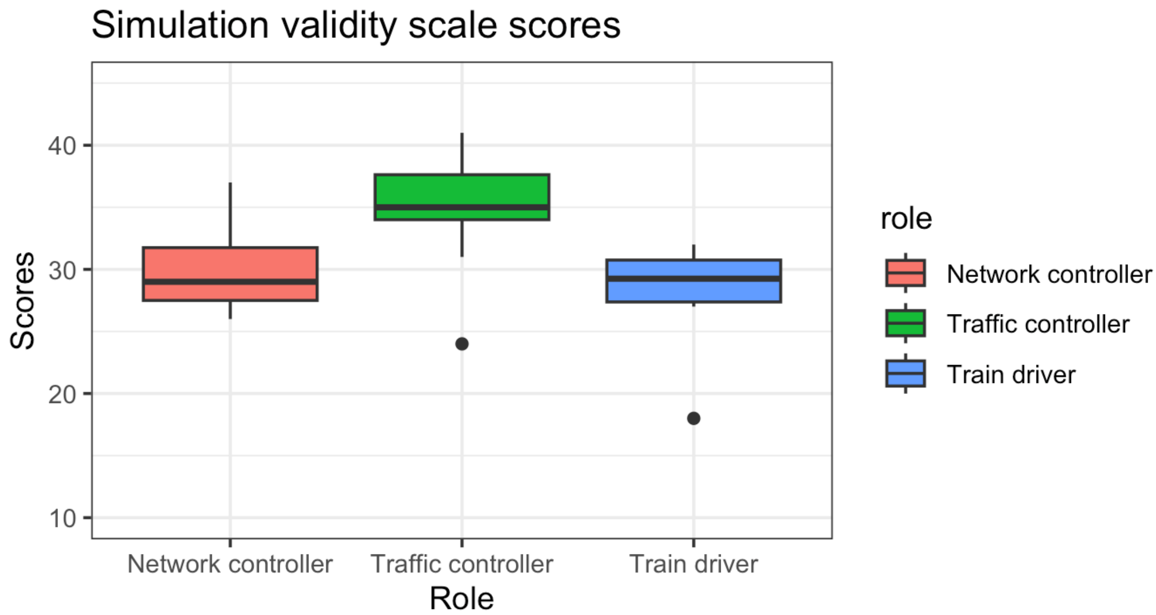

Figure 4: Simulation validity scale scores by each operator type

Communication and collaboration

Between train driver and traffic controller

Results from the observations and debriefing indicate that there is no change in communication and collaboration between train driver and traffic controller in the simulation.

Between traffic controller and network controller

Some changes in tasks and responsibilities impact their collaboration, but not extensively.

ATO simulation experience

Only a small number of train drivers perceive ATO as advantageous - mainly as a form of cruise control and for its ability to operate more consistently. Traffic controllers and network controllers experienced that the current ATO information integration in the TMS leads to an increased number of actions and therefore a higher workload. This is in line with the survey results regarding workload collected throughout the sessions.

Discussion

Overall, the simulation provided train operators with insight into how ATO implementation could affect their workflow and allowed them to assess its costs and benefits. The qualitative results show varying perspectives on how ATO will impact work operations, ranging from positive to negative, with most responses falling between indifferent and negative.

We recognise the limitations present in this study. During debriefing sessions, traffic/network controllers indicated that they needed to adjust to the unfamiliar workstation, particularly regarding the geographical area and timetable. This unfamiliarity, combined with simulation bugs, could have influenced the perceived workload. Specifically, we believe the workload was overestimated, particularly in the first and second sessions. Some train drivers also reported experiencing a higher workload due to missing information in the simulation environment (e.g., no location signs on the track). Nevertheless, this simulation serves as a first step in recognising and identifying human factors issues for the future of automated systems in railways

References

Brandenburger, N., Dreßler, A., & Grippenkoven, J. D. (2023). *DLR Workload Assessment Tool (DLR-WAT)-Official English Version*. Retrieved from https://elib.dlr.de/201450/1/2023_DLR-WAT_Brandenburger_Dressler_Grippenkoven.pdf

Balfe, N., Sharples, S., Wilson, J.R., 2018. Understanding is key: An analysis of factors pertaining to trust in a real-world automation system. *Hum. Factors* 60 (4), 477–495.

Farrington-Darby, T., Wilson, J.R., Norris, B.J., Clarke, T., (2006) A naturalistic study of railway controllers. *Ergonomics, 49*(12–13), 1370–1394.

Hillege, R.H.L., Lo, J.C., Janssen, C.P., Romeijn, N. (2020). The Mental Machine: Classifying Mental Workload State from Unobtrusive Heart Rate-Measures Using Machine Learning. In: Sottilare, R.A., Schwarz, J. (eds) Adaptive Instructional Systems. HCII 2020. Lecture Notes in Computer Science, vol 12214. Springer, Cham. https://doi.org/10.1007/978-3-030-50788-6_24

Körber, M. (2018, August). Theoretical considerations and development of a questionnaire to measure trust in automation. In *Congress of the International Ergonomics Association* (pp. 13-30). Cham: Springer International Publishing.

Kusumastuti, S. A., Kolkman, T. H., Lo, J. C., & Borsci, S. (2025). Charting the landscape of rail human factors and automation: A systematic scoping review. *Transportation Research Interdisciplinary Perspectives*, *30*, 101350.

Lewis, J. R. (2002). Psychometric evaluation of the PSSUQ using data from five years of usability studies. *International Journal of Human-Computer Interaction*, *14*(3-4), 463-488.

Lo, J. C., Sehic, E., Brookhuis, K. A., & Meijer, S. A. (2016). Explicit or implicit situation awareness? Measuring the situation awareness of train traffic controllers. *Transportation research part F: traffic psychology and behaviour*, *43*, 325-338.

Lo, J. (2020). *Gaming Simulation and Human Factors in Complex Socio-Technical Systems: A Multi-Level Approach to Mental Models and Situation Awareness in Railway Traffic Control.* [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/ 10.4233/uuid:93bba631-f675-4b28-8780-3bd9d686680a.

Nederlandse Spoorwegen. (2024, December). *Spoorkaart 2025: hier te downloaden*. https://nieuws.ns.nl/spoorkaart-2025-hier-te-downloaden/

Sharples, S., Millen, L., Golightly, D., Balfe, N., 2011. The impact of automation in rail signalling operations. *Proceedings of the Institution of Mechanical Engineers, Part f: Journal of Rail and Rapid Transit 225*(2), 179–191.