Modified pluralistic walkthrough for method evaluation in manufacturing

Peter Thorvald*a, Jessica Lindblom*a, Sebastian Schmitzb

*aUniversity of Skövde, Högskolevägen, Skövde, 54128, Sweden
bFIR at RWTH Aachen, Campus-Boulevard 55, Aachen, 52074, Germany

Abstract

This paper presents a mixed evaluation design including quantitative and qualitative data in order to validate the scope, validity, and reliability of a recently developed CLAM method and tool. A central part of the evaluation was the use of a modified pluralistic walkthrough that was created to better suit the industrial domain. The evaluation of CLAM was performed in a demonstration factory focusing on manufacturing of street scooters and electric pedal cars. The modified walkthrough was conducted on the shop floor, during on-going assembly, which stresses the representativeness of a real-world setting of manufacturing. The modified pluralistic walkthrough functioned accurately as an evaluation method of CLAM and the exhaustive comments and data from the user representatives were very beneficial. The quantitative results show convincing evidence of CLAM’s utility for assessing cognitive load while the qualitative results indicate that some modifications have to be made in order to enhance the actual tool in order to avoid misinterpretations resulting in misleading outcomes.

Keywords: CLAM; modified pluralistic method; cognitive workload; cognitive load method evaluation; manufacturing; assembly.

1. Introduction

The Cognitive Load Assessment for Manufacturing (CLAM) tool has been developed during the last couple of years and has previously been presented in several publications [1, 2]. CLAM is a proactive tool for evaluating cognitive load in manufacturing domains and it can be characterized as an inspection method created for people with

* Corresponding author. Tel.: +46500448543.
E-mail address: peter.thorvald@his.se
little to no experience in assessing cognitive load. The tool consists of a score sheet where the evaluator indicates quantitative assessments on 12 separate factors, relevant to the cognitive load of a manufacturing task [1, 2]. These factors are then weighted and summarized to present a general level of the cognitive load that is imposed on the worker. As a first step towards validating and adjusting factors and weights of CLAM, this paper presents method development and evaluation of CLAM.

2. The iterative method development process

Blandford and Green [3] present some underpinning principles for methodology development, including development and testing. They particularly stress that the overall methodology development process is time-consuming, requires large amount of work, and mental effort (and other resources). Consequently, the aim and intended benefits of the method should be clearly defined. The intended direct benefits with CLAM are that the method should be taken up and used in practice by non-experts, i.e. a so-called inspection method[4]. Hence, the aim is to make a difference in reducing cognitive load in the shop floor personnel in manufacturing. Another intended, indirect benefit is research related. From a research perspective, the development and testing of CLAM contributes to gain a better understanding of developing methodologies in general and condensing/applying situated, embodied, and distributed theories of cognition. Blandford and Green [3] discuss the need to consider some requirements when the intended method should be taken up and used in practice. They present some identified criteria, including validity, scope, reliability, usability, learnability, and feasibility of encapsulating a theory (for more details, see Blandford and Green [3].

Additionally, Nilsson [5] highlights several aspects to consider when assessing a method in general. First, the method could be more or less structured, by providing “how to” knowledge for its users/evaluators. She argues that a method, for instance, can be an aid to structure the process, to organize activities and results, or to make progress. It should be noted, however, that Nilsson [5]emphasizes that it is important to keep in mind that a method is not a simple recipe for success and should be used with good judgment. Consequently, a more formal and structured method can restrict new ideas and findings because the scope of interest is more rigid and defined. Secondly, the effectiveness of a method depends on its ability to fulfil its promises in terms of purpose and goal (i.e. related to validity and scope). Thirdly, the construct validity is related to the provenance of the underlying theoretical framework of the method (i.e. similar to feasibility of encapsulating a theory), and finally the reliability is related to the repeatability of the results when used by different users/evaluators [5].

To summarize, these above presented criteria and aspects serve as significant and general guiding questions in the different phases of the development process and particularly in the testing of CLAM.

Blandford and Green [3]argue that the “testing phase” of the methodology is dependent on the requirements of the method. Therefore, the first step is to decide on what criteria to use, define their content and the weightings between them. In order to evaluate a method properly, Blandford and Green [3]present what they refer to as the ‘PRET A Rapporter’ framework’, which consists of six questions that one might consider when testing a method, which include purpose of the evaluation, data-gathering and analysis techniques, and reporting of findings (for more details, see Blandford and Green [3].

To conclude, the testing of a method is of major important because how a method is used depends so heavily on who is using it. The knowledge and expertise of the users/analysts working with the method simply cannot be ignored [3].

2.1. Method and performance: Evaluation design

In order to test the developed method, a mixed evaluation design including quantitative and qualitative approaches was used. The reason for this mixed design was to obtain credibility and scientific rigor via different kinds of triangulation approaches [6]. We triangulated different methodological approaches in our evaluation design, i.e., method triangulation (e.g., field visits, inspection method), and data triangulation (e.g., participant observation, field notes, subjective measurements, quantitative production data, interviews, and video–recordings). By triangulating methods and data collection sources the accuracy (e.g., scope), credibility and scientific rigor (e.g.,
reliability and validity) of the findings increases and enables the developers to test for consistency between the different sources when evaluating the CLAM.

In order to properly evaluate CLAM, the chosen approach is aligned with the ‘PRET A Rapporter’ framework [3]. In this paper we focus on the field visits and the modified pluralistic walkthrough.

The purpose of the evaluation was to investigate the overall credibility of CLAM according to some selected criteria. First, it is necessary, but not sufficient, to confirm the scope, validity, and reliability of the intervals and metrics for the four scoring levels in CLAM (high, moderate, low, and very low cognitive load), and the weighting of factors in the tool for assessing cognitive load, before continuing with more focused user research. Hence, the evaluation consists of two different levels of analysis, the overall credibility of the CLAM framework as such (metrics of the intervals and scoring levels, and the weighting of factors), in combination with the overall usability and learnability of the CLAM tool and its handbook. Taken together, the planned mixed evaluation design is suggested to cover the above mentioned criteria.

Field visits: learning from observations

Field visits move research into, for instance, offices, homes, and factories, in order to understand how and why people do what they do [7]. This way of conducting user research is usually performed via different kinds of observations (data collecting techniques), ranging from shadowing selected users from a distance (non-participatory) to the researchers being actively participating in the users’ activities [6]. By observing potential users in their context, the researchers gain first-hand access to the users’ experience and seeing it with their own eyes. In so doing, the researchers will be provided with information about the users’ environment, the problems they encounter and how they cope with them. Field visits uncover what people actually do, and not what they say that they do. By experiencing their reality alongside them, the researchers gain a better and more thorough understanding of their circumstances, work practices, and insights into their situation [7].

According to Blandford and Green [3], it is of major importance when testing a method that is to be taken up and used in practice, to understand the requirements of that particular practice. Therefore, we used field visits for observing a selected number of workstations in a manufacturing environment in order to evaluate CLAM, especially the workstation module. In addition to observations, field visits have the advantage to perform informal interviews in the moment of actions, which provides the researcher with complementary and accurate data. These unstructured interviews [6] provide the researcher the possibility to gather further data, depending on the unfolding situation, that appears to be relevant. In this field visit the interviews were mainly unstructured and informal, but additional interviews were also performed with the purpose to decrease the risk of misunderstandings and pursue a deeper understanding about relevant aspects of the factors in CLAM.

Pluralistic walkthrough

In order to validate the scope, validity, and reliability of CLAM, a modified pluralistic walkthrough [8] was performed. Generally speaking, a pluralistic walkthrough is characterised as an inspection method (focusing on usability) where a group of stakeholders, with varying competence (users, management, developers), are gathered together to review a product design. This way of working, increased efficiency and utility, compared to traditional inspection methods, enabling a greater number of usability problems to be found [8].

The original pluralistic walkthrough has the following five defining characteristics: Firstly, a pluralistic walkthrough has the inclusion of different competences in the same walkthrough. Second, the scenario is presented in the same order in which they would appear in the final product design via some low-fi prototypes/hard copy panels (screens). The evaluators confront the prototypes similarly as they would during a positive conduct of the specific task, as currently designed. Third, the participants should assume the role of the intended user, stressing ‘inspector empathy’. In other words, the developers and usability/human factors professionals should try to put themselves in the shoes of the users when conducting the walkthrough. Fourth, each participant writes down, in as much detail as possible, what they would do for each step in the presented scenario (presented via the low-fi prototypes), before any common discussion. Fifth, the discussion always starts with the user representative, and when his/her comments are exhausted the human factors/usability experts and developers are allowed to present their opinions [8].
The pluralistic walkthrough is a group activity [8]. First, all the participants are provided with the instructions and the ground rules, and the task descriptions and prototype package. Next, a product expert (usually the designer) offers a brief overview of the key product concepts. This overview is intended to be a stand-in for any written presentation that would accompany the final product design. After that, the walkthrough begins with the first prototype, the participants individually write down their responses, and the walkthrough administrator announces the “right/intended” answer. This is followed by the user representative starting to verbalize and discuss potential problems while the product developer remains quiet and the human factors/usability professionals facilitate the discussion. By so doing, the represented users’ identified problems are not influenced by the product developer. It should be noted, however, that the representative user might be unwilling to make critical comments if the product developer is present and/or has a negative attitude. Therefore, it is of importance that the developer assumes a positive attitude of welcoming comments that are all intended to maximize the quality of the final product. Once the discussion is in progress, the product developer is allowed to join in, often explaining the intention behind the current design. The developers should be prepared in advance that the comments from the other participants can be experienced as major and iniquitous criticism. The presence of the different competences in a pluralistic walkthrough creates a potential that often lead to creative and collaborative design solutions. After each task, the participants sometimes are provided to fill in a questionnaire that concerns the usability of the prototype they just have inspected [8].

Pros and cons of pluralistic walkthroughs

Pluralistic walkthroughs provide performance and satisfaction data to developers from both users and human factors/user specialists at an early stage in the development process before the hi-fi prototype is available. The method is rapid and reduces time by cutting down the test-redesign-retest cycle by generating immediate feedback and discussion on design problems with developers, human factors/user professionals, and users present simultaneously. Sometimes it also generates design solutions on the fly. Regarding cost effectiveness, it gathers data from multiple participants in one session. The more persons involved looking for problems, having a diverse range of competences and perspectives, the higher the probability of finding issues to consider. The group atmosphere encourages collaborative, constructive comments from users, developers, and other member(s) of the product development team. Summing up, pluralistic walkthroughs generate valuable quantitative and qualitative data on users’ actions by written responses, and the product developer gain access to common usability issues or concerns about the design.

Pluralistic walkthroughs also have some cons [8]. First, the interdisciplinary team of participants has to be present at the same time; otherwise the whole underlying concept is ruined. The pace of the walkthrough has to progress as slowly as the slowest participant in the team. The group exercise nature prevents the team from starting the discussion before everybody has written down their comments. This may decrease the intended flow of interaction with the product, and also faster participants can experience the walkthrough session as lengthy. The pre-designed lo-fi prototypes of the product design only provide a fixed and predetermined way of interacting with the product design. This precludes the participants from exploring and investigating alternative paths of the same task, which can lead to additional learning insights about the product design. It might be the case that there are several correct paths of interaction available and participants that did not pick the one that the walkthrough administrator selected must “reset” and continue along the selected path. Additionally, it can be a hard experience for the product developer(s) to receive criticism face to face, and they may not feel comfortable when severe usability issues are presented [8]. Finally, it is a time consuming task to put together and analyze the written comments after the actual walkthrough, and find a conclusive design solution, based on analysis of the written reports.

3. Evaluation of CLAM

3.1. The Aachen demo plant

The major part of the evaluation study was conducted at the Aachen demo plant (see http://www.fir.rwth-aachen.de/en/campus/demonstrationsfabrik), where manufacturing of the StreetScooter and the MAXeKart
took place. At the point of time for the visit, only the manufacturing of the MAXeKart was up and running, and therefore the focus was mainly on the manufacturing of the MAXeKarts. The StreetScooter was still in the late prototype stages of production and was therefore only studied in a limited way. At this facility, three major workstations of varying difficulty and content were the selected units of analysis.

3.2. Field visits

During the initial parts of the evaluation, most time was spent on a factory visit. It consisted of an overall guided tour of the demo plant, presentation of the different workstations, the current production and production rate. In addition, the evaluators acquainted themselves with the assembly workers (all of them were men), who had varying experience of manufacturing. Thereafter, the evaluators observed the workers in real work situations at the different work-stations, and asked some questions in order to understand the current work processes and the assembly tasks. Most time were spent on the manufacturing of the MAXeKarts. Besides getting to know the process more thoroughly, the evaluators had to investigate and analyze the different work-stations in order to decide on which parts that should be used as testbed for the modified pluralistic walkthrough.

The whole assembly line of the MAXeKarts consists of two major phases, pre-assembly and end assembly. The pre-assembly consists of three workstations. At these workstations, pre-assembly of different parts to be assembled in the end assembly phase are put together. Given the current low production rate, the three pre-assembly workstations were evaluated as an overall “pre-assembly line” since the assembly workers follow the product via the different pre-assembly stations. In total, three assembly workers are allocated at the MAXeKarts assembly line, and at the time of evaluation, it took about 1 h, 20 minutes to assemble one car, however as production has been ramped up, production of one MAXeKart is expected to take in total around 1 hour, circa 20 minutes per end assembly station.

The major tasks conducted at the pre-assembly are subassemblies that are later mounted to the car in the end assembly workstations. These assemblies include hubs, front and rear axels, steering linkages etc. As in pre-assembly, the “end” assembly phase consists of three workstations (Figure 1).

![Figure 1. View of the threeworkstations used in sub assembly.](image)

The major tasks in end assembly consist of the mounting of subassemblies to the main chassis of the MAXeKart, assembly of breaks, wheels, electrical motor etc. see Figure 2,

![Figure 2. Detailed views of some of the assembly tasks performed in end assembly.](image)
As in pre-assembly, given the current low production rate, the three end assembly work-stations were evaluated as an overall “pre-assembly line” since the assembly workers follow the product via the three different end assembly stations.

Moreover, some parts of the manufacturing of the StreetScooter will be included in the evaluation of CLAM. At the moment, the StreetScooter is only in prototype production since it is still in the development stages of industrialization, and the chosen workstation is welding, although this task cannot be considered as a fully traditional assembly task. However, we wanted to incorporate it in the evaluation given the fact that it included some parts of assembly, e.g., parts identification, physical constraints and use of instructions, which are central for the CLAM.

3.3. Modified pluralistic walkthrough

The evaluation of CLAM by using the modified pluralistic walkthrough (designed by the authors) was performed on the second day at the Aachen visit. Generally speaking, it follows the major procedure for a pluralistic walkthrough but has been adapted and redesigned to a manufacturing context. Instead of having only one representative for the end user, two end user representatives were involved in the modified walkthrough. The level of expertise for the user representatives was high, representative 1 was the plant manager of the Aachen demo plant, he holds a PhD in manufacturing, doing research in the field of production engineering with a focus on sociotechnical aspects, and he also works as a European consultant in manufacturing. He was considered an expert in the field. Representative 2 has an MSc in industrial engineering, and is now working as a research assistant (PhD candidate). He was considered knowledgeable in the field. Hence, both user representatives matched perfectly with the intended end user group for CLAM[1, 2]. In line with this remark, the walkthrough was conducted on the shop floor, during ongoing assembly, which stresses the representativeness of a real-world setting of manufacturing and thus the validity, reliability and scope of the evaluation of CLAM.

Both user representatives were introduced to the modified pluralistic walkthrough at the beginning of the visit and the instructions and accompanying handbook for CLAM were given to them beforehand. Before the evaluation started, the user representatives were asked if they had any questions or if something in the handed out procedure and/or material was unclear. As they had no objections, the walkthrough started. All sessions were video-recorded and field notes were taken. In total, two evaluators (both human factors specialists and representatives of the CLAM development team, one of whom also had the role of the coordinator of the walkthrough) and the two user representatives described above participated. As declared in the instructions, the user representative started to address his/her opinions, and in this case it was user representative 1 who started, followed by user representative 2. After that, the walkthrough coordinator and finally the representative for the CLAM development team spoke. Before the discussion began, each participant (in total four participants) had rated the actual task in the provided material (printouts of CLAM and the handbook) for each factor in CLAM. The walkthrough was performed at pre-assembly (MAXeKart), end assembly (MAXeKart), and welding (StreetScooter). In sum, three iterations were conducted with four participants, meaning that 12 assessment sessions of CLAM were performed in total. After the last iteration, welding, a debriefing session was performed in an office environment at the demo plant.

The first iteration (pre-assembly) lasted about 1.5 hours, since there was a need to straighten things out in the description of the factors as well as in the rating levels of CLAM. The major issues were clarifications of some descriptions and concepts used, and when these concerns were sorted out, highly meaningful comments and ideas were discussed within the walkthrough team. The subsequent iterations (end assembly and welding) went much faster (approximately 15-20 minutes each) since the clarifications were already made. The debriefing session lasted approximately 45 minutes. Generally speaking, the modified pluralistic walkthrough functioned accurately as an evaluation method of CLAM. Furthermore, the initial results indicate that CLAM possesses the sought after validity, reliability, and scope for assessing cognitive load in shop floor personnel in manufacturing and the need for such an inspection method that could be used by production engineers and/or work-station designers was in great demand, i.e. CLAMs intended utility was confirmed.
4. Evaluation results

The major aim of the factory visit was to get to know the manufacturing facility better and to select the sections that should be used in the evaluation study. Therefore, the focus in this chapter is on the evaluation results of the modified pluralistic walkthrough. Table 1 shows the results (in numerical values) for the three iterations of the modified pluralistic walkthrough.

Table 1. Evaluation result of CLAM for the three chosen tasks. Two users (U1 & U2) and two evaluators (E1 & E2) were present. Evaluator 1 was the evaluation coordinator.

| CLAM Evaluation factors | Pre-Assembly | End Assembly | Welding |
|-------------------------|--------------|--------------|---------|
|                         | U1   | U2   | E1   | E2   | U1   | U2   | E1   | E2   | U1   | U2   | E1   | E2   |
| Saturation              | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    |
| Variant Flora           | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 4    | 4    | 4    | 4    |
| Batching of variants    | N/A  | N/A  | N/A  | N/A  | N/A  | N/A  | N/A  | N/A  | No   | No   | No   | No   |
| Level of difficulty     | 3    | 2    | 2    | 2    | 4    | 4    | 4    | 4    | 4    | 4    | 4    | 4    |
| Difficulty of tool use  | 3    | 3    | 3    | 3    | 2    | 2    | 2    | 2    | 3    | 3    | 3    | 3    |
| Level of attention required | 1    | 1    | 1    | 2    | 4    | 4    | 3    | 3    | 3    | 3    | 3    | 3    |
| Number of tools available | 3    | 3    | 3    | 3    | 3    | 3    | 3    | 3    | 2    | 2    | 2    | 2    |
| Mapping of workstation  | 2    | 2    | 2    | 2    | 2    | 2    | 2    | 2    | 3    | 3    | 3    | 3    |
| Parts identification    | 2    | 2    | 3    | 2    | 2    | 2    | 3    | 2    | 3    | 3    | 3    | 3    |
| Quality of instruction  | 2    | 2    | 3    | 2    | 2    | 2    | 3    | 3    | 4    | 4    | 4    | 4    |
| Information cost        | 2    | 2    | 2    | 2    | 3    | 3    | 3    | 3    | 4    | 4    | 4    | 4    |
| Poke-a-Yoke and constraints | 4    | 4    | 4    | 4    | 4    | 4    | 4    | 4    | 2    | 2    | 1    | 2    |

As depicted in the pre-assembly results, the interrater reliability is shifting between very high and excellent, so the obtained result indicates that the CLAM tool provides reliability, validity and accurate scope. Given the fact that the production rate was low, it was not estimated that the assessment of cognitive load should be high in general, and the results in Table 1 are aligned with this argument. However, there is some room for improvements in CLAM that we address in the end of this section.

As depicted in the end assembly results, the interrater reliability is here also shifting between very high and excellent, so the obtained result indicates that the CLAM tool provides reliability, validity and accurate scope. Just as with pre-assembly, a low assessment of cognitive load was expected, and the results again are along the lines of this expectation. The obtained assessment of cognitive load is higher for end assembly than pre-assembly, which is in line with the results from the observations and informal interviews at the field visit as well as the comments addressed during the discussion in the walkthrough team.

Finally, the result from the walkthrough concerning welding is depicted to the right in Table 1, and given the fact that this workstation is not considered to be a representative assembly task, the results indicate that CLAM has the potential to be used in similar tasks, a wider general use of CLAM may be possible. As depicted in the table, the interrater reliability is almost a perfect match, unless for the factor “Constraints and Poke-a-Yoke”, so the obtained result also indicates that the CLAM tool provides reliability, validity and accurate scope.

In summary, the quantitative results from the modified pluralistic walkthroughs show convincing evidence of CLAM’s validity, reliability and scope for assessing cognitive load. However, the qualitative data points out that some modifications have to be performed, in order to enhance the tool in order to avoid misinterpretations and subsequently resulting in misleading outcomes.

The results of the evaluation showed a need for a few changes in factors and evaluation levels as well as the resolution of the levels in general across the tool. Further, several minor changes to specific factors have been made...
and one of the factors, batching of variants has been removed. Several factors and assessment bases have been clarified and the level N/A (not applicable) has been added to all the factors.

5. Discussion and conclusions

The overall impression and conclusion of the evaluation of CLAMis that the tool is valid in its construction and that there is a true need for the tool on the market. The results show several opportunities for improvements to both the tool and the handbook. Perhaps one of the more major issues found with the current version of the tool, and the manor in which it was evaluated, is the fact that the tool and the handbook are two separate units. In the evaluation, the handbook and the assessment forms were all printed on paper, a fact that made it slightly complicated to move back and forth between the two. So, when questions arose in the assessment that would require the consulting of the handbook, participants were more prone to discuss their translation of the short assessment description found in the tool. This is potentially a large flaw of the tool, but one that can be easily remedied. First of all, using an electronic version of the tool would make these situations more rare. Performing the assessment on a computer or a PDA of some sort would enable quicker access to the handbook and introduce the opportunity for hyperlinks that connect the tool with direct links to the relevant parts of the handbook. As the tool will be released online, this will be realized. Secondly, the assessment forms of the tool will be extended to include more detailed descriptions of the factors within the tool so as to reduce the need to consult the handbook for information.

In summary, the evaluation was very beneficial to the development of clam and the modified pluralistic walkthrough served our purposes very well. It was very advantageous to perform the evaluation on the shop floor rather than in an “offline” office environment and the gathering of both qualitative and quantitative data enabled deeper analysis of the gathered data, leading to valuable improvements of the next version of CLAM. Thus, we now feel certain that CLAM covers the intended scope and offers good credibility, validity and reliability.

Acknowledgements

The authors wish to thank the users, managers and other representatives of the FIR at RWTH Aachen demonstration plant as well as the Sense&React Consortium for their input and support. The Sense&React project is funded by the EU grant FP7-314350.

References

[1] J. Lindblom, and P. Thorvald, Towards a framework for reducing cognitive load in manufacturing personnel. Advances in Cognitive Engineering and Neuroergonomics, 2014. 11: 233-244.
[2] P. Thorvald, and J. Lindblom, Initial Development of a Cognitive Load Assessment Tool. Advances in Cognitive Engineering and Neuroergonomics, 2014. 11: 223-232.
[3] A. Blandford, and T.R. Green, Methodological development, in Research methods for human-computer interaction, C. P and C.A. L, (Eds.). 2008, Cambridge university press: New York. pp. 158-174.
[4] J. Nielsen, and R.L. Mack, Usability inspection methods, 1994, New York: Wiley & Sons.
[5] M. Nilsson, Capturing semi-automated decision making: the methodology of CASADEMA. 2010.
[6] Patton, M.Q., Qualitative research. 3rd ed. 2002, London: Sage Publications.
[7] E.M. Goodman, Kuniavsky, and A. Moed, Observing the User Experience: A Practitioner's Guide to User Research. Professional Communication, IEEE Transactions on, 2013. 56(3): 260-261.
[8] R. G. Bias, The pluralistic usability walkthrough: coordinated empathies, in Usability inspection methods, J. Nielsen and R. Mack, (Eds.). 1994, John Wiley & Sons, Inc.: New York. pp. 63-76.