Experimental study to investigate mental workload of local vs remote operator in human-machine interaction

Melanie Zimmer, Ali Al-Yacoub, Pedro Ferreira, Ella-Mae Hubbard and Niels Lohse

Loughborough University, Wolfson School, Intelligent Automation Centre, Loughborough, UK

ABSTRACT
A new Coronavirus disease 2019 has spread globally since 2019. Consequently, businesses from different sectors were forced to work remotely. At the same time, research in this area has seen a rise in studying and emerging technologies that allow and promote such a remote working style; not every sector is equipped for such a transition. The manufacturing sector especially, has faced challenges in this respect. This paper investigates the mental workload (MWL) of two groups of participants through a human-machine interaction task. Participants were required to bring a robotised cell to full production by tuning system and dispensing process parameters. Following the experiment, a self-assessment of the participants' perceived MWL using the raw NASA Task Load Index (RTLX) was collected. The results reveal that remote participants tend to have a lower perceived workload compared to the local participants, but mental demand was deemed higher while performance was rated lower.

1. Introduction

Under COVID-19, several industries have realised the need for remote working practices. While for some sectors the move from working locally to remotely has been swift, others have faced more challenges doing so. Remote working brings its own unique challenges for workplaces, not just for companies but also for individuals. This holds particularly true in manufacturing where particular processes, such as corrective maintenance or ramp-up, require unpredictable physical work. Any delay here can lead to significant delays in terms of time and, consequently, result in a financial loss. Under these restrictions, the idea of remote collaboration or a remote expert working with a local operator to achieve a common goal has become a more accepted one. Given these assumptions, the presented research conducted a manufacturing ramp-up experiment that extends the paper originally presented at the 18th International Conference in Manufacturing Research ICMR 2021 (Zimmer et al., 2021). In a remote expert scenario, the remote experts provide their knowledge by interacting with their co-workers and systems using technology (Wang et al., 2021). While for the local operator the focus will be more on the physical aspects of the task at
hand, the remote operator’s role is more the one of a decision-maker. The quality of the decisions made by the remote expert, or their general performance, not only depends on their knowledge but also their mental workload (MWL). By knowing more about the MWL of the remote expert, additional aspects can be addressed, such as ‘How complex are the tasks that the operator is required to perform? Can any additional tasks be handled above and beyond those that are already imposed? How many people are needed to successfully carry out the task?’ (Young et al., 2015). Therefore, operator and system performance can be predicted based on the quantified mental cost of performing tasks (Kantowitz, 2000). But the impact on the cognitive overhead of a remote expert is yet understudied. In this paper, the perceived MWL of human operators is analysed for remote and local participation after bringing an industrial robotised gluing workstation to full production. It is hypothesised that the task is less demanding for remote participants in terms of perceived workload than undertaking this task locally. One of the well-established ways to measure the MWL is the NASA-Task Load Index (NASA-TLX; Hart & Staveland, 1988), which is based on six sub-scales associated with the MWL, namely mental, physical, and temporal demand (related to demands imposed on the human), as well as performance, effort and frustration (related to interaction of the human with the task). The rest of the paper is structured as follows: a brief literature review is presented in Section 2. Section 3 outlines the methodology that was applied for this study, including a description of the task, recruited subjects and experimental setup. Section 4 describes the experimental procedure and data collection. The results and discussion are presented in Section 5. Finally, Section 6 provides conclusions and an outlook for future work.

2. Literature review

2.1. Mental workload

Due to the growing use of computerised and semi-automated technologies in both administrative and manufacturing tasks, the notion of MWL has grown in importance to address the difficult demands on the human’s mental or information-processing abilities (ILO Encyclopaedia of Occupational Health and Safety, 2021; Yagoda, 2010). Thus, people are becoming increasingly conscious of cognitive limits, which can have a significant detrimental influence on productivity results (Thorvald et al., 2017) and safety (Gualtieri et al., 2022). Huey and Wickens (Huey & Wickens, 1993) have shown that high task demands lead to reduced performance and productivity, increased response times and errors, as well as changes in task performance strategies. MWL is often considered as a multi-dimensional concept, which is characterised by the operator and task as well as the environmental context (Young et al., 2015), and, as such, no single definition of the term ‘mental workload’ can be found in the literature (Cain, 2007). However, the number of tasks, the time required to do these tasks, as well as the subjective experiences of the human all seem to be associated aspects (Lysaght et al., 1989). Under consideration of these aspects, Young and Stanton (as cited in (Stanton et al., 2005), Ch. 39) describe the MWL as ‘the level of attentional resources required to meet both objective and subjective performance criteria, which may be mediated by task demands, external support,
and past experience’. To address and increase the key targets of ergonomics, such as the level of efficiency, satisfaction, safety and comfort in the workplace, assessing the MWL is a critical component in the enhancement of human-machine interfaces (Rubio et al., 2004).

### 2.2. Cognitive load in human-machine interaction

As humans may engage more with scalable industrial robots in future work systems, Kaufeld and Nickel (Kaufeld & Nickel, 2019) conducted a mixed reality task environment study with 20 participants to examine the impact of varied design criteria, related to human factors and ergonomics as well as occupational safety and health, on human MWL in human-robot interactions (HRIs). The experimental scenario depicts a production setting in which a human operator interacts with two virtual robots while executing grammatical reasoning tasks. Robot autonomy was constrained in this low level of robot autonomy (LORA) situation, and they were required to respond to human task demands. The authors concluded that combining a lower LORA with audio-visual information regarding upcoming HRI led humans to be less distracted from task performance, resulting in less impairment in operator workload.

Gualtieri et al. 2022 (Gualtieri et al., 2022) studied the role of cognitive ergonomics in HRI. Here, a set of cognitive ergonomics variables (amongst them cognitive workload), which have been identified in the literature, have been experimentally validated. The experiment involves three different scenarios: Only the essential features required to meet safety standards have been implemented in the first scenario, while interaction between the human and the robotic system are kept to a minimum. In the second scenario, additional features have been presented to enhance the quality of HRIs, but without allowing the participant to choose features such as type of command and robot speed. The implementation in the third scenario included more advanced features than the other scenarios, and the participants could interact with the robot physically or through a virtual push button. The results showed that when the three scenarios were used, the cognitive experience of 14 participants in the collaborative assembly was promoted by improving interaction settings and workstation characteristics. The most noticeable enhancements occur for changes between scenarios 1 and 2 according to the quantitative and qualitative data. Higher robot autonomy, greater synchronisation with robot duties, the ability to have more control over the system, and improved knowledge of workstation and robot condition are likely to be the key contributors to this development.

In their experimental study, Fraboni et al. (Fraboni et al., 2022) evaluated how collaborative robotic system characteristics affected workers’ reported cognitive workload, visual attention and usability. Individual dissimilarities among their 14 participants implied that robots should be able to tailor their behaviour to the needs of each individual. Furthermore, their results showed that using more human-like movement patterns, such as Minimum Jerk Trajectories, as well as giving the operator the possibility to set the robot’s pace and choose the favoured mode of interaction with the robot, increased usability and reduced perceived workload as a sense of familiarity and predictability could be achieved.
In a systematic literature review, Nelles et al. 2019 (Nelles et al., 2019) reported on the available research on evaluation metrics for human well-being in HRI. In this literature review, the authors emphasized that the experimental design, the questionnaires, and the measures employed are heterogeneous. Well-being is the state of feeling comfortable, healthy, and happy, and it is connected with the ability of an individual to manage stress.

2.3. Local and remote collaboration in manufacturing tasks

In a local and remote operator setting, the local operator is immersed in the environment where the problem occurs while also having the ability to undertake physical changes, in contrast to the remote operator who is more knowledgeable in terms of how to address the problem under consideration (Gurevich et al., 2015). There are different ways how both operators can work together to achieve their common goal. In terms of a remote expert and local operator working collaboratively, this can simply be done in the form of conversations and information exchange between two workers working on an assembly task (Flor, 1998). Those conversations mainly identified the tasks’ goal, the instruction for tasks, and the tasks’ completion. Further studies introduced a typical collaborative work that involves sketches and writing (Tang, 1991). It has been noted that this standard view supported the work process and communicated information adequately. Other studies tried to determine which visual information provides a benefit for a coworking team. The study by Kraut et al. (Kraut et al., 2003) examined an operator’s performance working alone on a bicycle repair task with a group of remote mechanics and a local operator who was working on the same task. The experiment presented evidence of the effect of a shared visual context in remote collaborative work. Gergle et al. (Gergle et al., 2013) further studied puzzle tasks in which the shared visual context was also crucial for situation awareness and a common understanding. Oftentimes, humans use gestures as an additional means to communicate their message. Thus, further research (Fussell et al., 2004) suggests that communication can be still more enhanced by showing the gestures of the expert as part of the environment the operator works in instead of providing complex descriptions.

Based on the presented introduction and the literature review, it is hypothesized that, in terms of perceived workload, the task to be undertaken is less challenging for remote participants than for local ones. In particular, physical demand and effort would be expected to be lower, but how will achieved results in the local and remote scenarios compare, also in light of perceived performance? Furthermore, the feedback from remote participants about their experience could highlight ways to enhance future remote interaction in industrial settings.

3. Methodology

3.1. Task

Participants were asked to tune a robotised Gluing Workstation (GWS) to volume production for three different products (Figure 1) while also assessing certain Key Performance Indicators (KPIs) related to functionality, quality and performance. This ramp-up process should be undertaken by participants as quickly as possible, meaning
using the least number of trials. There is no limitation on how many trials can be undertaken per ramp-up process, but for each product type at least one product needs to be produced considering that a defined target product quality as well as other KPIs need to be reached for each one. A walkthrough of the individual steps of the task is given in the section ‘Experimental Procedure’. After the practical part of the experiment was finished, participants were asked to fill in an online post-questionnaire to be able to capture more subjective feedback from the participants.

3.2. Subjects

For this study, participants have been assigned to two different scenarios that can be described as follows:

**Group A:** Participants in this group took part in the experiment in the lab environment. Change actions were done by the participants themselves.

**Group B:** Participants in this group took part in the experiment remotely. Therefore, experiment investigator undertakes the physical changes to the setup for them.

Differences and commonalities have been highlighted later in this section, see, Figure 6.

Participants were recruited from technical and non-technical backgrounds having little to no knowledge about the setup and its behaviour. In total, 6 female participants and 10 male participants took part. Participants’ age ranged from 20–29 (7x) to 30–39 (8x) and 60 and above (1x) years. Finally, 7 participants rated their experience with technological equipment very good, 7 good, 1 satisfactory, and 1 poor with the number of years worked in the field of automation or engineering ranging from 1 to 30 years.

3.3. Experimental setup

The main hardware components for the GWS setup are an ABB IRB 120 6-axis industrial robot with a two-finger SCHUNK gripper to manipulate a metal workpiece. To make the dispensing process possible, an automated time-pressure dispensing unit (Fisnar JB1113N) was used, which was connected to a syringe that has been mounted to the surrounding cell and contains the dispensing material. In addition, a Raspberry Pi3 has been fitted to the robot cell providing temperature and humidity data about the environment. The experimental setup also contained two computers, where one machine was used to run the robot control and local Graphical User Interface (GUI), and the other one hosted the GUI for remote participation where users could choose the parameter settings for the setup and any instructions to the local operator that were then both displayed to
the local operator through the local GUI. To ensure that remote participants would have the best possible and realistic experience, four cameras live-streaming different aspects of the process and a microphone were included in the setup (Figure 2).

An overview of the experimental variables is provided in Table 1. The robot can be parameterised to move at different speeds. Technically, industrial robot arms usually move in a trapezoidal manner, in which speed increases linearly until it reaches maximum speed, then stay constant until the robot gets close to the goal point, where it starts linearly decelerating until it reaches zero. The robot speed parameter in our experiment defined the maximum speed the robot can reach each trapezoidal, which impacts the overall duration of the task execution. Also, speed directly impacts the mechanical energy used to indicate the consumed energy to perform the task. The step size parameter indicates how close the robot can reach the target point within the trajectory. The attached gripper has open and close states, which are mainly needed to collect and release the metal workpiece. The dispensing process can take on the discrete values 5, 10, 16, 20, or 30 psi for setting the dispensing pressure for the glue dispenser. Moreover, the dispensing needle can be chosen to gain different line widths of 0.84, 1.2 or 1.6 mm.
4. Experimental procedure

This section will accommodate a walkthrough of the experimental procedure, along with a description of the implemented GUI. The GUI enables the participant to interact with the system, collect data and information, and provide feedback information to the participant in form of KPIs. The following subsections include the different stages of the ramp-up experiment, namely the pre-experimental, start, adjustment, test, assessment, end as well as post-experimental stages.

4.1. Pre-experiment

At the time when participants were recruited, high-level information about the system setup and an overview of the procedure was provided. On the day of participation and before the experiment started, a brief introduction explaining the aim of the experiment, the meaning of the ramp-up process, and a brief verbal explanation of the setup, its interaction, and expected behaviour were given. The introduction was always the same across participants in their respective groups; however, an additional explanation about the remote setup was given to participants in Group B and any answers before the experiment that participants might have had could differ.

4.2. Start of experiment

Once the participant was ready to start the experiment, the developed start page (Figure 3) was visible on the computer monitor. The participant was required to enter the provided user ID into the GUI. For remote participation, four live streams of the system were visible.
to the participants to the left of the start page and across the other views. At the top, the setting for the dispensing unit can be seen. The second stream shows the overall setup, whereas the third allows viewing the dispensing task closer. The final stream shows the workpiece holder and thus, the participants can see any product produced in the particular trial. Once the participant decided to start the actual experiment by clicking the ‘START Experiment’-button, the adjustment page is displayed (Figure 4).

4.3. Adjustment stage

After the start of the ramp-up process itself, the created adjustment page was presented to the participant (Figure 4). Hither, the participant had to choose and fill in the changeable equipment and process settings. As described in the experimental setup earlier, specific parameters for both processes and equipment modules could be freely selected to allow a specific range of adjustment. Based on the pattern chosen in the drop-down box, the picture with a good example outcome of the product was illustrated. In addition to the picture, environmental information in the form of current temperature and humidity

Figure 4. Adjustment page front end.
values and KPIs for the task duration, cycle time and mechanical energy were made available. Both should assist the participant in choosing the appropriate adjustment settings to deliver the desired result. If further information about the setup or potential troubleshooting issues were required, participants could consult a manual that had been collated prior to the experiment containing information about the different modules and troubleshooting advice. Participants in Group A were required to undertake any physical changes to the system themselves before the system could be tested under the current settings as the participation took place in person. For Group B, this did not apply due to remote participation. Here, once participants had made all the required choices and were ready to test their settings by clicking the 'Test Skill' button, the local experiment investigator received a notification indicating that the system’s physical changes should be made. Additional instructions could be given to the experiment investigator through a textbox in the developed GUI to request a replacement of the nozzle or a cleaning of the nozzle tip, for example, in case issues were encountered on the physical setup that would affect the performance of the task. These requests were then submitted along with the chosen parameters to the experiment investigator once the 'Test Skill'-button was clicked. By using this textbox approach, interaction between the remote participant and experiment investigator was minimised to ensure bias was reduced, and participants across the groups were treated as equally as possible. The remote participant could follow all the interactions occurring by the local operator with the system through the live stream. The experiment investigator would then send the system to test.

4.4. Test stage

The task execution could either directly be observed by the participants when taking part locally or be followed through the installed cameras. The robot was first sent from its standby position for the test run to pick up the fixed metal plate. After the plate had been collected, the robot moved to the dispensing position, where the dispensing unit was triggered to release air pressure, and the robot performed the chosen pattern. Once the robot had completed this task, the dispensing unit received a command to discontinue the airflow. The finished workpiece was put back to the holder and released by the robot, where the installed camera took a picture of the produced outcome before the robot went back to its standby position.

4.5. Assessment stage

Following the dispensing task, the assessment view (Figure 5) would emerge. This view enabled the participant to evaluate the different outcomes of the trial run, such as an assessment of the achieved KPIs, which were calculated for each trial, compared to the target KPIs, the product quality, the process and equipment functionality, as well as the process and equipment performance. For all the mentioned aspects and any additional other comments, the participant could provide free-form text to capture the human experience. In addition, the equipment and process functionality were also captured using a simple radio button (yes/no). The product quality and equipment and process performance could also be rated with a drop-down box value (very bad, bad, okay good, very good). In terms of product quality, good quality was defined by straight and continuous lines, with no excessive dispensing material and close similarity to the given target pictures.
As shown in Figure 5, this participant had chosen the frame product for this trial run. In this run, the required KPIs were not quite achieved (i.e. slightly higher duration, lightly lower cycle time and mechanical energy as well as bad product quality). After a visual inspection by the operator, the produced product was also deemed unsatisfactory due to excessive material resulting in curly lines. This means that the selected parameters/adjustments by the participants to create a product have not yet been produced the targeted qualities; thus, changes are needed.

At this point, the participant would be required to do another trial (‘Make Adjustment’ button), which would take the participant back to the adjustment view (Figure 4). In the case where the required functionality, product quality and performance were achieved for all three products, the experiment could be ended (‘END Experiment’ button).

4.6. End of experiment

Once the participant had determined to conclude the experiment, the end page would appear, thanking the participant for their time and officially designating the finish of this part of the experiment.

4.7. Post-experiment

After this part of the experiment was finished, participants were asked to fill in an online post-questionnaire. The post-questionnaire consisted of different types of questions, such as multiple choice, matrix, drop-down, open-ended and demographic questions. Participants, amongst others, had to rate which produced outcome was the best for each product type and why (open questions), declare which information they used for the decision-making process (multiple choice), describe how they experienced the ramp-up
Figure 6. Flow chart for experimental procedure, highlighting commonalities and differences between Groups A and B.
process (multiple choice) as well as written assessment (open questions), and provide (anonymised) personal information about their knowledge of the English language, their occupation, age range as well as rate their expertise with technological equipment and automation. Most importantly for this study, ratings (drop-down list) for the measures of the RTLX were also included. In addition, participants in Group B were also asked several questions about their experience with their remote participation (multiple choice, matrix, open questions). The analysis of the post-questionnaire is provided in the next section.

To summarise, the diagram shown in Figure 6 graphically illustrates the general flow of the experimental procedure for the different stages to highlight the commonalities and differences for Groups A and B. As can be seen, most steps are the same for both groups. Slight changes, however, exist in the induction, where some additional information about the remote setup was given to participants in Group B. Furthermore, differences existed in the undertaking of hardware changes and sending the system to test, where the in-presence experiment investigator assisted the participants in Group B. And, finally, the post-questionnaire slightly differed for Group B, where additional questions about the remote experience were added.

5. Results and discussion

In total, 282 change-cycles related to this experiment’s ramp-up process were obtained from 16 participants, where 104 were carried out in Group A and 178 in Group B. Figures 7 and 8 illustrate the final products produced by local participant number 1 and remote participant number 9, respectively. This indicates that local and remote participants could achieve similar end products.
To obtain the perceived workload estimations for the ramp-up experience during this experiment and help assess the proposed decision-support’s usability and effectiveness from the participants’ viewpoint, the widely accepted and used NASA-Task Load Index (NASA-TLX; Hart & Staveland, 1988) was employed in its raw or unweighted form in an online questionnaire. The index takes into consideration the six various subscale categories mental demand, physical demand, temporal demand, performance, effort, and frustration, which are scored by the individual directly after performing the task on a scale from 0 (good) to 20 (poor) for performance and 0 (low) to 20 (high) for the other categories. These are valuable insights that cannot undoubtedly be achieved through the other collected experimental data. However, using this measure is also not entirely without limitations as participants may not fully recall their experience. Figure 9 reveals the averaged results for the individual subscales across both groups.

Averaging the different subscales for the various groups to obtain an overall score results in $M_A = 11.08$ and $M_B = 7.8$ for Group A and B, respectively. As was assumed, this reveals that the perceived workload was overall less for the remote participants, while the local participants felt higher workload based on their results. The effort was deemed lower by the remote participants ($M_{B\_Effort} = 8.3, SD_{B\_Effort} = 4.19$) than for the local participants ($M_{A\_Effort} = 11.67, SD_{A\_Effort} = 2.58$). A big difference between local
and remote participation can be seen in the physical demand ($M_{A,\text{Physical}} = 7.17$, $SD_{A,\text{Physical}} = 6.79$; $M_{B,\text{Physical}} = 1.6$, $SD_{B,\text{Physical}} = 2.07$), which might also link to the temporal demand, which was higher in the local group as well ($M_{A,\text{Temporal}} = 11.33$, $SD_{A,\text{Temporal}} = 5.43$; $M_{B,\text{Temporal}} = 4.7$, $SD_{B,\text{Temporal}} = 4.79$). The mental demand was surprisingly deemed slightly higher in the group of remote participants ($M_{A,\text{Mental}} = 11.5$, $SD_{A,\text{Mental}} = 4.89$; $M_{B,\text{Mental}} = 12.2$, $SD_{B,\text{Mental}} = 4.26$). This high mental demand might happen due to the need for remote participants to be engaged in the experiment. The individual’s own performance was specifically well-rated for local participation participants compared to the remote group ($M_{A,\text{Performance}} = 9.83$, $SD_{A,\text{Performance}} = 6.82$; $M_{B,\text{Performance}} = 8.3$, $SD_{B,\text{Performance}} = 5.46$). This could be because the local operators have better feedback about the product and feel more immersed in the environment.

The highest average across the different scales was obtained in the frustration category, particularly for Group A participants ($M_{A,\text{Frustration}} = 15$, $SD_{A,\text{Frustration}} = 4.56$; $M_{B,\text{Frustration}} = 11.7$, $SD_{B,\text{Frustration}} = 4.6$). On the one hand, this could be explained by the in-person participation for the first group, which had to do the physical changes to the system themselves. On the other hand, this might also indicate that the immersion of the remote participants in the experiment was not as high.

A two-sample t-test assuming equal variances was also performed to determine if there was a statistically significant difference in perceived MWL between local and remote participants. Results show that significant effects were found for perceived workload measured by the RTLX between the two groups ($t(14) = 2.43, p = 0.03$).

Figure 9. Average scores with standard deviation for subcategories of raw NASA-TLX for both groups.
Other results obtained from the post-questionnaire are reported in the following. Surprisingly, the remote participants and local participants highlight almost the same remarks about the difficulties they have faced during the experiments, for example, several participants from remote and local groups have stated that they had difficulties in finding logical relations between parameters and outcomes. Also, when participants have been asked what additional information could be added to the setup that they think would support their decision-making process, 20% and 33.33% of remote and local participants have mentioned that access to historical data would assist them in performing the task and accelerate the ramp-up process.

Another interesting aspect, when participants have been asked about the information, they have used to bring the modules to the production phase, both groups relied on KPIs and the product quality as feedback. While remote participants have also indicated that they used the manual as an information source, this was not observed in local participants. Moreover, when participants were asked whether they used any strategy to tackle the task, answered yes. 83.33% of local, and 100% of remote participants. This indicates an excellent engagement from the remote participants. Most of the participants highlighted that they fixed some and experimented with the other parameters.

Remote participants stated that they were involved in the setup for various reasons. 60% of remote participants have stated that visual feedback and system responsivity have helped them to be engaged with the experiment. At the same time, 90% of answers indicate that system responsivity is why they felt engaged with the experiment. 20% have highlighted that audio, visual, and system responsivity assists them in being engaged during the experiment. During the experiment, it was noticed that the audio quality was not ideal due to background noise of other industrial equipment so that auditory clues about the dispensing unit could not always be heard. This was confirmed as several participants have suggested improving the setup by including a better audio quality. Also, 30% of remote participants have suggested using Virtual Reality (VR) for a better engagement with the remote setup. Moreover, 80% of remote participants have deemed their experience with the setup relaxed, while 20% have considered it boring and 20% have considered it stressful.

6. Conclusions and future work

In this research, the perceived mental workload for a robotised dispensing process in a local and remote context was examined using the raw NASA-TLX (RTLX). Results from the RTLX revealed that the perceived workload was overall lower for remote participants. However, surprisingly the perceived mental demand was rated higher while the perceived performance was rated lower in comparison to the local participants. Nonetheless, literature might provide an explanation: As Cain (Cain, 2007) states, despite the challenges of retaining focus in monitoring tasks, the workload can be seen as modest. Therefore, while well known, the disconnect between workload and efficiency remains poorly understood. What is more, NASA-TLX is not specifically tailored to the needs of manufacturing application overall. In manufacturing applications, assembly operations’ increasing complexity and dynamic demands lead to risen cognitive load in assembly workers. Previous studies have outlined the complexity of an assembly worker’s situation in terms of difficulty and speed of work, and there have been a few attempts to understand the cognitive load. Therefore, another cognitive
load assessment method, such as the *Cognitive Load Assessment in Manufacturing* (CLAM; Thorvald et al., 2017, 2019) could be further investigated in the future. Despite NASA-TLX being the basis for the development of CLAM, CLAM employs manufacturing-specific terminology, primarily being focused on assembly. The contribution of that study is that it reduces the cognitive load of workers on the shop floor while performing assembly tasks. The CLAM method does not mandate any expert knowledge, as the focus has been on the practitioners, applicability, and usability of the tool in practice.

A limitation of the presented study can be seen in the lack of a control group, and, thus, a missing baseline to compare the obtained results with. However, participants had been recruited to have little prior knowledge about the task and the setup itself, so a within-subject design would have allowed participants to build some knowledge from one scenario to the other. In addition, the practical part of the study, i.e. without the induction and post-questionnaire, could in cases be quite time-consuming as well, resulting in an average duration for participants in Group A of 68.2 minutes and 62.7 minutes for participants in Group B. Future studies could consider this in their design.

Using the perceived workload as a measure can highlight specific issues in terms of the operator’s physiology or safety but applying self-reflective metrics such as the NASA-TLX are not time-sensitive for time-critical operations. Hence, including data from eye movements or heart rate can be promising approaches, which will be further examined as part of this work in the future. Furthermore, experiences such as virtual or augmented reality (AR) also seem interesting avenues. Buchner et al. 2021 (Buchner et al., 2022), for example, presented a study that aims to systematically answer the question of the impact of AR on cognitive load, including performance, when used in learning settings. The results show that in comparison to other technologies, such as technical manuals, 2D displays, audio instructions and immersive VR, AR appears to be less cognitively demanding and leads to higher performance. The crucial conclusions of this study are that, depending on the technology used, cognitive load can either be increased (e.g. paper/display based) or decreased (e.g. AR).

Additionally, the paradigm of Industry 4.0 incorporates a transfer towards intelligent operations (Thoben et al., 2017), in which artificial intelligence, robotics, and automation enhance human capacities and counteract their weaknesses. This will facilitate the processes to have higher levels of safety, enhanced productivity and reduced mental load of human operators, which are fundamental elements for more intelligent workplaces.

**Disclosure statement**

No potential conflict of interest was reported by the author(s).

**Funding**

This work was supported by the Engineering and Physical Sciences Research Council [EP/L014998/1].
References

Buchner, J., Buntins, K., & Kerres, M. (2022). The impact of augmented reality on cognitive load and performance: A systematic review. *J. Comput. Assist. Learn.*, 38(1), 285–303. https://doi.org/10.1111/jcal.12617

Cain, B. (2007). A review of the mental workload literature. Defense Research and Development Canada Report #RTO-TR-HFM-121-Part-II, D. R. Dev. Toronto, no. 1998, pp. Toronto, Toronto, http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA474193

Flor, N. V. (1998). Side-by-side collaboration: A case study. *Int. J. Human—Computer Stud.*, 49(3), 201–222. https://doi.org/10.1006/ijhc.1998.0203

Fraboni, F., Gualtieri, L., Millo, F., De Marchi, M., Pietrantoni, L., & Rauch, E. (2022). Human-robot collaboration during assembly tasks: the cognitive effects of collaborative assembly workstation features, Black, N.L., Neumann, W.P., Noy, I. *Proc. 21st Congr. Int. Ergon. Assoc. (IEA 2021)*. IEA 2021. *Lect. Notes Networks Syst.*, vol. 223.

Fussell, S. R., Setlock, L. D., Yang, J., Ou, J., Mauer, E., & Kramer, A. D. I. (2004). Gestures over video streams to support remote collaboration on physical tasks. *Human–Computer Interact.*, 19(3), 273–309. https://doi.org/10.1007/s10606-007-9010-9

Gergle, D., Kraut, R. E., & Fussell, S. R. (2013, January). Using visual information for grounding and awareness in collaborative tasks. *Human–Computer Interact.*, 28(1), 1–39. https://doi.org/10.1080/07370024.2012.678246

Gualtieri, L., Fraboni, F., De Marchi, M., & Rauch, E., ‘Evaluation of variables of cognitive ergonomics in industrial human-robot collaborative assembly systems’, Black, N.L., Neumann, W.P., Noy, I. *Proc. 21st Congr. Int. Ergon. Assoc. (IEA 2021)*, pp. 266–273, 2022.

Gurevich, P., Lanir, J., & Cohen, B. (2015). Design and implementation of teleadvisor: A projection-based augmented reality system for remote collaboration. *Computer Supported Cooperative Work (CSCW)*, 24(6), 527–562. https://link.springer.com/article/10.1007/s10606-015-9232-7

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In P. A. Hancock & N. B. T.-A. in P. Meshkat (Eds.), *Human mental workload* (Vol. 52, pp. 139–183). North-Holland.

Huey, B. M., & Wickens, C. D. (Eds.). (1993). *Workload transition - implications for individual and team performance*. National Academy Press.

ILO Encyclopaedia of Occupational Health and Safety, ‘Mental workload’. https://www.iloencyclopaedia.org/k2-feed1/item/628-mental-workload (accessed Apr. 09, 2021)

Kantowitz, B. H. (2000). Attention and mental workload. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 44(21), 456–459. https://doi.org/10.1177/15492004002421

Kaufeld, M., & Nickel, P. (2019). Level of robot autonomy and information aids in human-robot interaction affect human mental workload—an investigation in virtual reality International Conference on Human-Computer Interaction., (Springer). 278–291.

Kraut, R. E., Fussell, S. R., & Siegel, J. (2003). Visual information as a conversational resource in collaborative physical tasks. *Human–Computer Interact.*, 18(1–2) 13–49. https://doi.org/10.1027/152327051HIC1812

Lysaght, R. J., Hill, S. G., Dick, A. O., Plamondon, Brian D., Linton, Paul M. (1989). Operator workload: Comprehensive review and evaluation of operator workload methodologies TR- 851 (Alexandria, VA, USA: ANALYTICS INC WILLOW GROVE PA). https://apps.dtic.mil/sti/pdfs/ADA212879.pdf
Nelles, J., Kwee-Meier, S. T., & Mertens, A., ‘Evaluation metrics regarding human well-being and system performance in human-robot interaction – a literature review’, Bagnara, S., Tartaglia, R., Albolino, S., Alexander, T., Fujita, Y. Proc. 20th Congr. Int. Ergon. Assoc. (IEA 2018). IEA 2018. Adv. Intell. Syst. Comput., vol. 825, 2019.

Rubio, S., Diaz, E., Martin, J., & Puente, J. M. (2004). Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods. Appl. Psychol, 53(1), 61–86. https://doi.org/10.1111/j.1464-0597.2004.00161.x

Tang, J. C. (1991). Findings from observational studies of collaborative work. Int. J. Man. Mach. Stud, 34(2), 143–160. https://doi.org/10.1016/0014-0139(91)90039-A

Thorvald, K., Wiesner, S., & Wuest, T. (2017). “Industrie 4.0” and Smart Manufacturing – A Review of Research Issues and Application Examples. Int. J. Autom. Technol, 11(1), 4–16. https://doi.org/10.20965/ijat.2017.p0004

Thorvald, P., Lindblom, J., & Andreasson, R. (2017). CLAM–A Method for Cognitive Load Assessment in Manufacturing. Advances in Manufacturing Technology XXXI(6), 114–119. https://doi.org/10.3233/978-1-61499-792-4-114

Thorvald, P., Lindblom, J., & Andreasson, R. (2019). On the development of a method for cognitive load assessment in manufacturing. Robot. Comput. Integr. Manuf, 59, 252–266. 0736-5845. https://doi.org/10.1016/j.rcim.2019.04.012

Wang, B., Liu, Y., Qian, J., & Parker, S. K. (2021). ‘Achieving Effective Remote Working During the COVID-19 Pandemic: A Work Design Perspective’. Applied Psychology, 70(1), 16–59. https://doi.org/10.1111/apps.12290

Yagoda, R., 'Development of the Human Robot Interaction Workload Measurement Tool (HRI-WM)', in Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2010, pp. 304–308.

Young, Mark S., Stanton, Neville A. Stanton, N., Hedge, A., Brookhuis, K., Salas, E., & Hendrick, H., Eds. (2005). Mental workload. Handbook of Human Factors and Ergonomics Methods. CRC Press. 416–426. https://www.taylorfrancis.com/chapters/edit/10.1201/9780203489925-50/mental-workload-mark-young-neville-stanton

Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: Mental workload in ergonomics. Ergonomics, 58(1), 1–17. https://doi.org/10.1080/00140139.2014.956151

Zimmer, M., Al-Yacoub, A., Ferreira, P., Hubbard, E.-M., & Lohse, N. (2021). Mental workload of local vs remote operator in human-machine interaction case study. Advances in Manufacturing Technology XXXIV. 33–38. https://web.archive.org/web/20210911020151id_/https://ebooks.iospress.nl/pdfdoi/10.3233/ATDE210008.