Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 59-10 (2025) 981–986

Evaluating the Effectiveness of Subjective Questionnaires for Assessing Cognitive Well-Being in Assembly Tasks

Eva Geurts * **Anouk Michiels** * **Gustavo Rovelo Ruiz** *

* *UHasselt - Hasselt University, Digital Future Lab, Flanders Make, Diepenbeek, Belgium (e-mail: firstname.lastname@uhasselt.be).*

Abstract: Recognizing Industry 5.0's emphasis on human-centric work, we explored the use of established questionnaires, such as NASA-TLX, SWAT, and IMI, to evaluate cognitive well-being in assembly-like tasks. While expensive and invasive sensors can provide detailed insights, our aim was to determine how effectively existing, accessible questionnaires can detect factors such as boredom, cognitive load, temporal demand, and frustration. This information serves as a relevant contextual resource, enabling manufacturing companies to identify the root causes of well-being threats on the workfloor, particularly those linked to specific tasks. The results demonstrate that these questionnaires can capture key well-being dimensions, making them valuable for industrial settings. This supports their potential as practical, non-invasive tools for monitoring work-related well-being, aligning with the goals of a human-centered industrial future.

Keywords: Industry 4.0; Industry 5.0; Well-being; Manufacturing industry; Assembly

1. INTRODUCTION

Monitoring well-being in the workplace has become increasingly important as industries transition towards Industry 5.0, where human-centric approaches and sustainability play a central role (Yang et al. (2024)). Assembly tasks, in particular, pose unique challenges to worker well-being, with potential impacts on physical and cognitive load. Traditionally, well-being research in industrial settings has often focused on ergonomics, addressing how physical strain impacts health and performance. However, cognitive aspects—such as stress, fatigue, boredom, and attention—are equally crucial, influencing overall job satisfaction and productivity (Antonaci et al. (2024)). Solutions exist that involve sensor-based methods to monitor these factors, offering detailed insights into the mental and emotional states of workers (Park et al. (2020)). Yet, such technologies are not always feasible for widespread application due to costs, complexity, and concerns over privacy (Li et al. (2021)).

This research addresses the need for efficient cognitive well-being monitoring by measuring mental load and stress in real production settings. It supports personalized and adaptive manufacturing models, building on insights from sensor data (Park et al. (2020)) and process mining (Iriondo Pascual et al. (2022)). To enable effective personalization, we explore integrating micro-surveys into work routines by evaluating existing questionnaires to determine their suitability for subjective cognitive well-being monitoring. This study examines using single questions from established tools such as NASA Task Load Index (NASA-TLX) to assess well-being in assembly contexts. While these tools measure cognitive load, workload, and motivation, their task-level applicability in industry remains underexplored. Leveraging these questionnaires offers a non-invasive, cost-effective approach suitable for various industries. To isolate cognitive effects, we conducted a study with 24 participants performing assembly-like tasks designed to minimize physical strain, ensuring questionnaire responses reflected cognitive impacts alone.

2. RELATED WORK

Prior work has explored stress prediction in the workplace using surrounding stress data, such as colleagues' stress levels and an individual's stress history, improving accuracy (Muñoz et al. (2022)). Other studies have used physiological measures like heart rate variability (HRV) and electrodermal activity (EDA). Durantin et al. (2014) found that HRV and functional near-infrared spectroscopy (fNIRS) are sensitive to mental workload, while Setz et al. (2010) distinguished stress from cognitive load using EDA with up to 82.5% accuracy. However, these methods remain invasive and present technological challenges. Thorvald et al. (2019) addressed rising cognitive load in complex assembly tasks and developed the Cognitive Load Assessment for Manufacturing (CLAM), a practical tool to evaluate cognitive load without expert knowledge. Although CLAM offers a promising approach to cognitive ergonomics, more efficient interventions are needed to improve productivity and efficiency.

Many questionnaires assess subjective workload and cognitive well-being. A widely used one is the NASA-TLX (Hart (2006)), which evaluates workload across six dimensions: mental demand, physical demand, temporal demand, effort, performance, and frustration. Its simplicity and adaptability have made it a standard tool. Another is the Subjective Workload Assessment Technique (SWAT)

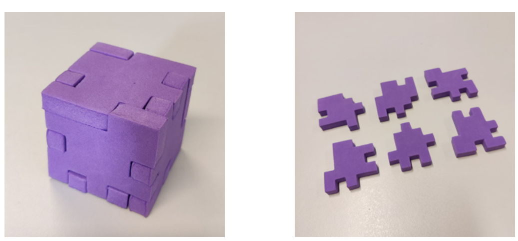
10.1016/j.ifacol.2025.09.166

(Reid and Nygren (1988)), which categorizes workload into time load, mental effort load, and stress load, making it particularly useful for time-sensitive tasks. The CLAM tool earlier described also integrates these questionnaires for cognitive well-being assessment.

While many technological solutions exist for detecting cognitive well-being factors like mental load, we aim to explore whether existing surveys can provide a low-cost, accessible solution for monitoring cognitive well-being in an assembly-like context.

3. METHODOLOGY

To simulate an assembly-like process while minimizing physical strain, we used Happy Cubes puzzles with six interlocking pieces that vary in complexity (András et al. (2013)). This ensured accessibility without requiring prior assembly knowledge. Two researchers tested all puzzles, recording the completion time to confirm significant complexity differences and select suitable tasks. Using Happy Cubes, we designed six distinct assembly tasks that target cognitive well-being factors such as frustration, satisfaction, boredom, cognitive load, and temporal demand. Each task explored how puzzle complexity and conditions influence these parameters. Figure 1 illustrates an assembled and disassembled puzzle (4cm x 4cm).



(a) Assembled variation of the puzzle

(b) Disassembled variation of the puzzle

Fig. 1. Example of an assembled and disassembled variation of the Happy Cubes puzzle.

3.1 Hypotheses

We put forward the following hypotheses to investigate parameters related to well-being that can be evaluated through subjective assessment (i.e. questionnaires):

H1: Questions related to **mental/cognitive demand** are suitable for detecting more complex tasks, as complex tasks require more cognitive demand.

H2: Questions related to **temporal demand** are suitable for detecting time pressure in assembly tasks, given that tasks with a time constraint lead to time pressure.

H3: Questions related to **frustration** are suitable for detecting assembly tasks that are unsolvable, as people struggle when they cannot complete a task.

H4: Questions related to **satisfaction** are suitable for detecting complex tasks, assuming that completing complex tasks leads to higher satisfaction levels than easy tasks.

H5: Questions related to **boredom** are suitable for detecting repetitiveness in assembly tasks, since repetitive tasks may lead to attention loss and boredom.

3.2 Study Apparatus

We defined six tasks to investigate our hypotheses:

Task A consists of two assignments that involve solving puzzles of **varying difficulty**. Task A1 features an easy puzzle, intended to provide a straightforward and engaging challenge that participants can complete with minimal cognitive effort. Task A2 ramps up the difficulty by presenting a more complex puzzle. Both tasks do not include instructions to guide participants or time constraints.

Task B involves another puzzle from a difficult level with **time pressure** introduced. A visible stopwatch is placed in front of the participants, and they are instructed to complete the puzzle as quickly as possible. The introduction of time pressure seeks to simulate a real-world environment where speed might be critical.

Task C is identical to Task B except that participants are given **explicit paper instructions** that illustrate step by step how to solve the puzzle. The puzzle contained the same complexity level as in Task B.

Task D is intended to be a frustrating experience, as participants are tasked with solving a puzzle that is inherently **unsolvable** due to the inclusion of one incorrect piece. Participants were instructed to stop attempting to solve the puzzle after 15 minutes (based on pilot studies), or earlier if they explicitly expressed, with reasons, that they had recognized the puzzle's unsolvability. No time pressure or instructions were present.

Task E involves an **easy** yet **repetitive** puzzle task with a specific variation: each puzzle piece has a figure printed on one side. Participants are instructed to first assemble the puzzle with all the figures facing inward. After completing this configuration, they must fully disassemble the puzzle before reassembling it with all the figures facing outward. This sequence—building the puzzle with figures inside, then outside—is repeated ten times in total, with a full disassembly required between each variation. The task aims to explore the impact of easy, but repetitive work.

Task F is intentionally designed to be cognitively challenging, requiring participants to tackle a complex puzzle while also estimating a two-minute duration without any external cues. This setup creates a **multi-tasking** scenario, demanding participants to split their attention between spatial problem-solving and time estimation.

3.3 Questionnaires

We selected the NASA-TLX (Hart (2006)), a subset of questions from the Intrinsic Motivation Inventory (IMI) (McAuley et al. (1989)), and the SWAT (Reid and Nygren (1988)) to assess cognitive well-being in relation to the tasks. The NASA-TLX (21-point scale) evaluates workload across six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration. The IMI (7-point scale) includes subscales on interest, competence, effort, value, pressure, and perceived choice, with only relevant subscales and questions included in our study: *This activity was fun to do*, *I thought this was a boring activity*, *This activity did not hold my attention at all*, *I think I am pretty good at this activity*, *I am satisfied with my performance at this task*, *I tried very hard on this activity*, *I did not feel nervous at all while doing this activity*, *I was very relaxed while doing this activity*, *I felt pressured while doing this activity*. The SWAT (3-point scale) measures time load, mental effort, and psychological stress. In addition, we developed a set of custom questions to explore alternative ways of assessing well-being: *Performing this activity was satisfying*, *The end result of this activity was satisfying*, *This activity was boring*, *This activity was frustrating*, *This activity was fun*, *I had to be focused to do this activity*, *I felt time pressure during this activity*, *I felt stressed regardless of the activity*, *This assembly activity was stressful*.

3.4 Protocol & Participants

Participants were individually invited and gave their informed consent before completing demographic questions and receiving a brief study explanation. We used a within-subject design with balanced Latin square randomization to counterbalance task order. For Task A, which included an easy (A1) and complex puzzle (A2), we randomized their order, but kept them under the same task to examine their relationship while avoiding distribution across all tasks. After each task, including A1 and A2, participants completed a set of questionnaires. Data for both A1 and A2 are analyzed separately. The experiment involved 24 participants (5 female, 19 male), recruited from a pool of students and researchers in our university, with an average age of 24 years (SD = 2.14) (Caine (2016)). The study was approved by the ethical committee. None of the participants were familiar with the Happy Cubes puzzles, and they all completed the six tasks.

4. RESULTS

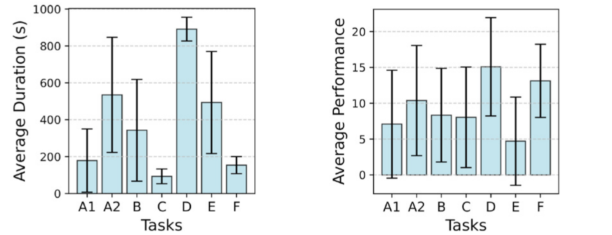
This section presents the results and discusses them in relation to the proposed hypotheses.

We evaluated the normality of the data using appropriate statistical tests. Given the ordinal nature of the data, we used the Friedman test for multiple comparisons and the Wilcoxon signed-rank test for pairwise comparisons. Where applicable, Bonferroni corrections were applied to adjust for multiple testing. Before presenting the results related to the hypotheses, we report on the overall performance of all tasks through the task duration that we measured during the experiment, and the perceived successfulness reported in the NASA-TLX for each task. Figure 2a shows an overview of the average duration per task and its standard deviation. Figure 2b shows the results for the perceived successfulness score as reported in the NASA-TLX per task.

4.1 Mental/Cognitive demand

H1: *Questions related to mental/cognitive demand are suitable for detecting more complex tasks.*

We investigated the two tasks of Task A to investigate this hypothesis given their difference in expected complexity, and investigated the outcomes related to *NASA-TLX Mental demand* and *SWAT Mental load*. The results of the Wilcoxon signed-rank tests show that Task A1 induces significantly lower subjective mental load compared to Task



(a) Task duration (b) NASA-TLX Performance

Fig. 2. Overall performance measures

A2. For the NASA-TLX mental demand category, Task A1 (median = 10.0) showed a significantly lower score than Task A2 (median = 13.0), W=18.0, p=0.0007. Similarly, for the SWAT mental load measure, Task A1 (median = 2.0) also demonstrated a significantly lower score compared to Task A2 (median = 2.0, but with noticeable individual differences), W=7.0, p=0.0016. These findings highlight the increased perceived mental load for Task A2, suggesting that it requires greater cognitive resources or effort compared to Task A1. The consistency across both the NASA-TLX and SWAT measures strengthens the evidence for this conclusion.

Similarly, we performed the same analysis for Task B and C since they only differ in the presence of instructions, making Task C significantly easier (as also visible in Figure 2a for the average completion time). The results of the Wilcoxon signed-rank tests support the hypothesis that Task C induces a significantly lower subjective mental load compared to Task B. For the NASA-TLX Mental Demand measure, Task B (median = 12.5) showed a significantly higher score than Task C (median = 5.0), W = 3.0, $p < 0.001$. Similarly, for the SWAT Mental Load measure, Task B (median = 2.0) demonstrated a significantly higher score compared to Task C (median = 1.0), $W = 7.5$, $p = 0.001$. The consistency across both the NASA-TLX Mental Demand and SWAT Mental Load measures highlights the evidence for this result. This aligns with the expectations that Task C, where instructions are present, places a significantly lower strain on participants' cognitive capacities. Given these results, we support hypothesis H1.

4.2 Time pressure/Temporal demand

H2: *Questions related to time pressure or temporal demand are suitable for detecting time pressure in assembly tasks.*

We used the three questions ("SWAT Time Load", "NASA-TLX Temporal Demand," and "Custom question: Time Pressure") to assess time pressure among the tasks. We compare Task B, C and F against A, D and E since they differ in the presence of time constraints. The median for each task per question is demonstrated in Figure 3, and reveals that there is a noticeable difference in NASA-TLX Temporal Demand and our custom question.

We performed Friedman tests, resulting in significant differences for each question:

NASA-TLX: Temp. Demand: $\chi^2(3) = 41.18$, $p < 0.001$
SWAT: Time Load: $\chi^2(3) = 39.08$, $p < 0.001$

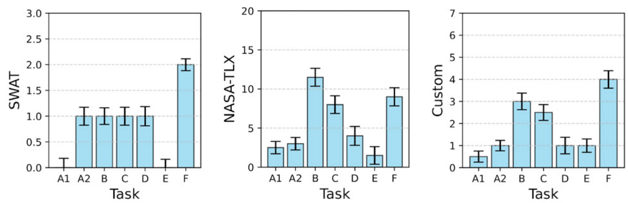


Fig. 3. Comparison of SWAT, NASA, and custom Time pressure medians

Custom: Time Pressure: $\chi^2(3) = 58.24$, $p < 0.001$

We conducted post-hoc pairwise comparisons using the Wilcoxon signed-rank test with Bonferroni correction ($p = 0.002$). The significant results of these pairwise comparisons are summarized in Table 1.

Table 1. Pairwise Wilcoxon signed-rank test results for NASA-TLX, SWAT, and custom question. (Task T = task with time pressure, Task NT = no time pressure)

Task T	Task NT	P-Value	Significant	W
NASA-TLX				
B	A1	$p < 0.001$	**True**	0.0
B	A2	$p < 0.001$	**True**	3.0
B	D	$p < 0.001$	**True**	16.0
B	E	$p < 0.001$	**True**	6.5
C	A1	0.002	**True**	29.5
F	A1	0.001	**True**	14.0
SWAT				
B	A1	0.004	**True**	0.0
F	A1	$p < 0.001$	**True**	5.5
F	A2	0.001	**True**	5.0
F	E	$p < 0.001$	**True**	0.0
Custom Question				
B	A1	$p < 0.001$	**True**	0.0
B	A2	$p < 0.001$	**True**	10.5
B	E	$p < 0.001$	**True**	3.5
C	A1	$p < 0.001$	**True**	4.0
C	A2	0.002	**True**	22.5
C	E	$p < 0.001$	**True**	19.5
F	A1	$p < 0.001$	**True**	16.0
F	A2	$p < 0.001$	**True**	2.5
F	D	0.001	**True**	14.5
F	E	$p < 0.001$	**True**	19.5

The post-hoc comparisons reveal that Task B shows a significant difference with every other task that does not have time constraints. The results for C and F show less encouraging results, which may indicate that limited complexity of task (Task C) and the 'unsolvability' of the task (Task F) impacts the findings. The SWAT time load reveals less promising results with nearly no significant differences among the tasks with and without time pressure, making it a less effective candidate for measuring temporal demand. The most significant differences can be found in our custom question related to time pressure. We only encountered no significant difference when comparing Task B and C to Task D. Our custom question shows to be a good candidate to capture time pressure in tasks (B, C, F) with a clear time constraint versus a solvable task without time constraints (A1, A2, and E). Since we did not encounter a significant difference in all comparisons, we do not support H2.

4.3 Frustration

H3: *Questions related to frustration are suitable for detecting assembly tasks that are unsolvable or cannot be successfully performed.*

We compare Task D with the other tasks since this task contains a wrong puzzle piece that makes it impossible to finish the puzzle. The duration of this task is also significantly longer than the other tasks (see Figure 2a). We performed a Friedman test for each to assess differences between tasks for the three questions related to frustration and psychological stress. The results showed significant differences for all three questions:

NASA-TLX: Frustration: $\chi^2(3) = 61.42$, $p < 0.001$
SWAT: Psychological stress: $\chi^2(3) = 50.35$, $p < 0.001$
Custom: This activity was frustrating: $\chi^2(3) = 65.07$, $p < 0.001$

Given the significant results, we conducted post-hoc pairwise comparisons using the Wilcoxon signed-rank test with Bonferroni correction ($p = 0.002$). The significant results of these pairwise comparisons are summarized in Table 2.

Table 2. Post-hoc Wilcoxon signed-rank test results for questions related to frustration.

Task 1	Task 2	p-value	Significant	W
NASA-TLX - Frustration				
A1	D	$p < 0.001$	**True**	8.5
B	D	0.001	**True**	30.5
C	D	$p < 0.001$	**True**	1.5
D	E	$p < 0.001$	**True**	0.0
D	F	$p < 0.001$	**True**	1.5
SWAT - Psychological stress				
A1	D	$p < 0.001$	**True**	10.0
C	D	$p < 0.001$	**True**	5.0
D	E	$p < 0.001$	**True**	2.0
Custom: This activity was frustrating				
A1	D	$p < 0.001$	**True**	8.0
B	D	0.001	**True**	25.0
C	D	$p < 0.001$	**True**	0.0
D	E	$p < 0.001$	**True**	0.0

The findings reveal that Task D generally resulted in higher levels of frustration and psychological stress across the measures (see Figure 4). However, the comparison with Task A2 did not reveal a significant difference. During our observations and looking at the performance reported in NASA-TLX (see Figure 2b), we also noticed that people struggled more to finish Task A2 compared to other tasks, probably also causing frustration. Given these findings, we cannot support H3, although the other tasks showed promising results. The NASA-TLX Frustation question seems to be most suitable to use for detecting frustration.

4.4 Satisfaction

H4: *Questions related to satisfaction are suitable for detecting complex tasks.*

We used the IMI - Satisfaction question and two of our custom questions related to satisfaction to investigate this hypothesis. We performed a Friedman test for each question to assess differences between tasks for satisfaction-

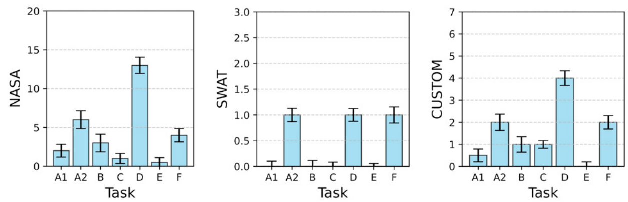


Fig. 4. Comparison of NASA, SWAT and Custom Frustration medians

related questions from the IMI and the custom questionnaire. The results showed significant differences for all three questions:

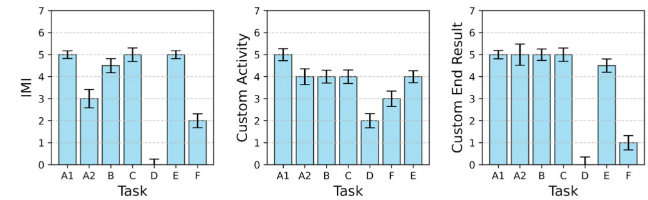
IMI: Satisfaction: $\chi^2(3) = 59.91$, $p < 0.001$
Custom: Performing this activity was satisfying: $\chi^2(3) = 32.85$, $p < 0.001$
Custom: The end result of this activity was satisfying: $\chi^2(3) = 47.53$, $p < 0.001$

Given the significant results, we conducted post-hoc pairwise comparisons using the Wilcoxon signed-rank test with Bonferroni correction ($p = 0.003$). The significant results of these pairwise comparisons are summarized in Table 3. We focus on comparing Task D (unsolvable puzzle) to Task A, B, C, and E since they were clearly solvable. Task F is left out, as it is doubtful whether this is perceived as a solvable task or not.

Table 3. Post-hoc Wilcoxon signed-rank test results - Questions related to satisfaction

Task 1	Task 2	p-value	Significant	W
IMI: I am satisfied with my performance				
A1	D	$p < 0.001$	**True**	2.0
A2	D	$p < 0.001$	**True**	28.5
B	D	$p < 0.001$	**True**	25.0
C	D	$p < 0.001$	**True**	0.0
D	E	$p < 0.001$	**True**	0.0
Performing this activity was satisfying				
A1	D	$p < 0.001$	**True**	12.0
B	D	0.001	**True**	16.5
C	D	$p < 0.001$	**True**	2.5
D	E	$p < 0.001$	**True**	14.0
The end result of this activity was satisfying				
A1	D	$p < 0.001$	**True**	7.5
A2	D	$p = 0.003$	**True**	22.0
B	D	$p < 0.001$	**True**	23.5
C	D	$p < 0.001$	**True**	6.0
D	E	$p < 0.001$	**True**	23.5

The findings demonstrate that Task D generally led to lower satisfaction across the measures (see Figure 5). Both *I am satisfied with my performance* and *The end result of this activity was satisfying* showed significant differences across all tasks, suggesting that these measures are more sensitive indicators of satisfaction than *Performing this activity was satisfying.* These results support the hypothesis that Task D induces lower satisfaction compared to the other tasks, leading us to confirm **Hypothesis H4**. However, as satisfaction may follow a U-shaped relationship with cognitive demand or complexity, this effect should be interpreted with caution (van Steenbergen et al. (2015)).



(a) IMI (b) Custom: Activity (c) Custom: Result

Fig. 5. Comparison of IMI and custom questions (activity, end result) medians for Satisfaction

4.5 Boredom

H5: *Questions related to boredom are suitable for detecting repetitiveness in assembly tasks.*

To assess differences across tasks for the boredom-related questions from the IMI and the custom question 'This activity was boring', we performed a Friedman test for each question, which revealed only a significant difference for the custom question;

IMI: Boring activity: $\chi^2(3) = 10.85$, $p = 0.093$
Custom: This activity was boring: $\chi^2(3) = 13.38$, $p = 0.037$.

However, no significant differences were found when conducting post-hoc pairwise comparisons using the Wilcoxon signed-rank test with Bonferroni correction. These results do not support **Hypothesis H5**, and therefore, we reject it. Boredom-related questions were not effective indicators for detecting repetitiveness in tasks. However, it is important to exercise caution in interpreting these results, as the tasks were new to participants. Thus, to more effectively investigate the repetitiveness of tasks, it may be necessary to include a greater variety of tasks in future studies.

4.6 Summary of Findings

Based on our findings, we recommend the following measures for assessing cognitive well-being in assembly tasks:

Mental Demand: NASA-TLX Mental Demand scale or SWAT Mental Load as a simpler alternative.
Temporal Demand: Custom question: *"I felt time pressure during this activity."*
Frustration: NASA-TLX Frustration scale.
Satisfaction: IMI Satisfaction scale or custom question: *"The end result of this activity was satisfying."*
Boredom: Further investigation needed.

5. DISCUSSION

Our experiment showed that both standardized and customized questionnaire items effectively detect cognitive well-being aspects in assembly tasks. These subjective measures complement objective data (e.g., sensor data), providing insights into operator experiences that affect performance. This demonstrates the potential of tailored questionnaires for well-being monitoring in industrial settings. The SWAT questionnaire, with its simple 3-point scales, is well-suited for micro-surveys on compact devices like smartwatches. However, it has limitations, and workstation screens could offer a better medium for more detailed questionnaires at appropriate times. The experiment used an assembly-like task designed to assess cognitive well-being while minimizing physical strain, enhancing the generalizability of our findings. Future studies should validate these results with a broader range of assembly tasks that include physical strain and focus more on objective performance metrics, ensuring the translation of findings to real industrial assembly settings.

Subjective data provides valuable context while also offering certain advantages and presenting challenges. One key advantage is privacy, as questionnaires can be managed locally, unlike sensor-based systems that often rely on cloud storage. However, data misrepresentation may occur if operators provide inaccurate responses due to job security concerns (Holden et al. (2015)), making trust-building essential for honest feedback. Additionally, time constraints pose a challenge, as collecting subjective responses can disrupt workflows; thus, identifying optimal moments and keeping questionnaires concise is crucial.

Future research directions include investigating the optimal timing and frequency for deploying micro-surveys to avoid overwhelming operators while ensuring sufficient data collection. Additionally, exploring correlations between factors such as perceived stress, performance, and effort by examining additional standardized questions could provide deeper insights. Also, developing methods to aggregate well-being data collected at the task level would enable analysis at the daily or job level and across multiple tasks, offering a more comprehensive understanding of cognitive well-being at work.

6. CONCLUSION

We performed an experiment (n=24) to investigate whether single questions from subjective questionnaires can be used to measure cognitive well-being in assembly tasks. The study revealed that customized and standardized questions provide valuable insights into cognitive well-being factors such as mental demand and time pressure which are critical for understanding factors influencing work performance. While challenges like data misrepresentation and time constraints exist, these can be mitigated through careful design and privacy-conscious implementation strategies. Future research should explore integrating these questions into micro-surveys and examining the relationship between subjective and objective well-being data.

ACKNOWLEDGEMENTS

This research was supported by Flanders Make, the strategic research centre for the manufacturing industry, in the project WELLFICIENCY.

REFERENCES

András, S., Sipos, K., and Soós, A. (2013). *Which is harder?-Classification of Happy Cube puzzles.* :.

Antonaci, F.G., Olivetti, E.C., Marcolin, F., Castiblanco Jimenez, I.A., Eynard, B., Vezzetti, E., and Moos, S. (2024). Workplace well-being in industry 5.0: A worker-centered systematic review. *Sensors*, 24(17), 5473.

Caine, K. (2016). Local standards for sample size at chi. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, 981–992.

Durantin, G., Gagnon, J.F., Tremblay, S., and Dehais, F. (2014). Using near infrared spectroscopy and heart rate variability to detect mental overload. *Behavioural brain research*, 259, 16–23.

Hart, S.G. (2006). Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 50, 904–908. Sage publications Sage CA: Los Angeles, CA.

Holden, R.J., McDougald Scott, A.M., Hoonakker, P.L., Hundt, A.S., and Carayon, P. (2015). Data collection challenges in community settings: insights from two field studies of patients with chronic disease. *Quality of Life Research*, 24, 1043–1055.

Iriondo Pascual, A., Smedberg, H., Högberg, D., Syberfeldt, A., and Lämkull, D. (2022). Enabling knowledge discovery in multi-objective optimizations of worker well-being and productivity. *Sustainability*, 14(9), 4894.

Li, C., Lin, S.H., and Chib, A. (2021). The state of wearable health technologies: a transdisciplinary literature review. *Mobile Media & Communication*, 9(2), 353–376.

McAuley, E., Duncan, T., and Tammen, V.V. (1989). Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: a confirmatory factor analysis. *Research Quarterly for Exercise and Sport*, 60(1), 48–58. doi:10.1080/02701367.1989.10607413.

Muñoz, S., Iglesias, C.Á., Mayora, O., and Osmani, V. (2022). Prediction of stress levels in the workplace using surrounding stress. *Information Processing & Management*, 59(6), 103064.

Park, S., Constantinides, M., Aiello, L.M., Quercia, D., and Van Gent, P. (2020). Wellbeat: A framework for tracking daily well-being using smartwatches. *IEEE Internet Computing*, 24(5), 10–17.

Reid, G.B. and Nygren, T.E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P.A. Hancock and N. Meshkati (eds.), *Human Mental Workload*, volume 52 of *Advances in Psychology*, 185–218. North-Holland. doi: https://doi.org/10.1016/S0166-4115(08)62387-0.

Setz, C., Arnrich, B., Schumm, J., La Marca, R., Tröster, G., and Ehlert, U. (2010). Discriminating stress from cognitive load using a wearable eda device. *IEEE Transactions on Information Technology in Biomedicine*, 14(2), 410–417. doi:10.1109/TITB.2009.2036164.

Thorvald, P., Lindblom, J., and Andreasson, R. (2019). On the development of a method for cognitive load assessment in manufacturing. *Robotics and Computer-Integrated Manufacturing*, 59, 252–266. doi: https://doi.org/10.1016/j.rcim.2019.04.012.

van Steenbergen, H., Band, G.P., and Hommel, B. (2015). Does conflict help or hurt cognitive control? initial evidence for an inverted u-shape relationship between perceived task difficulty and conflict adaptation. *Frontiers in Psychology*, 6, 974.

Yang, J., Liu, Y., and Morgan, P.L. (2024). Human-machine interaction towards industry 5.0: Human-centric smart manufacturing. *Digital Engineering*, 100013.