Evaluating worker-centered smart interventions on the shop floor

Marlene Schafler[1], Lea Hannola[2], Francisco José Lacueva-Pérez[3], Matjaž Milfelner[4], Melanie Steinhüser[5], Miguel Angel Gracia[3]

Smart Production and Human-Centered Solutions, Virtual Vehicle Research Center, Austria[1]
Lappeenranta University of Technology, Lappeenranta, Finland[2]
Department of ICT-Multimedia, ITAINNOVA, Spain[3]
Department for the research and development, EMO - Orodjarna d.o.o., Slovenia[4]
Department of Informatics, University of Zurich, Zürich[5]

Abstract

This paper presents the evaluation strategy and the first results we obtained when we used the FACTS4WORKERS evaluation framework. The purpose of the framework is to prove whether the project interventions achieve the expected results, which are: improving workers' job satisfaction, increasing innovation and problem solving skills as well as enhancing productivity. Because of the diversity of the industrial partners and of the workplaces where the interventions are going to be implemented, the different languages, legal and cultural environments the framework was conceived as general as possible to be adapted to any particular case. We present here one example for using the framework, the first results of these measurements and the feedback the evaluation provides both for supporting the decisions about the interventions and about the framework itself.

1 Introduction

This study is a part of the on-going FACTS4WORKERS (F4W) project, which develops worker-centered solutions that support the inclusion of increasing elements of knowledge work on the shop floor of smart factories. Originally, the Smart Factory initiative (Zuehlke, 2010) came from industrial and academic partners for creating, demonstrating and researching a test environment for factory technologies of the future. The main goal of the evaluation framework is to demonstrate the achievement of the three project objectives: 1) the increase of worker satisfaction, 2) the improvement of innovation and problem-solving skills, and 3) the increase of workers' productivity. The evaluation framework was created for demonstrating and evaluating the benefits introduced in factories when worker-centered digital interventions are deployed. It has two main purposes: 1) to define metrics and

Veröffentlicht durch die Gesellschaft für Informatik e. V. 2017 in
M. Burghardt, R. Wimmer, C. Wolff, C. Womser-Hacker (Hrsg.):
Mensch und Computer 2017 – Workshopband, 10.–13. September 2017, Regensburg.
 https://doi.org/10.18420/muc2017-ws04-0395

methods for measuring the impact of the smart factory on workers and organizations; 2) to describe proper methodologies for an iterative evaluation.

Although, there exist several models for measuring job satisfaction (e.g. Spector, 1985) as well as technology success and acceptance (e.g. Davis et al., 1989, DeLone & McLean, 2003), there is no framework reported especially for evaluating smart interventions on the shop floor. Thus, this paper makes a theoretical contribution by introducing an evaluation framework for production environments and practical contributions by highlighting the first results of the digital interventions in a smart factory. In this paper, the term "smart interventions" comprises the implementation of digital tools (e.g. tablets or smart phones) and software applications (F4W solution), which are provided to the workers in order to enhance their job satisfaction and personal skills, and also the changes in their work practices that these digital interventions will cause.

Once the first release of the framework was created, it has been tested and evolved on the developments of the specific use cases defined by the industrial partners. The purpose of the evaluation is to directly give feedback for the development with required improvements on deployed capabilities or indirectly by identifying new needs or scenarios. The objective of this paper is to present one example for using the evaluation framework, the first prototype evaluation results of the measurements in EMO - Orodjarna and the feedback that the evaluation provides both for decision support about the interventions and about the framework itself.

2 Evaluation Strategy

2.1 Evaluation Framework Overview

The F4W evaluation framework (F4WEF) aims at providing a set of measurements which prove whether interventions on shop-floor achieve the expected impacts. The framework was introduced by Lacueva et al. (2016) and Hannola et al. (2017) and it is twofold. The first component deals with the validation of the information system (IS) solutions in terms of measuring to what degree the artifacts meet workers' expectations concerning the (system, information and/or interaction) quality. Which aspects of the quality can be measured, and conclusions can be drawn depends on the time and on the maturity of the artifacts: a mock-up/demonstrator shows the proof of concept, functional prototypes provide the proof of value, whereas pilots are supposed to demonstrate the proof of use (Gable et al, 2008). Figure 1 shows that time and maturity are main aspects of the evaluation (Venable et al., 2012), moving from the validation of the artifacts in the beginning of the project to the impact analysis (IA) when mature solutions are provided. IA is the second component of our evaluation. It assesses the impact of interventions on individuals, which is measured by the dimensions autonomy, competence, variety, relatedness and protection; and the impact on organizations which is measured by efficiency and quality (Heinrich et al, 2015).

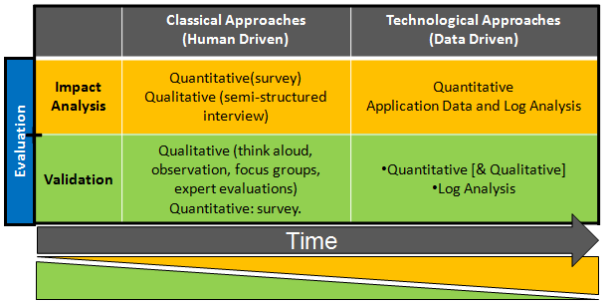

Figure 1 Formulation, tools and methods of F4WEF

Also related to time and maturity is the kind of tools the F4WEF proposes for evaluating an intervention. As shown in Figure 1, we divide these tools in two sets. The first one is called *classical approaches* (CA). CAs are human driven and set the focus directly on the workers. Both quantitative (surveys) as well as qualitative data (interviews, observation, etc.) are gathered to validate the artifacts and to assess their impact. The second set of tools tries to take advantage of the deployed IS solutions. These tools are data driven, which is why we call them technological approaches (TA). System logs and application data are used for measuring the quality of a system, its performance and supported tasks, or the quality of the production.

While TAs reduce the intrusion level in workers' daily tasks, they can compromise the anonymity. Ethical and legal issues as well as idiomatic problems lead us to decide providing a broad set of evaluation tools. Within the course of these decisions, the F4WEF was defined as a flexible framework that can be adapted to any industrial scenario. It is designed to be practical in its application as well as to support the scientific recognition of our work. The main example of this balance is the possibility to use both methods, surveys and guided interviews, for assessing the impact of the interventions. While survey results can be quantified easily, they are scalable and they can be translated to different languages, guided interviews do not scale so well. However, interviews provide other valuable insights of the workers' feelings with the intervention. Furthermore, they provide information about the understanding of the questions included in the questionnaires which is important for the validation process of the F4WEF itself.

Evaluation is a process which is going to be executed through all the project life cycle. Validations of artifacts will be linked to particular interventions, while the impact assessment process will be "*independently*" executed at given points of time before and after each intervention. By proceeding in this way, F4WEF contributes to two of its objectives: firstly, give feedback to developers in order to make improvements at the deployed version; secondly, assessing the change in impact dimensions (ID) due to interventions. Both objectives also give rise to the main goal of the framework, to support decisions about next steps of the project: if the impact results as expected, a decision about continuing/stopping the project could be based on the distance to the desired impact and, in the case of deciding to continue, the requirements for the next iteration would be defined; if the impact does not result as expected, the feedback from workers would contain the cause.

2.2 Evaluation Strategy

In the previous paragraphs, we showed that the evaluation framework is composed of the validation of the artifacts and of the analysis of the interventions' impact on workers and organizations. We also showed that time and artifact maturity influence the evaluation results as well as the conclusions which could be obtained from the results.

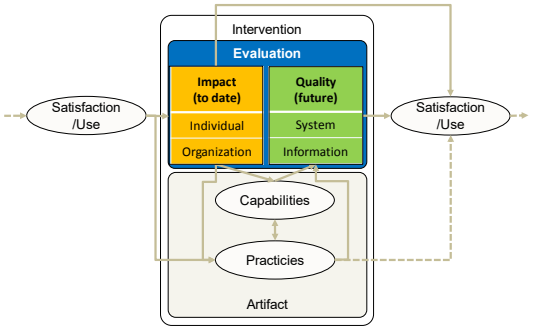

Figure 2 Relation between previous and current intervention impacts

Figure 2 is adapted from the IS impact measurement model introduced by Gable et al. (2008). It shows how the impact of previous interventions influences current interventions, how the quality of deployed artifacts will influence the impact as well as the relation between the quality of the artifacts, the impact and the capabilities and practices provided by the artifacts. It also supports the creation of a "general schedule" of the evaluations: they are going to be performed linked to deployments of artifacts, IA is going to be performed before and after the intervention, while validation will be performed during interventions.

For measuring the impact of a given intervention the F4WEF proposes to perform measurements whenever an intervention would take place. On early development stages these measurements will provide us with feedback about the framework itself (i.e. if the questions are correctly understood or not) and they will also support a baseline for comparing the state before starting the project and after the last intervention. As the artifacts get more mature, the change in the workers' satisfaction, their innovation and problem solving skills or their productivity due to the intervention can be obtained by comparing the measurements before the intervention took place and after some time. However, we are not shielded from biases due to external events which are not related to the intervention. To solve this issue, the framework proposes to use a control group (CG). By constituting CGs, the particular legal and ethical issues to be considered are related to the factory where the evaluation is going to be executed. They have an influence on the selection of the tools and the methods which will be used for the evaluation. Moreover, the selection of these tools and methods will also be influenced by the concrete use case, the environmental factors of the workplaces where the IS artifact is going to be used or the maturity of the artifacts.

In order to adapt the provided set of tools to a given deployment scenario the F4WEF divides the evaluation process in 3 phases or steps. Preparation as the first phase, considers the factors previously introduced for selecting the tools and methods that best fit to perform an

evaluation in a given scenario. In coordination with the development team, it also created a scheduling of the interventions and evaluations to be executed. In the second phase, the evaluation is planned and conducted. Before each intervention, the selection of tools, both for validating and analyzing the impact, is reconsidered and adapted to the concrete artifact to be evaluated. The concrete objectives are fixed and a guide to perform the evaluation is created. Finally the data obtained from the evaluation are processed and used during the conclusion phase in order to determine what to do next.

The evaluation procedure is documented in a guideline which will be used by the evaluation team and which describes the following issues:

- An introduction to the project in general and evaluation goals in particular as well as an explanation of the rights of the workers;
- The worker consent document in English and also in the workers' language;
- The questionnaire/interview guide;
- The guide to perform the validation customized for the artifact to be evaluated;
- The recommended execution plan.

The validation results can be quickly transferred to the development teams using the projects tracking tools. After the second assessment of the impact, conclusions about the expected impact can be made. The validation results support decisions about the next steps in the project. In the next section we exemplarily illustrate evaluation results as obtained at EMO Orodjarna.

3 Initial evaluation results of EMO - Orodjarna

EMO - Orodjarna manufactures and assembles a product range of tools for transforming sheet metals which are exported mainly to the automotive industry. The SME employs specialists that are working with modern technology and software on the shop floor. In order to enhance and develop their established labor practices in the future, the EU project FACTS4WORKERS should support and in fact prove worker-centered improvements as well as organizational improvements. During the requirements management phase the company defined three problem scenarios within two worker centered contexts-of-use (Denner et al., 2015). The first context-of-use deals with missing awareness during the assembly process and unclear quality control responsibilities. The second context-of-use considers the challenge of dealing with too many unpredicted problem issues that affect for the most part the maintenance team. After defining activity scenarios for the selected problems, mock-ups were designed and evaluated by four workers in the first run (proof of concept). The feedback of this initial intervention incorporated specific requirements for developing the software prototype proposed for the shop floor. The prototype was installed on a server of the IT infrastructure of the company whereas the developed software building blocks (software applications) were delivered and provided via a novel OS virtualization (docker container).

The first prototype intervention on site took place in June 2017. 15 employees were involved in the evaluation process from which 9 users tested the functionalities of the prototype. Six employees were dedicated as the CG. Before testing the prototype, all participants had to fill out the questionnaire for the IA. This procedure contributes to obtaining the state of the art regarding work satisfaction, problem solving and innovation skills and productivity in general. In figure 3, the IA results are demonstrated. By the F4W group it can be shown that competence, protection and relatedness followed by autonomy are supposed to be the main affected dimensions. In contrast, the CG which is not involved in the project, stated that the workers' competence followed by autonomy will be mainly influenced by the software.

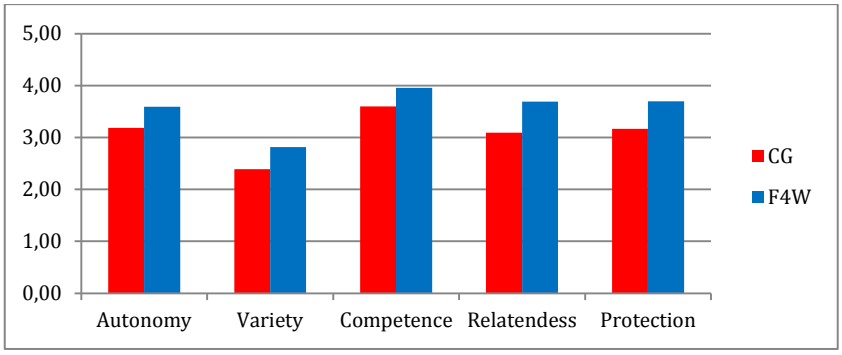

Figure 3 impact analysis chart one

The second IA chart in figure 4 shows that the system has more effects on quality than on efficiency whereas the CG has similar opinions such as the F4W group. Regarding these results, it can be assumed that there will be a rise in quality issues but not so much in efficiency by using the software.

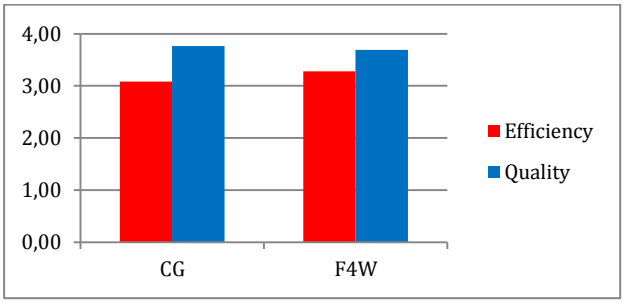

Figure 4 impact analysis chart two

The expectations of the mock-up intervention differ to theses outcomes. Four workers were asked and they had a higher expectation regarding protection and relatedness followed by efficiency and autonomy. The implemented prototype supports the workers in their working practices that are described in the first context-of-use. During the preparation phase, we defined 'evaluation stories' the test users had to perform on the smart device. The validation

(proof of use) was executed with the think aloud method without interrupting the test user during the exercise. We also decided to walk through the artifact in a collaborative way with the test users and recorded the screen in order to follow up the movements and clicks of the user. Then the test users were required to fill out the UMUX-lite (figure 5) questionnaire.

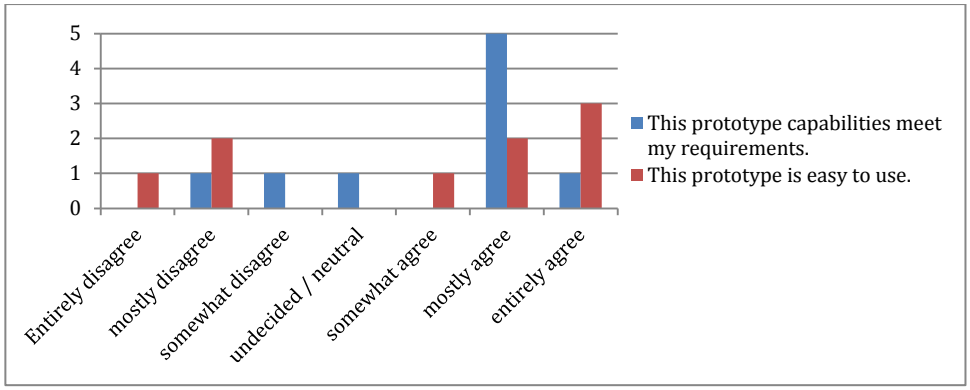

Figure 5 UMUX-lite results

After the testing phase, the employees were interviewed in group sessions. They agreed that the prototype meets their requirements and demands. In general, it is easy to use and very helpful in their daily tasks on the shop floor. Especially the functionality of error handling which supports a smooth and quick work processing is a big advantage for the workers. Also, the traceability of parts and the transparent documentation of responsibilities of positions improve the planning of their own work processes and work practices. We recognize that the answers from the interviews differ from the answers in the UMUX-lite questionnaire. Three test users did not agree with the simplicity of the application. Therefore, we have to reorganize the interview sessions in order to go more into detail on an individual level.

The main changes in the production and assembly process because of interventions are: 1) all relevant information are in one place, 2) better traceability and review of the work, and 3) supported decision making and faster solving of errors during the assembly process.

4 Lessons Learned

The evaluation aims at reviewing the progress of the implementations, in order to understand what is working well and what is working less well. Continuous testing and tracking is essential to improve the implementation strategies and finally to achieve better results. To this day, after the first evaluation rounds have been carried out on-site with the workers, we are able to reflect our evaluation strategy and recap the lessons we have learned so far. One of the main issues that needs to be mentioned at this point is flexibility: Each industrial partner operates in a different setting, shaped by its particular organizational structure and culture, specific needs, different languages, and legal conditions. This circumstance calls for

evaluation tools being highly adaptable to the specific situations. A question that is of high relevance in one use case, for example, might be inappropriate in another. Hence, the framework needs to allow for sufficient freedom to ask use case specific questions, and at the same time it has to include universal measures that allow a project wide comparison of results.

Furthermore, the internationality as reflected by the project results in challenges concerning different languages. The evaluation strategy in general, as well as the measurement instruments in particular, have been developed in the project language English. However, not only translating the questions into the workers' native languages (German, Slovenian and Spanish) but also translating their answers back into English turned out to be a challenging task. We have to ensure that no information gets lost or misinterpreted on either way. Closely involving the facilitators, the local contact persons, turned out to be essential in this context to gather valid evaluation results and, in consequence, they have to be trained to understand the evaluation, to be able to explain it to the workers and also to execute correctly the evaluation tasks.

Another main challenge we are facing in the project is the small number of pilot users of the systems. On the one hand, to be able to obtain honest answers and significant results, confidentiality has to be ensured what is particularly challenging with a small sample. To address this, it is essential to make clear that we do not aim at judging the workers or their personal performance but just the impacts of our project's activities. Furthermore, anonymity is preserved where possible. Personal evaluation activities are conducted by project members who are not directly associated with the company, but instead by company-external researchers, for example the use case leaders who keep the raw data and just share the analyses. On the other hand, the small sample sizes lead to difficulties when it comes to interpret the evaluation results. We are aware that we will not obtain representative results from a survey that is filled out by less than ten people. However, by including qualitative elements to the evaluation, we get around that issue and are able to enrich the results with statements that are interpretable and allow for drawing valid conclusions.

Finally, we have to admit that developing and conducting the evaluation is an on-going process, underlying constant feedback loops before obtaining a final set of validated measurement items. For example, during the first evaluations, it turned out that not all asked questions were understood completely or in the same way in the different companies. Therefore, adaptions have to be possible and transparent to every party. Additionally, progresses in the particular implementation situations need to be taken into account. E.g., whereas information about the system functioning are highly relevant at early project stages, changes in work practices as well as the achievement of expected impacts is of particular interest after a system has been used for a while.

5 Conclusion and Outlook

As demonstrated in this article, the evaluation framework that has been developed is appropriate to measure the impacts of the F4W project interventions. Analyses of the first rounds of evaluation show, that our strategy takes into account the peculiarities of the project and is able to produce valuable results. We have also highlighted some of the challenges we are facing and how to overcome these. It is clear that the evaluation strategy is not fix. It has to remain flexible in order to be able to quickly react to changing circumstances. However, in the end, we expect to obtain comparable evaluation results that allow for drawing conclusions about how to design worker-centered digital workplaces that lead to high job satisfaction, improved problem-solving and innovation skills and also to increased productivity.

6 Acknowledgements

This study has received funding from FACTS4WORKERS – Worker-centric Workplaces for Smart Factories - the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 636778.

7 Literature

Davis, F.D., Bagozzi, R.P., Warshaw, P.R. (1989). *User Acceptance of Computer Technology: A Comparison of Two Theoretical Models*. Management Science, 35 (8), 982-1002.

DeLone, W.H., McLean, R. (2003). *The DeLone and McLean Model of Information Systems Success: A Ten-Year Update*, Journal of Management Information Systems, 19 (4), 9-30.

Denner et al. (2015). *First version of requirements of workers and organisations.* Project Report - FACTS4WORKERS: Worker-Centric Workplaces in Smart Factories.

Gable G.G., Sedera D., Chan T. (2008). *Re-conceptualizing information system success: the IS-Impact Measurement Model*, Journal of the Association for Information Systems, 9(7), 377-408.

Hannola, L., Lacueva-Pérez, F.J. (2017). *An evaluation framework for worker-centric solutions in production environments*. 24th International Conference on Production Research, Poznan, Poland.

Heinrich, P.; Richter, A. (2015). *Captured and structured practices of workers and contexts of organizations*. Project Report -FACTS4WORKERS: Worker -Centric Workplaces in Smart Factories.

Lacueva, F.J.; Mayo, S; Heinrich, P.; Moertl, P.; Hannola, L.. (2016): “Evaluation Framework”. Project Report - FACTS4WORKERS: Worker-Centric Workplaces in Smart Factories

Spector, P. E., 1985, Measurement of Human Service Staff Satisfaction: Development of the JS Survey. American Journal of Community Psychology, 13 (6), 693-713.

Venable, J. Pries-Heje, J., Baskerville, R., 2012, A comprehensive framework for evaluation in design science research, International Conference on Design Science Research in Information Systems, Springer Berlin Heidelberg, 423-438.

Zuehlke, D., 2010, SmartFactory - Towards a factory-of-things, Annual Reviews in Control, 34, 129-138.