WORKINGAGE

Smart Working Environments for All Ages

D9.2 Final Assessment Report

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N. 826232

WP9 – Pilots

D9.2 – Final Assessment Report

Project number:	826232
Project acronym:	WorkingAge
Project title:	Smart Working Environments for All Ages
Author(s):	Vera Rick (RWTH)
Partners contributed:	A. Bonanos (EXUS), V. Mpakalis (EXUS), Alberto Navarro (ITCL), Julen Rostan (ITCL), J. Melús Carruez (ITCL), V. Ronca (BS), V. Scotti (POLIMI), P. Loygue (GC), R. Antolín Pérez (INTRAS), H. Sagha (AUD), H. Gunes (UCAM), I. Maya (MUTUA), R. Almeida (INTRAS)
Delivery date:	M42 (July 2022)
Document date:	28/07/2022
Version:	1
Revision:	14
Remarks:	
Status:	⦿ PU (Public) ○ PP Restricted to other programme participants (including the Commission Services) ○ Restricted to a group specified by the consortium (including the Commission Services) (please specify the group) ○ Confidential, only for members of the consortium (including the Commission Services

DOCUMENT REVISION LOG

VERSION	REVISION	DATE	DESCRIPTION	AUTHOR
1	0	01/04/2022	Draft	V. Rick (RWTH)
1	1	28/06/2022	Input ZeroMQ, Gesture component input, input voice analysis submodule	A. Bonanos, V. Mpakalis (EXUS), V. Scotti (POLIMI)
1	2	01/07/2022	Input primary user statistics, tool evaluation, eye tracking subsystem, WAOW tool questionnaires, secondary & tertiary users evaluation	V. Rick (RWTH)
1	3	05/07/2022	Input environmental sensor	A. Navarro (ITCL)
1	4	08/07/2022	Input body pose	J. Rostan (ITCL)
1	5	12/07/2022	Input neurometric subsystem, DSS, questionnaires	V. Ronca (BS), V. Rick (RWTH)
1	6	13/07/2022	Input Wi-Fi, subjective well-being evaluation, formatting	P. Loygue (GC), R. Antolín Pérez (INTRAS), V. Rick (RWTH)
1	7	14/07/2022	Input activity tracker, WAOW app, first review, formatting	J. Melús Carruez (ITCL), V. Rick (RWTH)
1	8	14/07/2022	Input noiseBox, introduction & summary	H. Sagha (AUD), V. Rick (RWTH)
1	9	15/07/2022	Input well-being questionnaire, review	R. Antolín Pérez (INTRAS), V. Rick (RWTH)
1	10	19/07/2022	UCAM input as well as review of the document	H. Gunes (UCAM)
1	11	22/07/2022	Review of the document	I. Maya (MUTUA)
1	12	27/07/2022	Review of the document and incorporation of review comments	R. Almeida (INTRAS), R. Antolín Pérez (INTRAS)
1	13	28/07/2022	incorporation of review comments, formatting, finalizing document	V. Rick (RWTH)
1	14	28/07/2022	Final review and formatting	M. van Gasteren (ITCL)

Executive Summary

The content of this deliverable is the summary of the results of in-company field tests conducted within the WorkingAge Project.

We aim to assess the sustainability, accessibility, usability, validity, and usefulness of the WorkingAge system according to the standards and indicators of the criteria defined in the evaluation protocol (D9.1). Thus, the deliverable will provide an overview of:

1. The assessment of the benefits for the occupational health and quality of life from the perspective of the primary user (section 2.3), including a pre-questionnaire (questionnaire to assess the situation before the WAOW tool was used) and a post-questionnaire (questionnaire to assess the situation after the WAOW tool was used) in order to assess working condition, health status, social behaviour, general well-being and user experience and to be able to provide useful insights in changes over usage time.

2. The evaluation of the entire WAOW tool as well as single components from the perspective of the primary users in short- and long-term test (section 2.4) taking into account factors of usability and acceptability of the tool as well as an evaluation of all individual submodules, including interventions. Furthermore, possible future approaches are discussed from the point of view of primary users. The two-fold data collection (during short- and during long-term tests) enables a time-dependent evaluation in order to obtain the best possible overview of the aspects and components.

3. Assessment of the validity and usefulness of the individual components of the WAOW tool based on the data collected with the primary users within short- and long-term field tests (section 2.5 – 2.19). All sub-modules and components are evaluated according to the D9.1 protocol. Additional tests were carried out to quantify the DSS performance as it is a core aspect of the WAOW tool. In this regard, it was chosen to feed the final version of the DSS with the mental, emotional, and physical states classifications of the participants involved in the In-Lab experimental phase of the WorkingAge project. An overview of the components evaluates within this deliverable is given in Table 1.

4. Finally, the deliverable will provide an evaluation of the WAOW tool from the perspective of secondary and tertiary users (section 3). The focus is on surveying the most important stakeholders in the implementation process of the WAOW tool, which are on the one hand managers, human resources etc. (decision-makers) and on the other hand the potential users of the tool. Factors that have an influence on the decision regarding implementation as well as factors that affect the actual use are considered.

All these steps helped to collect feedback and comments on usability and validity to ensure the best possible provision of insights and input for the use and further development of the WAOW tool.

Table 1: Overview of individual components of the WAOW tool

Submodule	Method	Parameter	Section
Gesture recognition submodule	Short-term test evaluation	facial recognition and gesture interaction	**2.5**
Posture recognition submodule	Short-term test evaluation & long-term test evaluation	Sitting and standing risk	**2.6**
Facial affect analysis submodule	Short-term test evaluation & long-term test evaluation	Emotional valence and arousal detection	**2.7**
Neurometrics submodule	Short-term test evaluation & long-term test evaluation	Mental workload, stress level and emotional state	**2.8.**
Eye tracking submodule	Short-term test evaluation & long-term test evaluation	Mental strain detection	**2.9**
Voice analysis submodule	Short-term test evaluation & long-term test evaluation	Emotional status	**2.10**
Noise analysis submodule	In-Lab test evaluation	Environmental noise and human noise	**2.11**
Activity tracker	Short-term test evaluation & long-term test evaluation	Heart rate, steps, calorie consumption, sleep	**2.12**
Smart scale	Short-term test evaluation & long-term test evaluation	Weight and BMI	**2.13**
Environmental sensor	Short-term test evaluation & long-term test evaluation	Light, thermohygro-metrics, CO_2 concentration	**2.14**
WAOW tools questionnaires	Long-term test evaluation	Physical, mental and emotional state	**2.15**
WAOW tools Wi-Fi network	valuation per pilot use case		**2.16**
ZeroMQ	Short-term test evaluation & long-term test evaluation		**2.17**
WAOW App	Short-term test evaluation & long-term test evaluation		**2.18**
Decision Support System (DSS)	Performance tests		**2.19**

Table of Contents

List of Tables

List of Figures

1. Introduction

The objective of this task is to evaluate the effectiveness of the WAOW tool's intervention plan tailored to the workplace, as well as its acceptability and ability to promote engagement and adherence to the proposed plan through the use of the WorkingAge application. This means that the cause-effect relationship between the impacts in different user domains and the use of the WAOW tool under real conditions is evaluated. This evaluation will allow to determine if the WAOW tool is well accepted and usable under real conditions, and to determine the maximum tolerable interference levels to ensure that the target users can use the WAOW application in their real workplaces.

This evaluation will be guided by the objectives originally established by the KPIs defined for the WA project as well as described in the Evaluation Protocol (D9.1). During the in-company pilots, target users were assessed pre- and post-intervention to determine the sustainable impact of WAOW tool use on the various target dimensions of the workplace-tailored intervention plan (e.g., dietary habits, physical activity, sleep quality, social participation, etc.), with a particular focus on the physical and mental health and overall well-being of target users. This pre- and post-assessment included validated scales or relevant sections for the assessment purpose of the WA project.

Compliance with recommendations were captured through WA system records.

A multiple case study approach will be conducted to analyse the effectiveness of the tailored recommendations in terms of appropriateness, reliability, usefulness, utilization, and objective measures of adherence, as well as satisfaction with the personalized intervention plan achieved by selecting the most relevant and useful SMART goal modules for each participant's characteristics.

2. Primary Users

2.1 Introduction

The following report gives an overview of the evaluations related to all components of the WAOW tool, examining on the one hand:

- the verification of the correctness of the components of the WAOW tool in the real application context
- Validate the correctness of the sub-modules that make up the WAOW tool.
- Measure the acceptability, usability, and effectiveness of the WAOW tool.

The whole evaluation process is divided into two parts, depending on the length of the evaluation period, i.e., short term and long term.

The first part (short-term tests) deals with the evaluation of the correctness and quality of the hardware and software components of the WAOW tool through what we call reliability tests, following the protocol in D9.1.

Furthermore, the following long-term tests aim in particular to confirm the results of the reliability tests over a longer period of time and to evaluate the impact of the actions proposed by the tool, i.e., to verify that the tool achieves the goal for which it was originally developed.

The tests and results described in this chapter focus on primary users (i.e., users of the WAOW tool) and rely on statistical correlation measures between the extracted high-level information and the well-being improvement and attitudinal effect KPIs as defined in D9.1.

The chapter starts by giving an overview of the primary users of the WAOW tool (chapter 2.2.), taking into account how long and to what extent the WAOW tool was used. The subjective evaluation of the WAOW tool will be discussed further, on the one hand to what extent a change or an influence on their own well-being can be recognized (chapter 2.3) and on the other hand how the users evaluate the WAOW tool with regard to acceptance and usability (chapter 2.4).

This is followed by an overview and analysis of each individual component of the WAOW tool, evaluating the individual sensors as well as the other components such as the network, communication, and questionnaires of the tool. Finally, an overview and comparison of the measured data is given in order to compare individual sensors with each other (chapter 2.20). This will provide a holistic overview of subjective and objective data collected during the pilot tests with the primary users to be able to draw a conclusion about the usefulness and effectiveness of the WAOW tool.

2.2 Sample description

2.2.1 Use Case: Office (Mutua Universal)

2.2.1.1 Overview

Mutua Universal (MUTUA) covers the office workplace (Figure 1). The predominant work tasks are use of computers and desk tasks activities; online meetings and phone calls are also an important part of the tasks. The tasks associated with the office workplace can be summarized as knowledge work. The object of this work is information that must be processed by the employee, information needs to be combined and knowledge is generated. Essentially, the work consists of the application of job-specific methods, problem solving, implementation of strategies, and decision making. Work is characterized through the use of computer programs and by predominantly visual and auditory information intake and output (e.g., reading, writing, communication). Relevant physical environment influences are limited to indoor aspects, such as illumination. Other physical conditions, such as climate, mechanical vibrations, noise exposure, surface temperatures, dirty/wet environment, weather conditions, etc. do only play a minor role for this workplace. From a psychological point of view, office work is typically characterised by regulated working hours, it is less flexible than the teleworking use case having reduced autonomy.

Figure 1: Mutua workplaces

2.2.1.2 Participants

MUTUA UNIVERSAL is the company that agreed to participate with the Office use case in the WA project with 15 volunteers. Due to sickness absence from work, 14 MUTUA employees finally participated in the WorkingAge pilot tests until the end. Of these fourteen users, seven were male and the other seven were female (Figure 2).

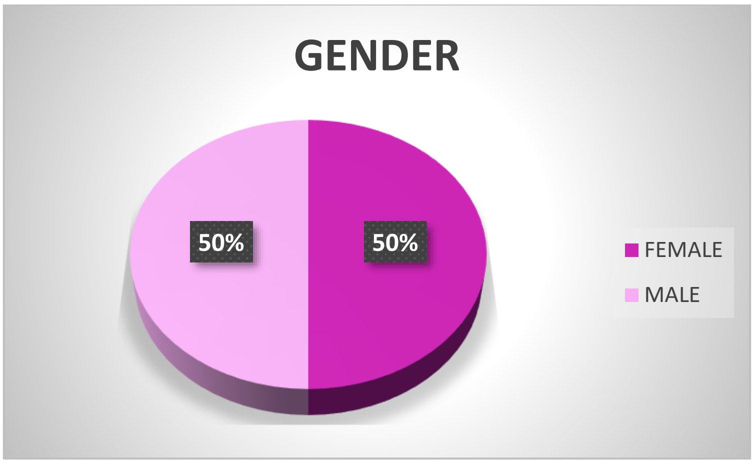


Figure 2: Office use case - gender

With regard to age, four users were under 45 years old and ten over 45 years old. A detailed description of the age range can be found in the following Figure 3.

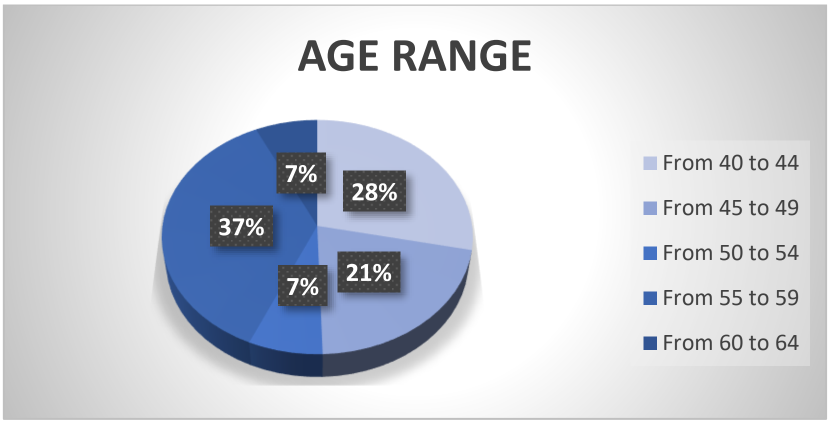


Figure 3: Office use case - age

As it is shown in Figure 4 14% of the participants, completed the post-secondary non tertiary education. Another 14% completed the short-cycle tertiary education. 43% of the volunteers have been graduated as bachelor or equivalent. And last, the 29% of the participants achieved a level of education of master or equivalent.

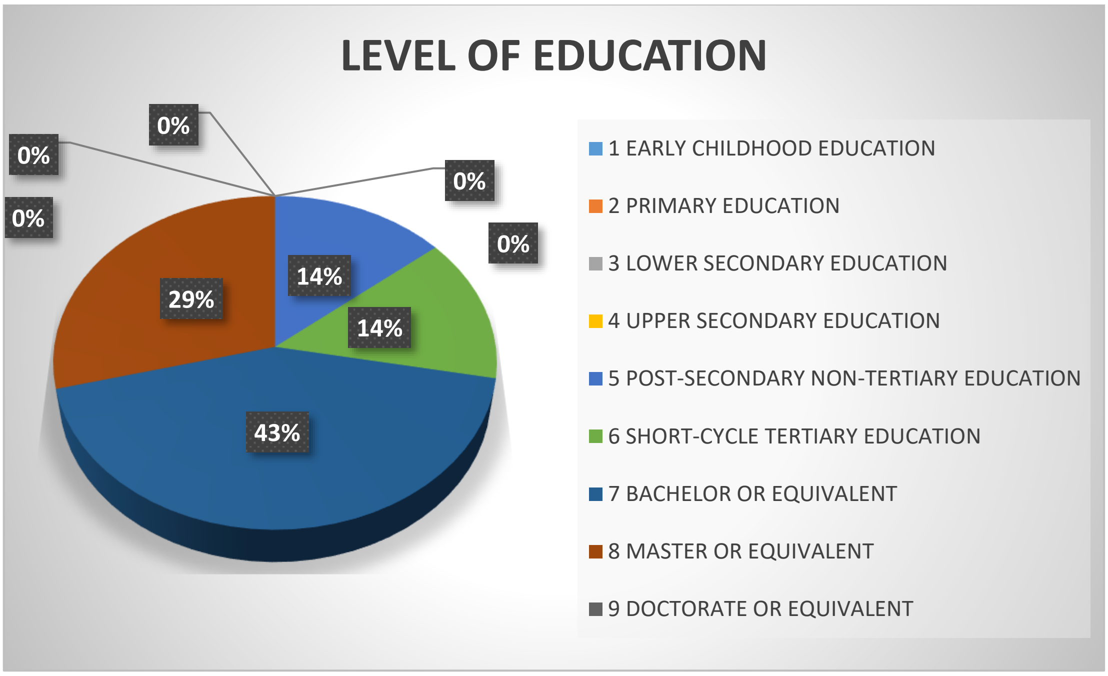


Figure 4: Office use case - level of education.

2.2.1.3 WAOW tool usage at MUTUA

The breakdown of sensors used from MUTUA employees was as follows: All users used the WAOW app, the smartband, the environmental sensors, the NoiseBox, the scale, the body posture recognition camera, the gesture recognition camera as well as the microphone for voice analysis. Five users used the eye tracking subsystem and three used the neurometric sensors (EEG headband and Empatica wristband).

Figure 5 shows the average days of use for each sensor used by participants. The only sensor that is missing is the NoiseBox. The NoiseBox recorded user data whilst being used, however, this data could not be evaluated due to wrong timestamps. Overall, the WAOW tool as used for more than five weeks (M = 37 days; SD = 12.9).

Deliverable D9.1, which is the basis for this final assessment report, describes additional tests related to the E112 and location service, a validation of the service is to be performed during short-term tests with a dedicated session in MUTUA. These have been moved forward into the integration testing period to bridge pandemic-related changes in schedules and to reserve resources for the upcoming effort related to the start of in-company testing. The associated results can be found in D8.2.

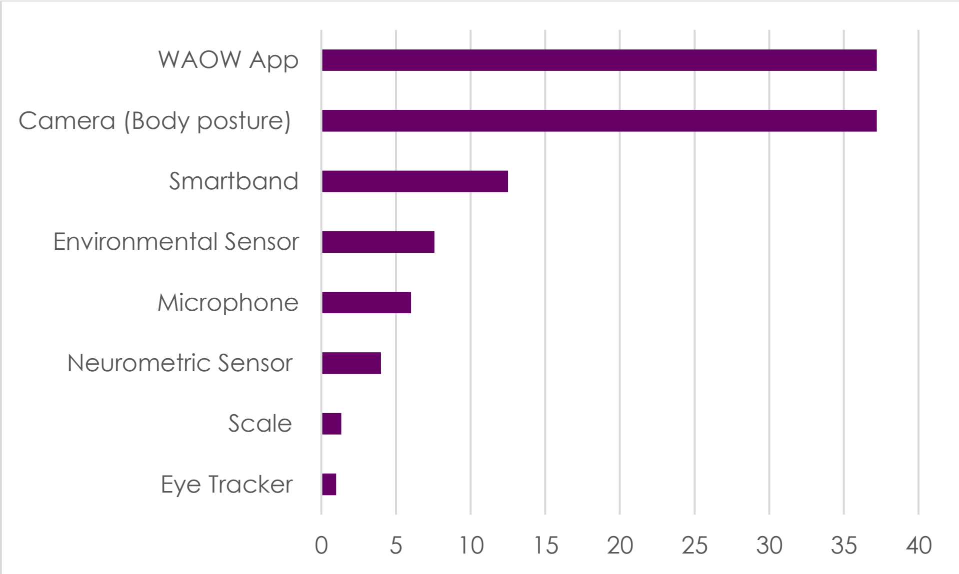


Figure 5: Average days of use for sensors (Office Use Case) – NoiseBox not evaluated due to missing Timestamps

Within the App users were able to choose a goal which they wanted to pursue. Most MUTUA employees chose a physical action goal.

2.2.2 Use Case: Production (Grupo Antolín RyA)

2.2.2.1 Overview

Grupo Antolín (GA) is one of the leading interior manufacturers worldwide. The workplaces studied in the WorkingAge project are in production environment, and the work tasks are primarily physical work activities. Production work involves performing movements, such as handling materials, and perceiving and responding to information, such as interacting with machines. In production halls, the physical environmental influences are much more important than in the office, as the worker is exposed to changing light or climate conditions or surface temperatures of materials, mechanical machine vibrations, noise, dirty and wet environments, chemicals (e.g. coolants). The following figures show an overview of the workplaces examined (Figure 6).

Figure 6: Grupo Antolín workplaces

2.2.2.2 Participants

For the Production use case, Grupo Antolín RyA, in Valladolid, accepted the challenge to collaborate with the WorkingAge project, contributing with 13 volunteers. All of them completed the pilots from the beginning to the end. All of the participants were over 45 years old, most of them over 50 years (62%) (Figure 7). Of these thirteen users, five were male and eight female employees, and all thirteen users were over the age of 45 years.

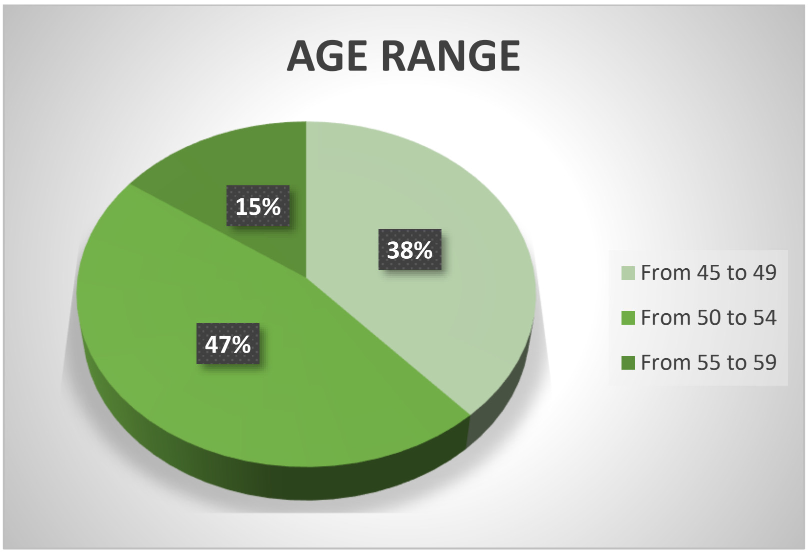


Figure 7: Production use case - Age range

62% of the users at Grupo Antolin were female, 38% male (Figure 8).

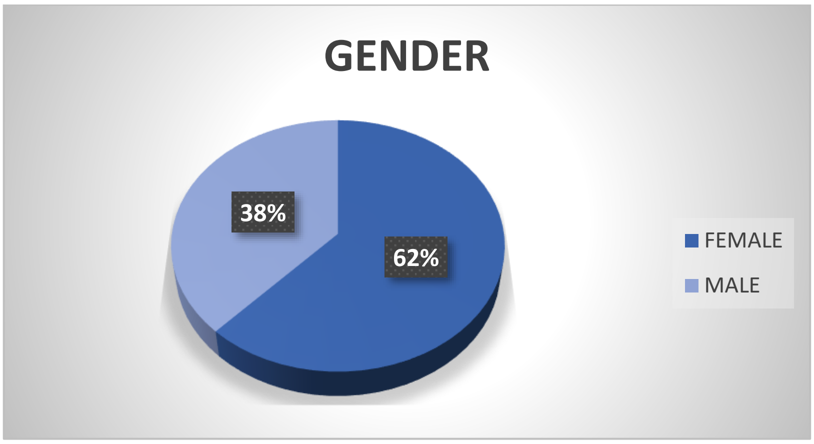


Figure 8: Gender. Production use case.

As it is shown in Figure 9, almost half of the participants completed early childhood education (46%), while 23% completed a primary education and furthermore 23% a lower secondary education.

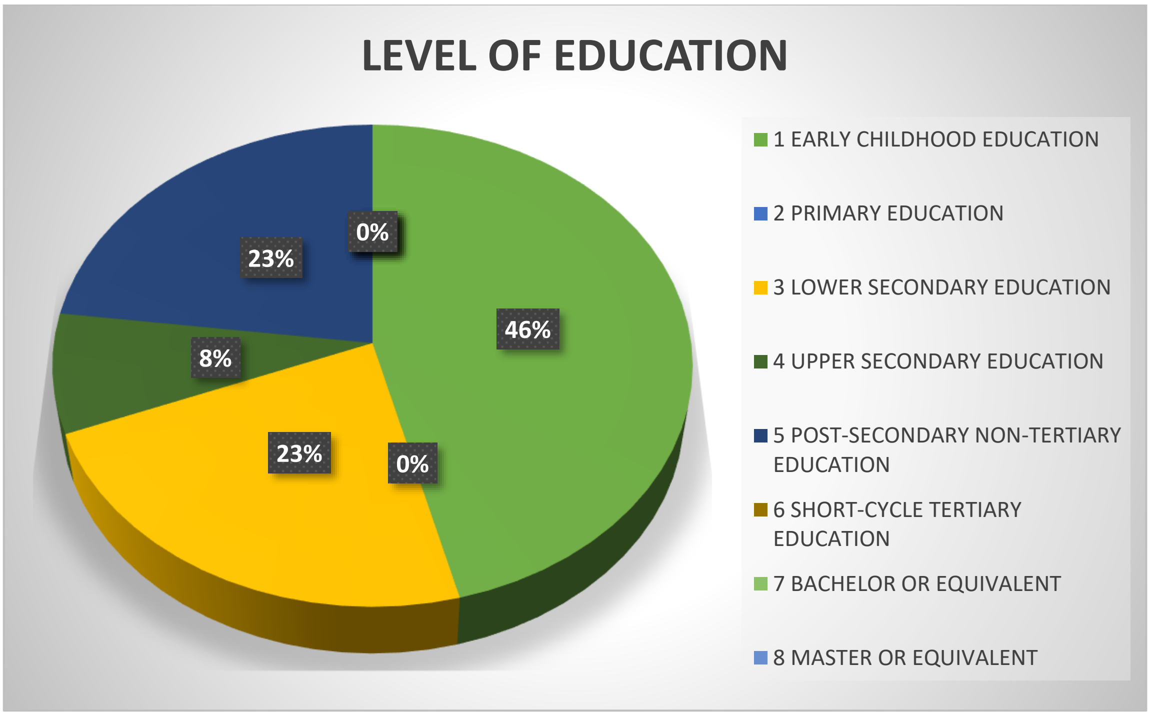


Figure 9: Level of education. Production use case.

Compared to the other use cases, it is the case in production that a wide variety of workplaces occur. Based on the participants and the possibilities of the WAOW tool, the following workplaces were considered: sewing workplace, assembly workplace, end-check workplace (Figure 10).

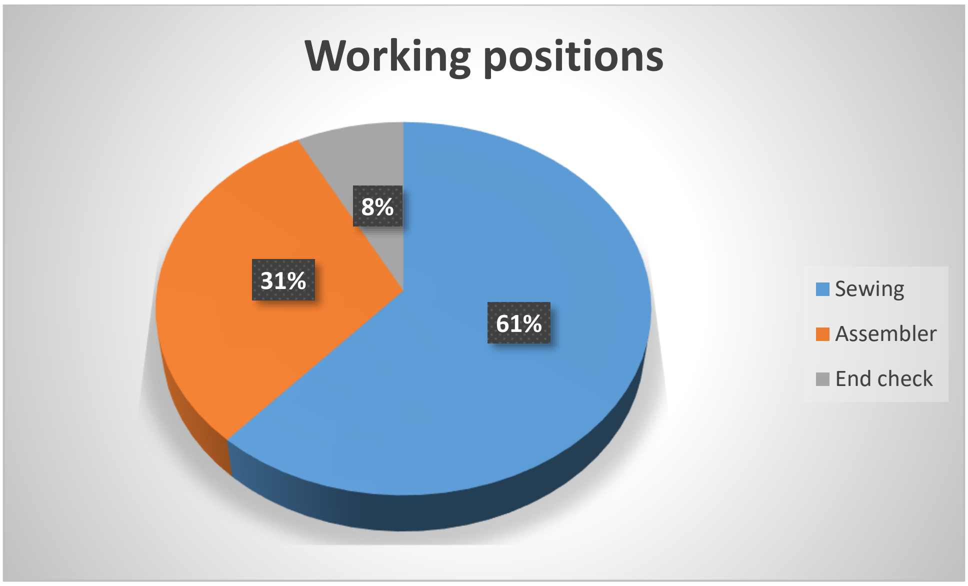


Figure 10: Working positions. Production use case.

2.2.2.3 *WAOW tool usage at GA*

The breakdown of sensors used in Grupo Antolin was as follows: All users used the WAOW app, the smartband, the environmental sensors, the noiseBox, the Scale, and the body posture recognition camera as well as the gesture recognition camera. Two users used the neurometrics sensors (EEG headband and Empatica wristband).

Overall, thirteen Grupo Antolín employees used the WAOW tool, however, of those thirteen only eleven could be fully evaluated. The WAOW tool was used for more than one month five weeks on average (M = 39.6 days, SD = 29.7). The sensor NoiseBox could not be evaluated due to missing Timestamps but was used by all users. Figure 11 shows the average days of use of sensors by Grupo Antolín employees.

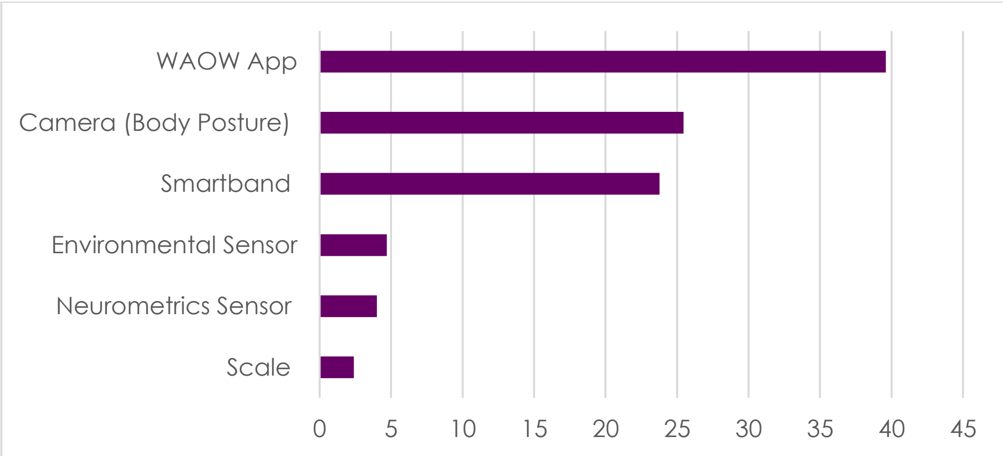


Figure 11: Average days of use for sensors (Production Use Case) – NoiseBox not evaluated due to missing Timestamps

Within the App users were able to choose a goal which they wanted to pursue. Most Grupo Antolín employees chose a sleep time goal.

2.2.3 Use Case: Teleworking (EXUS)

2.2.3.1 Overview

EXUS is an enterprise software company specializing in credit risk management, digital transformation services and innovation management. Due to the COVID-19 pandemic, EXUS hat transitioned to a remote working environment. The teleworking workplace and the tasks associated with it can be summarized as knowledge work. The object of this work is information that must be processed by the employee, information needs to be combined and knowledge is generated. Essentially, the work consists of the application of job-specific methods, problem solving, implementation of strategies, and decision making. Work is characterized by the use of computer programs and by predominantly visual and auditory information intake and output (e.g., reading, writing, communication). The fundamental differences from a traditional office workplace are changes in the mental and physical stress that distinguish a telecommuting workplace from a traditional office workplace. From a psychological point of view, the growing flexibility and autonomy can have a positive impact, but the lack of social support, possible isolation and a blurring boundary between work and private life can have a negative impact. With regard to physical stress, it is apparent that many teleworkers have poorer ergonomically designed workplaces than traditional office workplaces, for example that there is not always sufficient leg and foot room, and that adjustable seating is not available (for a more detailed description see D2.1.). Therefore, the analysis of the WAOW tool data will focus on these aspects in particular.

2.2.3.2 Participants

Nine EXUS employees have agreed to participate in the WorkingAge pilot tests. Of these users, five were male and four female employees, three under the age of 45 years, six over the age of 45 years (M = 47.33 years, SD = 4.82).

Age range: Two of the participants were under 40 years old, but as their age was close to 40, are being included in the first range "From 40 to 45".

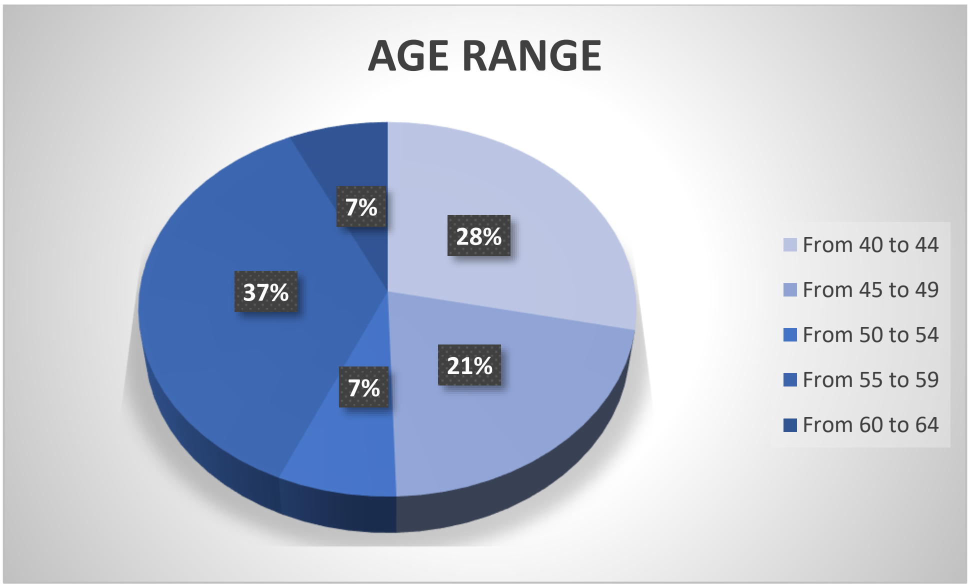


Figure 12: Age range. Office use case.

The number of participants divided by gender is exactly the same, 50% are female, and 50% are male.

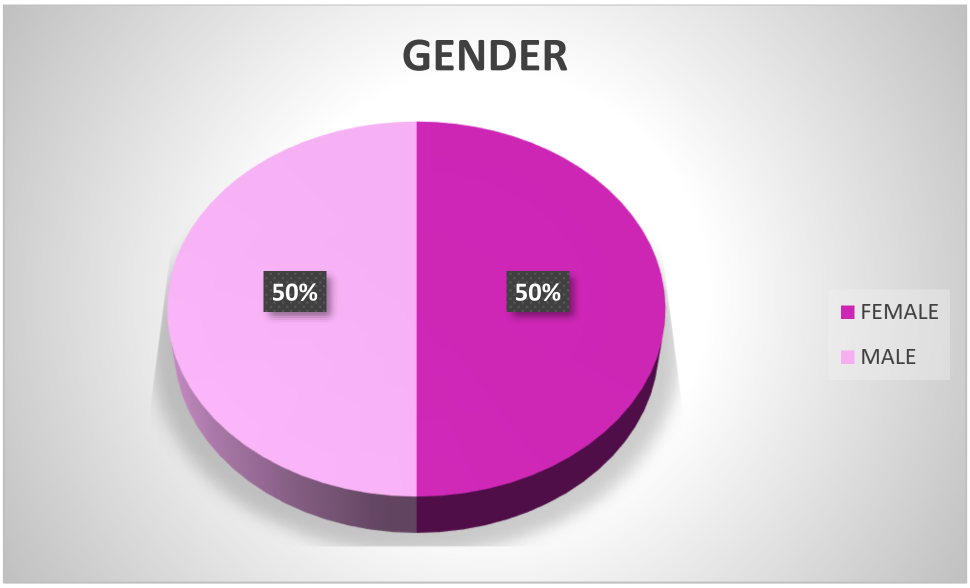


Figure 13: Gender. Office use case.

As it is shown in the following Figure, almost half of the participants completed a post-secondary non-tertiary education (43%), 29% completed a primary education while 14% completed an early childhood education as well as 14% a short-cycle tertiary education.

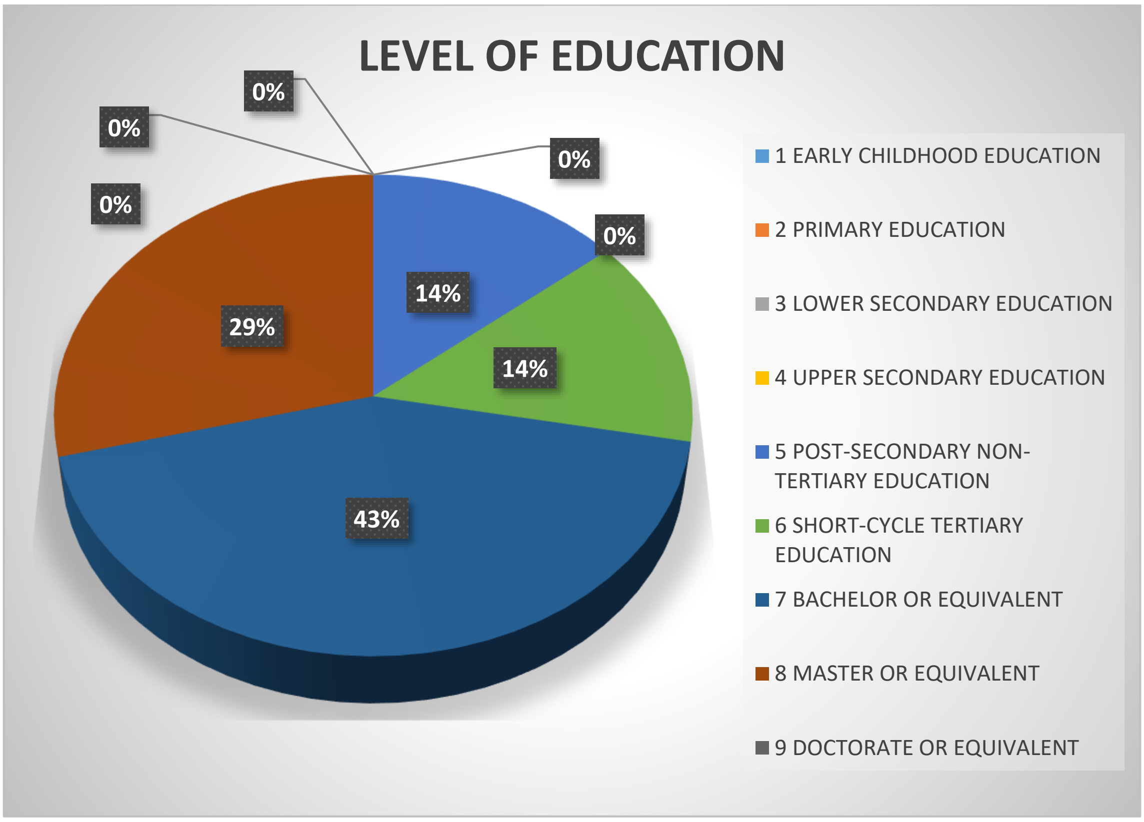


Figure 14: Level of education. Office use case.

2.2.3.3 WAOW tool usage at EXUS

The breakdown of sensors used in EXUS was as follows: All users used the WAOW app, the smartband, the environmental sensors, the noiseBox, the scale, and the body posture recognition camera as well as the gesture recognition camera. Seven users used the microphone for voice recognition and two users used the neurometric sensors (EEG headband and Empatica wristband).

Nine EXUS employees have agreed to participate in the WorkingAge pilot tests. The EXUS employees used the WAOW tool for more than two months on average (M = 68.1 days). The most used sensor amongst EXUS employees was the Microphone for voice analysis (M = 24.6 days, SD = 33.7), followed by the Smartband (M = 16.4 days, SD = 19.3). The NoiseBox could not be evaluated due to missing Timestamps but was used by all users. Figure 15 shows the average days of use for the sensors.

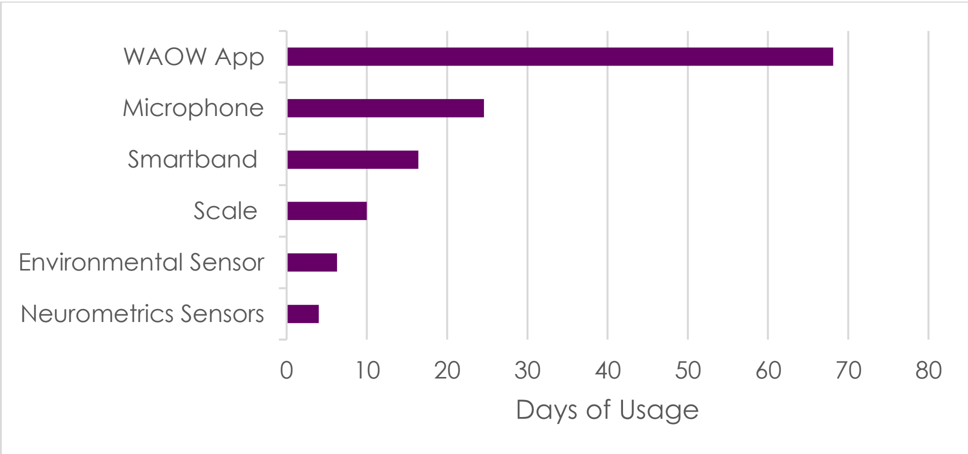


Figure 15: Average days of use for sensors (Teleworking Use Case) - NoiseBox not evaluated due to missing Timestamps

Within the App users were able to choose a goal which they wanted to pursue. Most EXUS employees chose a nutritional goal.

2.3 Subjective Well-Being

During the In-Company pilots, target users were assessed before and after the intervention to detect the sustained impact of the use of the WAOW tool on different target dimensions of the workplace-tailored intervention plan (e.g. nutritional habits, physical activity, sleep quality, social participation, etc.), with a specific focus on target user physical and mental health and overall well-being. These pre and post assessment include validated scales or relevant sections of them for the assessment purpose of the WA project.

The effectiveness of the WAOW tool workplace-tailored intervention plan was assessed as well as its degree of acceptability and ability to motivate the engagement and adherence to the proposed plan through the use of the WA application. This implies evaluating the cause-effect relationship between the impact produced in different user domains and the use of the WAOW tool under real conditions.

This assessment was guided by the objectives initially established through the KPI's set for the WA project.

2.3.1 Method

Using the WAOW tool the participants were asked through the link person in each company to complete the pre and post questionnaire.

The access to these questionnaires were located in the profile screen. See Figure 16 below.

Figure 16: Profile screen. Access to the pre and post questionnaires.

The participants completed questions related with the following areas (the different scales and surveys and their references are explained in the document D9.1_Annex 1 'EVALUATION PROTOCOL FOR LONG-TERM FIELD TESTS'):

1. *Physical Working Conditions*
2. *Psychosocial working conditions*
3. *Physical status*
4. *Cognitive/Mental status*
5. *Affective/Emotional status*
6. *Social relationships*
7. *General wellbeing & Quality of life*
8. *Health-related physical activity*

9. *Life-changing*
10. *User experience assessment*

To find out if the expected changes in the evaluation were due to the use of the WAOW application and its advice, or to another external circumstance, the following question was included with the option to complete the reason for it if the participant so wished.

- *Was there any significant situation in the last month affecting your life?*

2.3.2 Results

2.3.3 *Use Case: Office (Mutua Universal)*

MUTUA UNIVERSAL is the company that agreed to participate with the Office use case in the WA project with 15 volunteers. One of them fell ill and was off since the beginning of the pilots, so he did not participate. So, the final number of participants were 14.

At the beginning of the pilots, we asked the volunteers to complete the pre-questionnaire online. When the pilots finalized, we asked the same online questions to measure the wellbeing improvement after using the WAOW Tool.

2.3.3.1.1 Physical Working Conditions

To the question "Does your job ever require that you wear personal protective equipment?", 86% of the participants answered negative and only a 14% answered affirmative, as it is shown in the Figure 17.

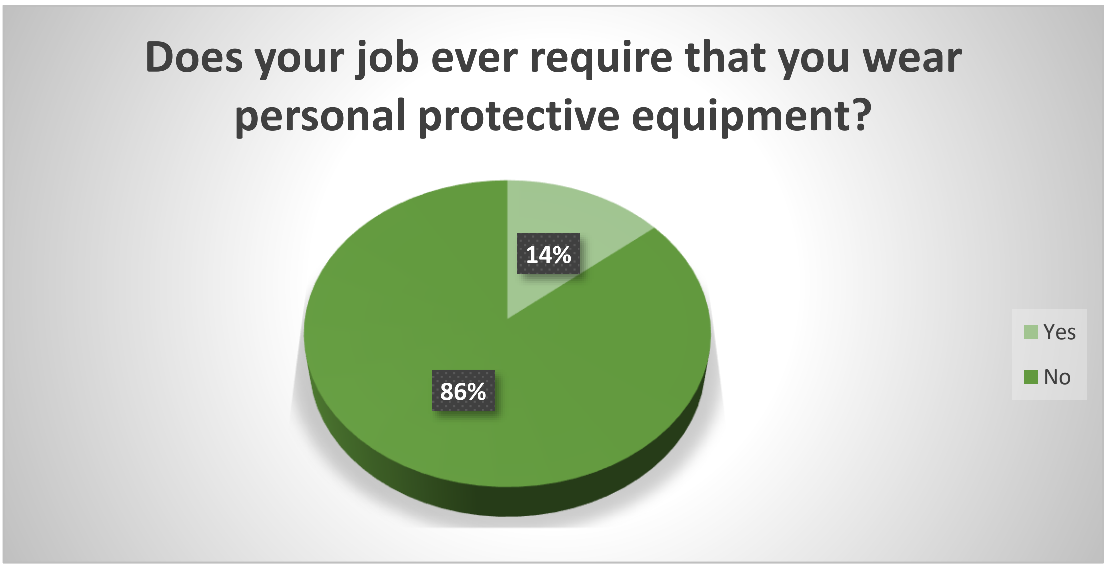


Figure 17: Protective equipment. Office use case.

100% of the workers who must wear this protective equipment, always use it.

Next are the answers of the office use case workers comparing the pre and post.

When the volunteers were asked if they health or safety were at risk because of their work, they gave the following answers:

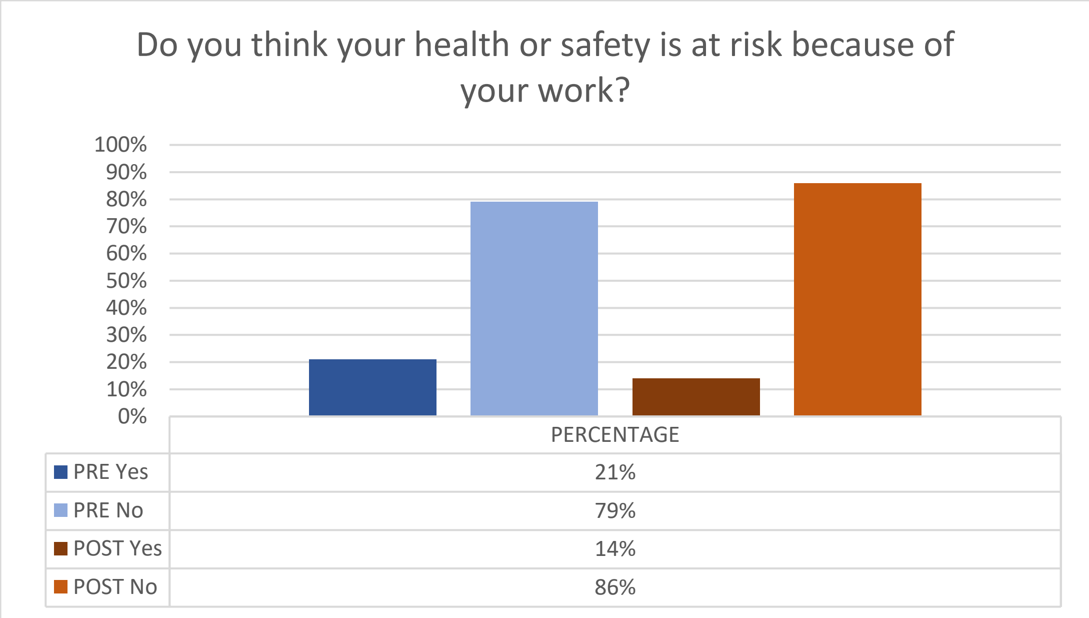


Figure 18: Risk because of work. Office use case.

The office volunteers were asked if their work affected their health. This is shown in the next figure.

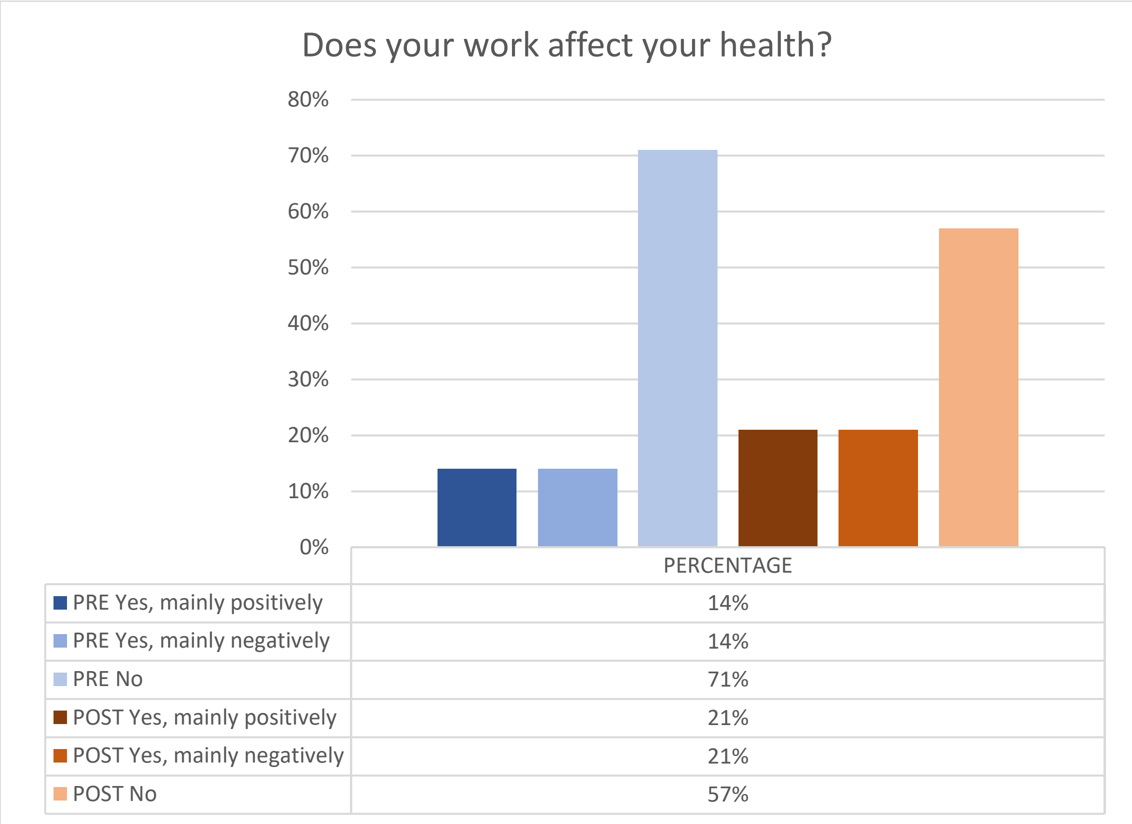
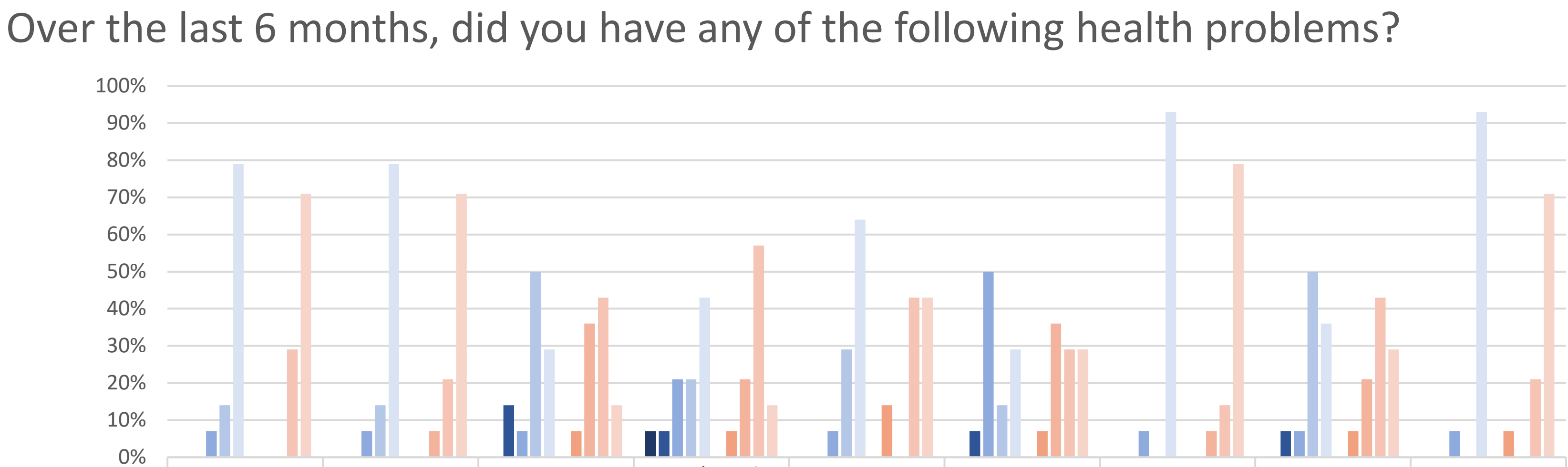


Figure 19: Health affected by the work. Office use case.

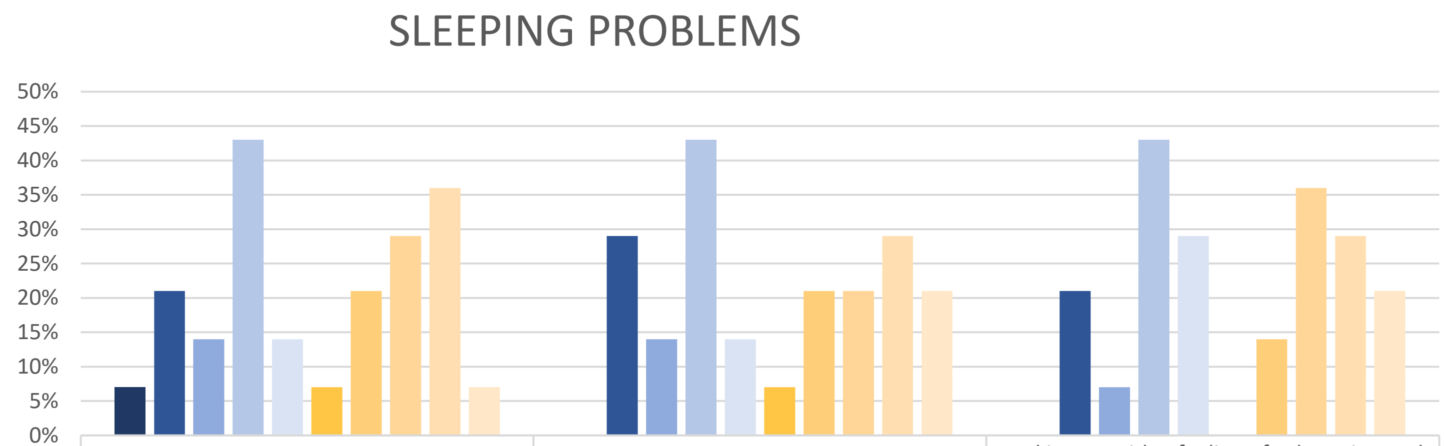
The following graph shows if the participants suffered from different health problems, comparing the pre and post (Figure 20).



	Hearing problems	Skin problems	Backache	Muscular pains in shoulders, neck and/or upper limbs	Muscular pains in lower limbs	Headaches, eyestrain	Injury	Overall fatigue	Other (Spontaneus)
PRE All the time	0%	0%	0%	7%	0%	0%	0%	0%	0%
PRE A large part of the time	0%	0%	14%	7%	0%	7%	0%	7%	0%
PRE Part of the time	7%	7%	7%	21%	7%	50%	7%	7%	7%
PRE A small part of the time	14%	14%	50%	21%	29%	14%	0%	50%	0%
PRE Not at all	79%	79%	29%	43%	64%	29%	93%	36%	93%
POST All the time	0%	0%	0%	0%	0%	0%	0%	0%	0%
POST A large part of the time	0%	0%	7%	7%	14%	7%	0%	7%	7%
POST Part of the time	0%	7%	36%	21%	0%	36%	7%	21%	0%
POST A small part of the time	29%	21%	43%	57%	43%	29%	14%	43%	21%
POST Not at all	71%	71%	14%	14%	43%	29%	79%	29%	71%

Figure 20: Health problems. Office use case.

The next question is "Over the last 6 months, how often did you have any of the following sleep related problems?". The volunteers were asked about 3 different aspects about sleeping problems:



	Difficulty falling asleep	Waking up repeatedly during the sleep	Waking up with a feeling of exhaustion and fatigue
PRE Daily	7%	0%	0%
PRE Several times a week	21%	29%	21%
PRE Several times a month	14%	14%	7%
PRE Less often	43%	43%	43%
PRE Never	14%	14%	29%
POST Daily	7%	7%	0%
POST Several times a week	21%	21%	14%
POST Several times a month	29%	21%	36%
POST Less often	36%	29%	29%
POST Never	7%	21%	21%

Figure 21: Sleeping problems. Office use case.

The next questions are about times when the participants might have been absent from work.

- Over the past 6 months how many days in total were you absent from work due to sick leave or health-related leave? (Figure 22)

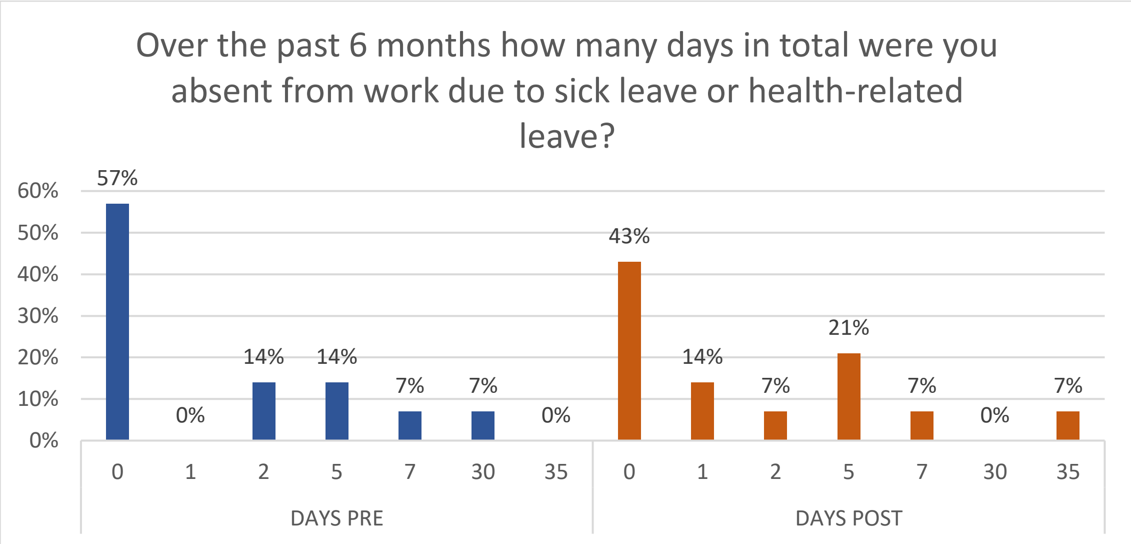


Figure 22: Absences from work. Office use case.

- Asked if these days were caused by accidents at work, the participants, answered the following:

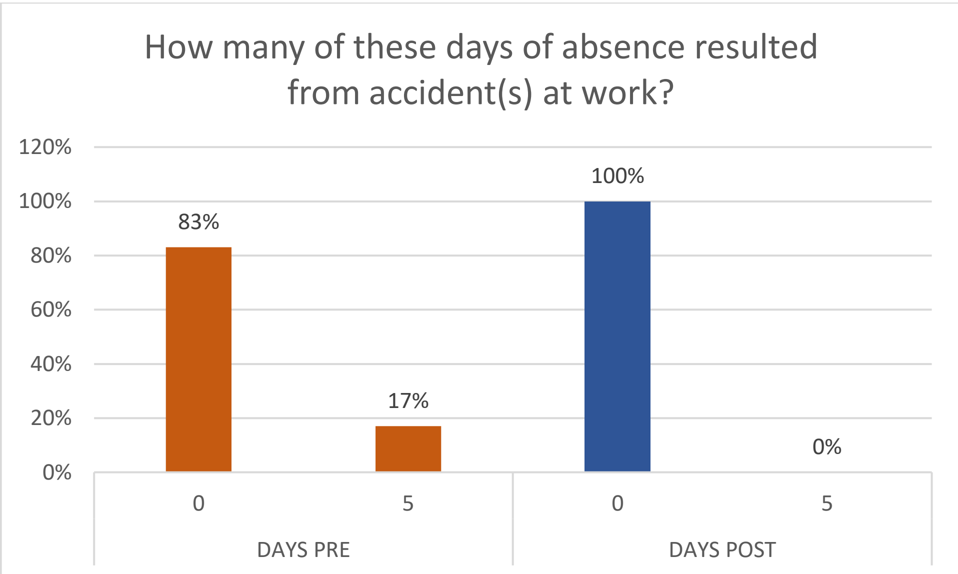


Figure 23: Accidents at work. Office use case.

- To the question if these days of absence resulted from health problems caused or made worse by their work (excluding accidents), the participants answered this way:

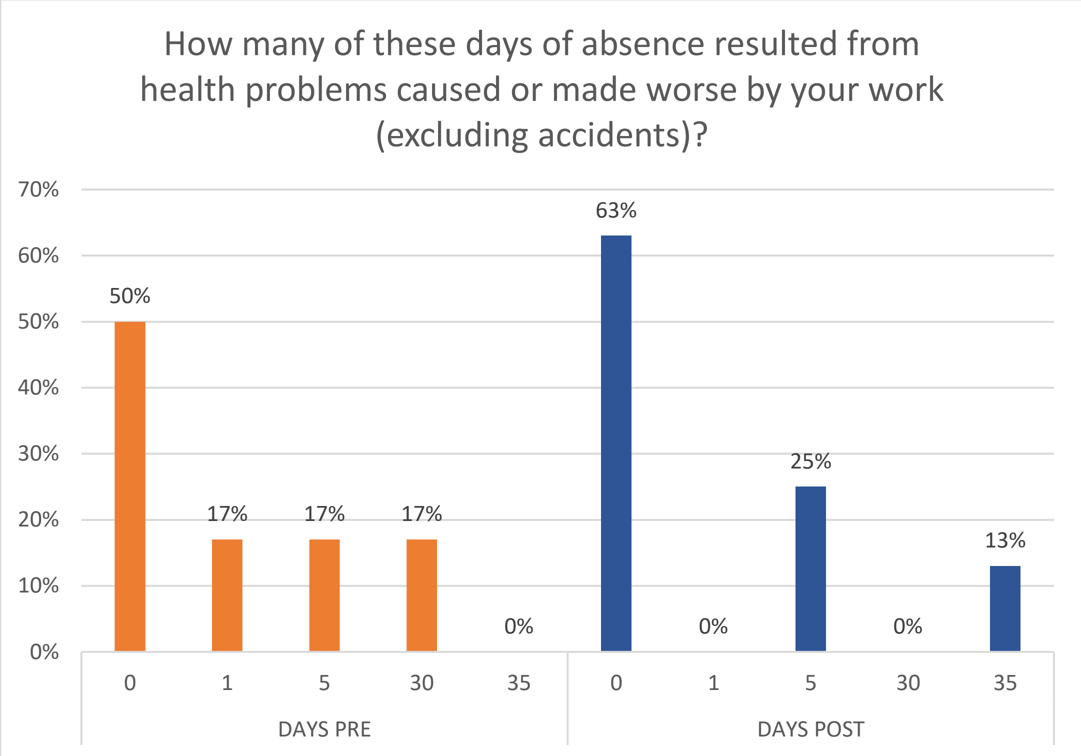


Figure 24: Health problems caused or made worse by work. Office use case.

- The volunteers were asked about the number of days that they work when they were sick. To have a more graphic view of the data, you can see the graph below (Figure 25).

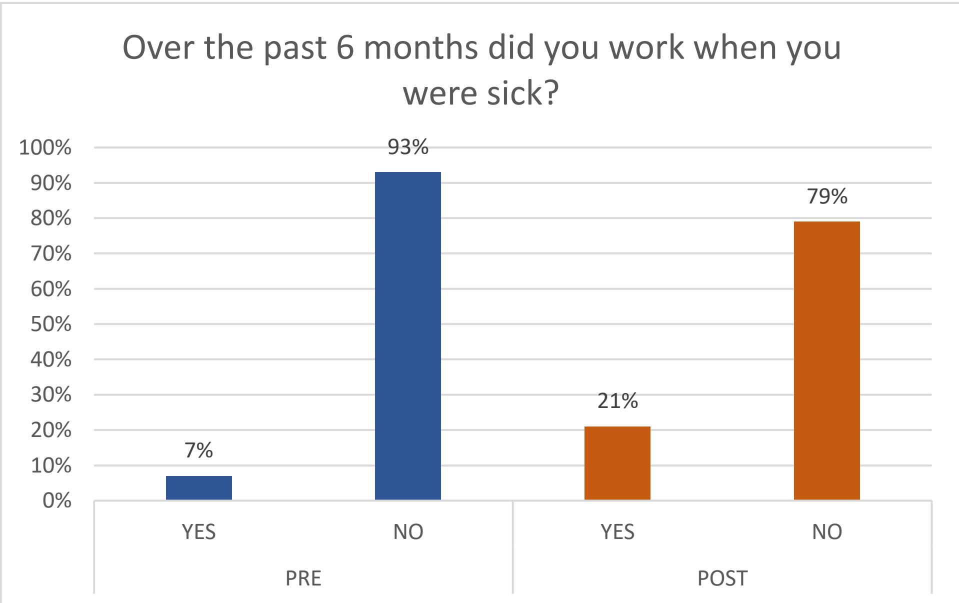
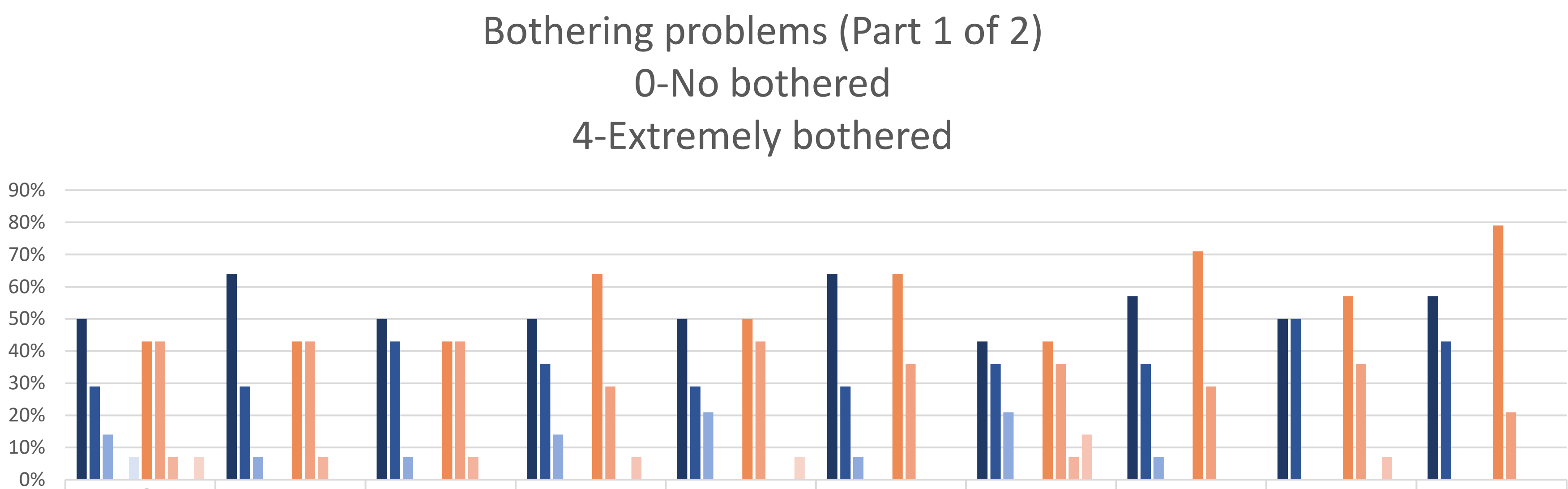


Figure 25: Working when sick. Office use case.

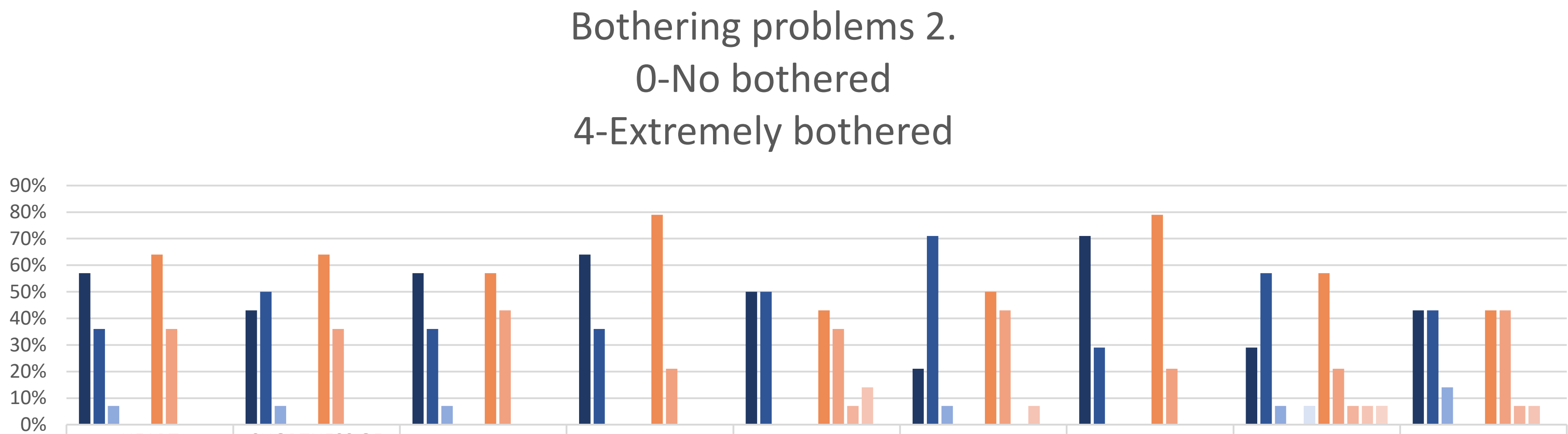
The 7% who answered yes to the previous question, worked 10 days when were sick. In the post-questionnaire, one of the workers who answered yes, worked 2 days when were sick. Two other volunteers worked 5 days in the same conditions.

To end with the questions about physical working conditions, the volunteers were asked about the physical symptoms they had suffered during the last few weeks and how much discomfort these problems caused them. The answering options are divided from 0 to 4, where 0 means that the participant has not been bothered by the problem, and 4 means that the problem has been an extreme bother. The graph is divided in two, due to the large number of symptoms related.



	WEIGHT CHANGE (gain or loss of 5 libs. or more)	CONSTIPATION	DIZZINES	DIARRHEA	FAINTNESS	NAUSEA AND/OR VOMITING	ACID STOMACH OR INDIGESTION	STOMACH PAIN (E.G., CRAMPS)	HOT OR COLD SPELLS	HANDS TREMBLING
PRE 0	50%	64%	50%	50%	50%	64%	43%	57%	50%	57%
PRE 1	29%	29%	43%	36%	29%	29%	36%	36%	50%	43%
PRE 2	14%	7%	7%	14%	21%	7%	21%	7%	0%	0%
PRE 3	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
PRE 4	7%	0%	0%	0%	0%	0%	0%	0%	0%	0%
POST 0	43%	43%	43%	64%	50%	64%	43%	71%	57%	79%
POST 1	43%	43%	43%	29%	43%	36%	36%	29%	36%	21%
POST 2	7%	7%	7%	0%	0%	0%	7%	0%	0%	0%
POST 3	0%	0%	0%	7%	0%	0%	14%	0%	7%	0%
POST 4	7%	0%	0%	0%	7%	0%	0%	0%	0%	0%

Figure 26: Bothering problems in the last 2 weeks 1. Office use case.



	HEART PALPITATIONS (Heart pounding/racing)	SHORTNESS OF BREATH when not exercising or working hard	NUMBNESS OR TINGLING in parts of your body	PAINS IN HEART OR CHEST	STUFFY HEAD OR NOSE	BLURRED VISION	NOSEBLEED	JOINT PAINS	COLD OR COUGH
PRE 0	57%	43%	57%	64%	50%	21%	71%	29%	43%
PRE 1	36%	50%	36%	36%	50%	71%	29%	57%	43%
PRE 2	7%	7%	7%	0%	0%	7%	0%	7%	14%
PRE 3	0%	0%	0%	0%	0%	0%	0%	0%	0%
PRE 4	0%	0%	0%	0%	0%	0%	0%	7%	0%
POST 0	64%	64%	57%	79%	43%	50%	79%	57%	43%
POST 1	36%	36%	43%	21%	36%	43%	21%	21%	43%
POST 2	0%	0%	0%	0%	7%	0%	0%	7%	7%
POST 3	0%	0%	0%	0%	14%	7%	0%	7%	7%
POST 4	0%	0%	0%	0%	0%	0%	0%	7%	0%

Figure 27: Bothering problems in the last 2 weeks 1. Office use case.

2.3.3.1.2 *Psychosocial working conditions*

To evaluate the psychological working conditions, the volunteers were asked the following questions in the pre and post questionnaires:

- "Is your work meaningful?"

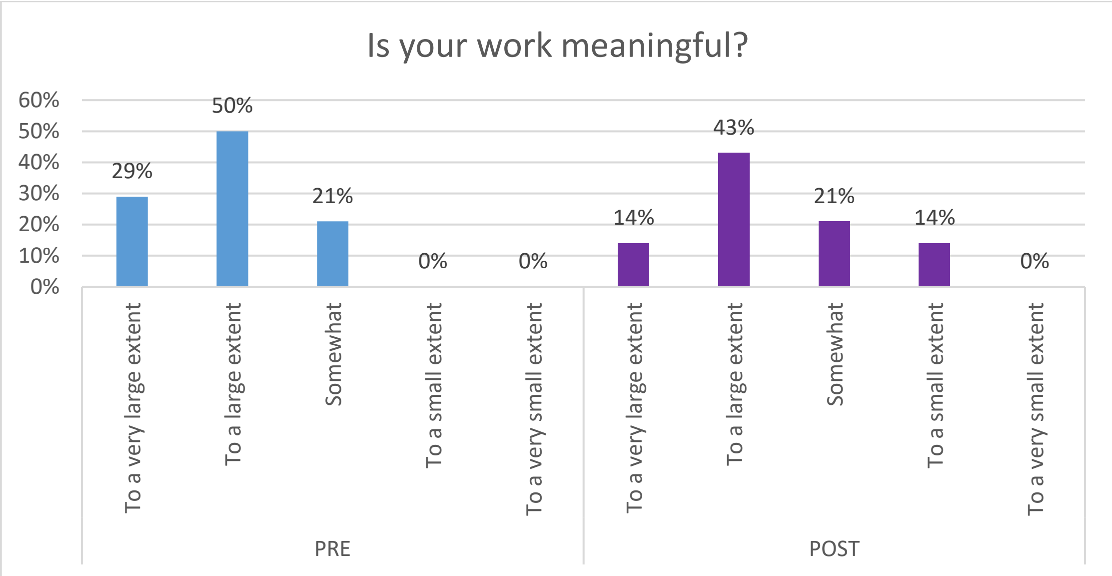


Figure 28: Meaning of work. Office use case.

- "How pleased are you with your job as a whole?"

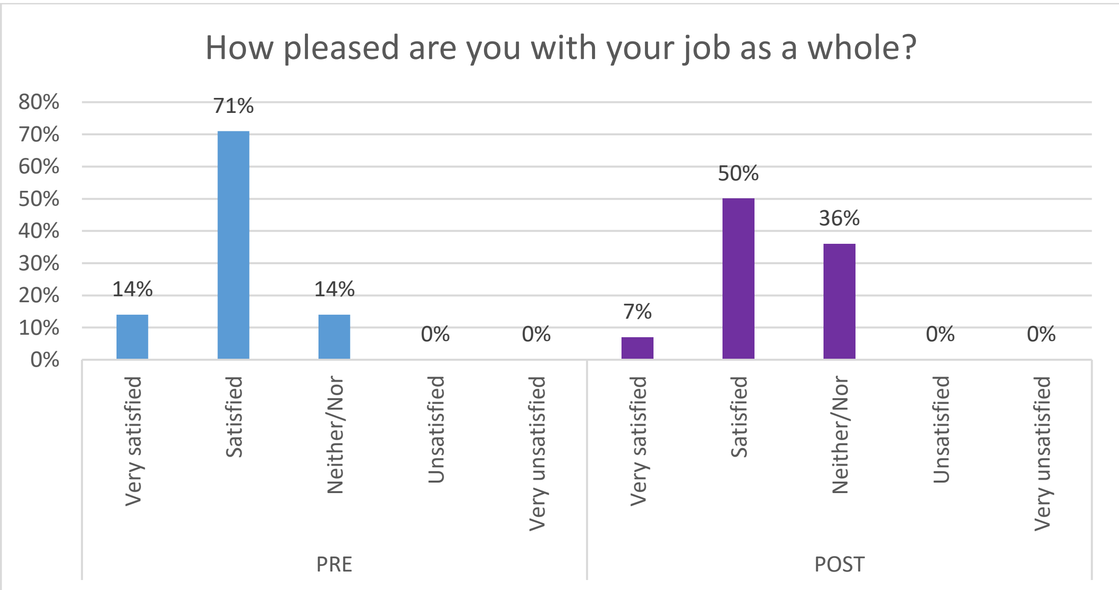


Figure 29: Job satisfaction. Office use case.

- "How often have you felt worn out?"

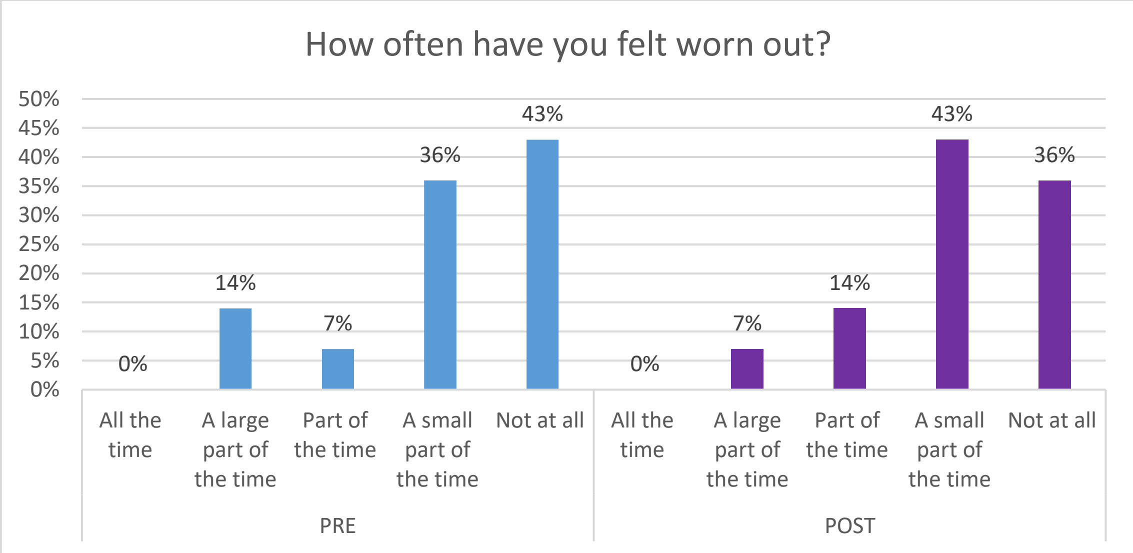


Figure 30: Burn out and stress 1. Office use case.

- "How often have you been physically exhausted?"

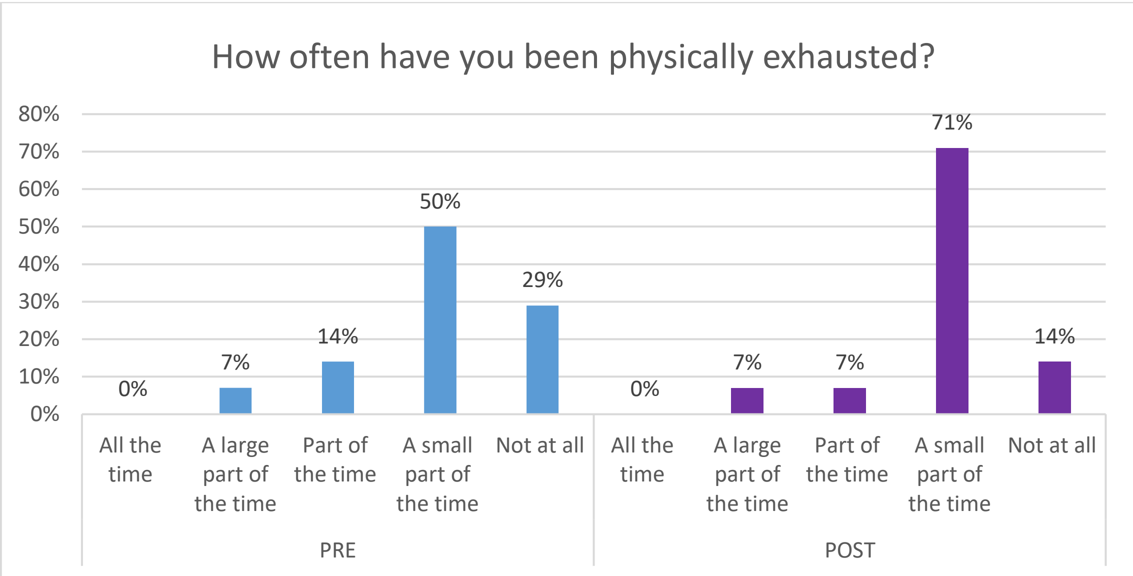


Figure 31: Burn out and stress 2. Office use case.

- "How often have you been emotionally exhausted?"

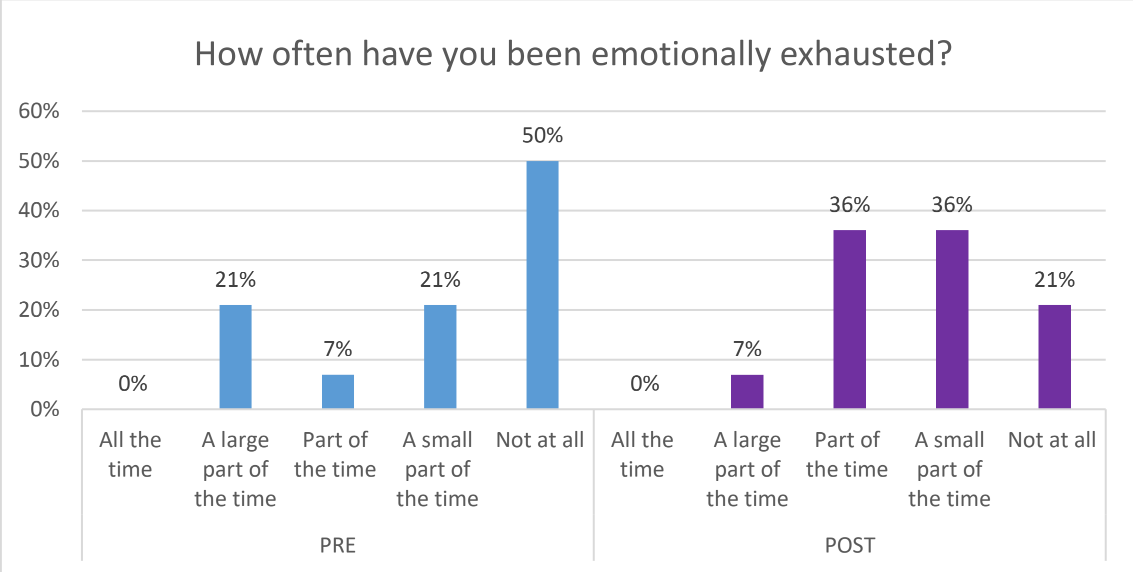


Figure 32: Burn out and stress 3. Office use case.

- "How often have you been tense?"

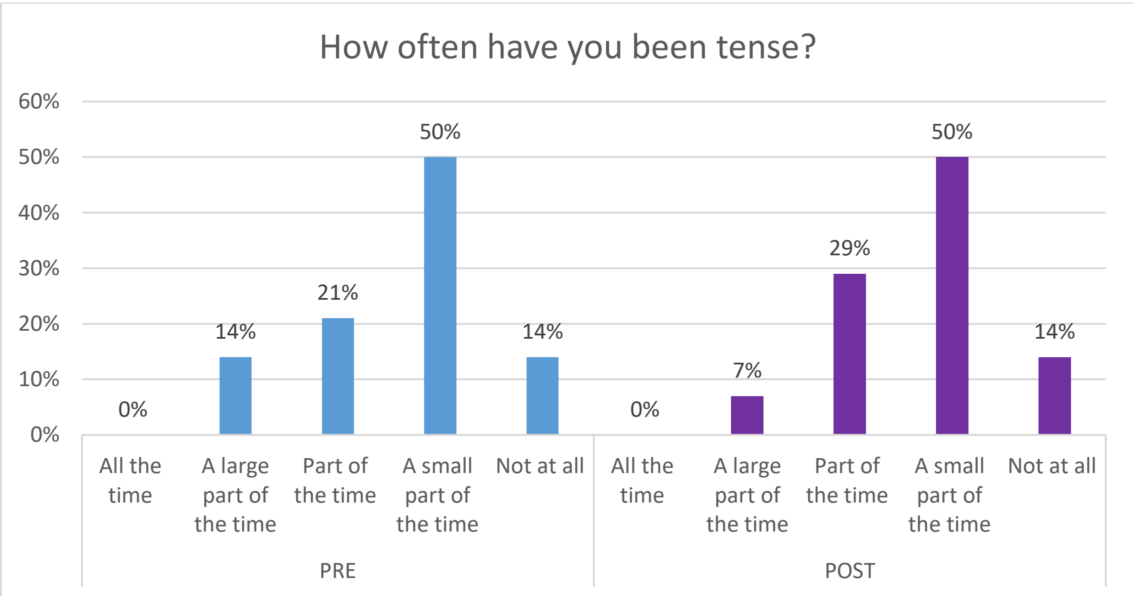


Figure 33: Burn out and stress 4. Office use case.

- "Is your work recognized and appreciated by the management?"

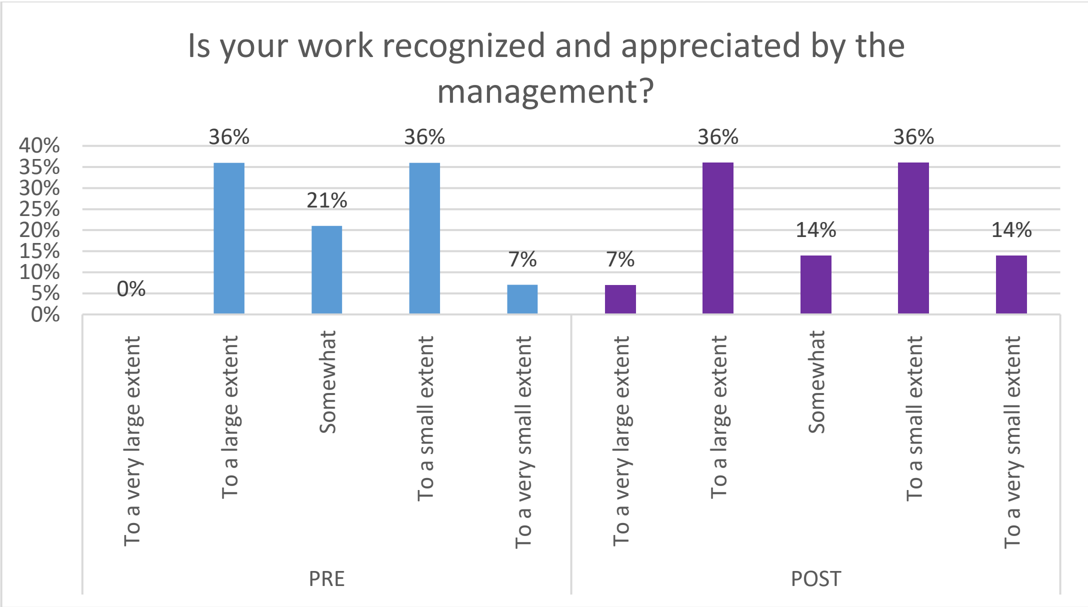
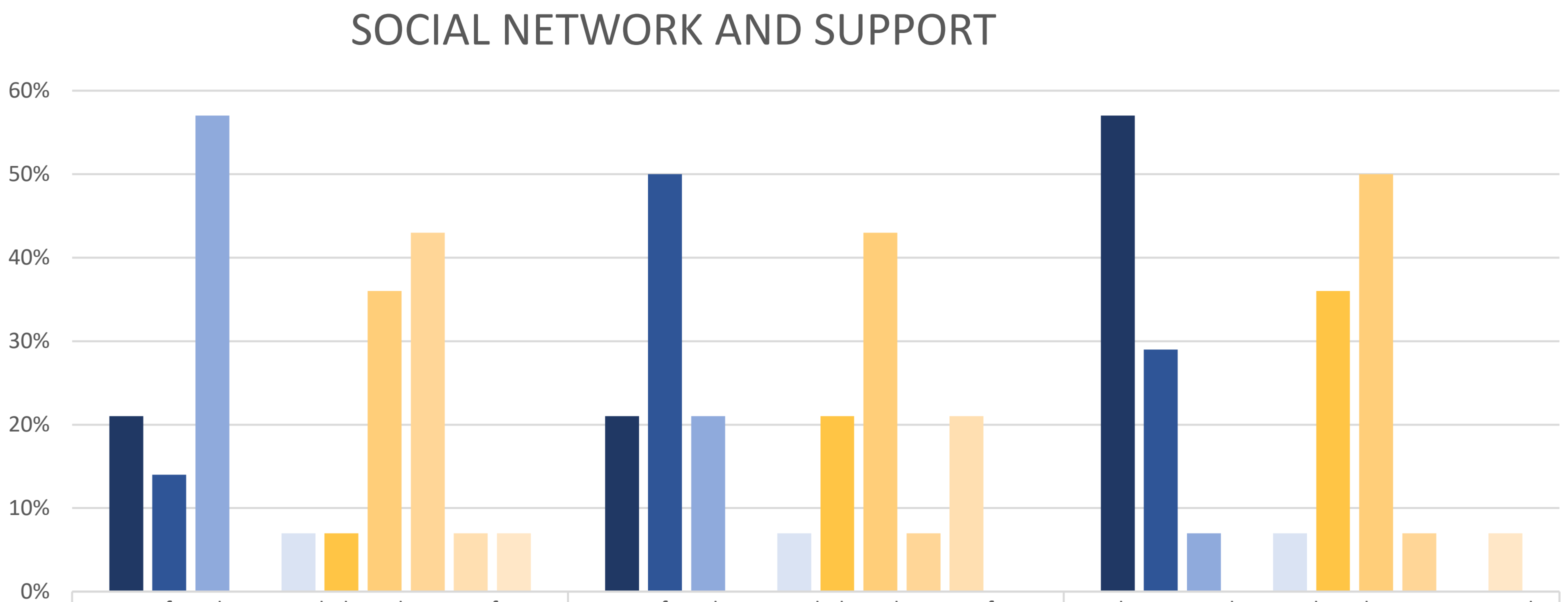


Figure 34: Social network and Support 1. Office use case

The next figure shows the social network and support felt by the volunteers in the pre and post questionnaire. To evaluate this aspect, three questions were asked to the participants.

- "How often do you get help and support from your immediate superior, if needed?"
- "How often do you get help and support from your colleagues, if needed?"
- "Is there a good atmosphere between you and your colleagues?"

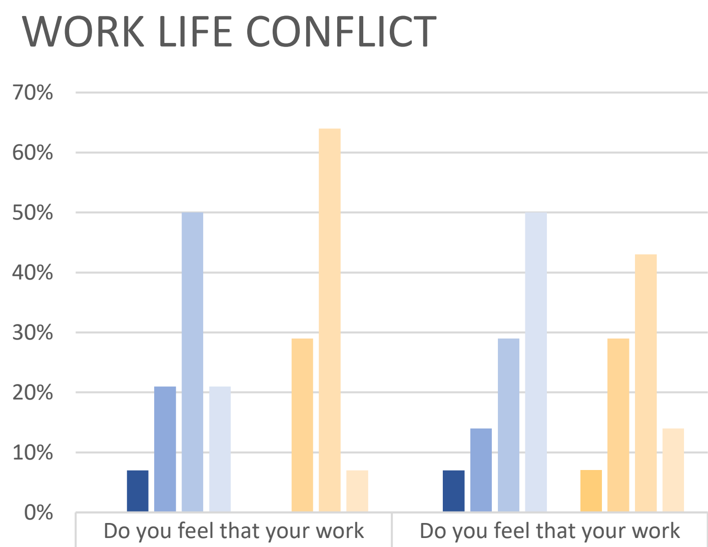


	How often do you get help and support from your immediate superior, if needed?	How often do you get help and support from your colleagues, if needed?	Is there a good atmosphere between you and your colleagues?
PRE Always	21%	21%	57%
PRE Often	14%	50%	29%
PRE Sometimes	57%	21%	7%
PRE Seldom	0%	0%	0%
PRE Never/hardly ever	7%	7%	7%
POST Always	7%	21%	36%
POST Often	36%	43%	50%
POST Sometimes	43%	7%	7%
POST Seldom	7%	21%	0%
POST Never/hardly ever	7%	0%	7%

Figure 35: Social network and support 2. Office use case.

We continue with two questions related to work life conflict. They can be seen in detail in the next graphic.

- “Do you feel that your work drains so much of your energy that it has a negative effect on your private life?”
- “Do you feel that your work takes so much of your time that it has a negative effect on your private life?”



	Do you feel that your work drains so much of your energy that it has a negative effect on your private life?	Do you feel that your work takes so much of your time that it has a negative effect on your private life?
PRE To a very large extent	0%	0%
PRE To a large extent	7%	7%
PRE Somewhat	21%	14%
PRE To a small extent	50%	29%
PRE To a very small extent	21%	50%
POST To a very large extent	0%	0%
POST To a large extent	0%	7%
POST Somewhat	29%	29%
POST To a small extent	64%	43%
POST To a very small extent	7%	14%

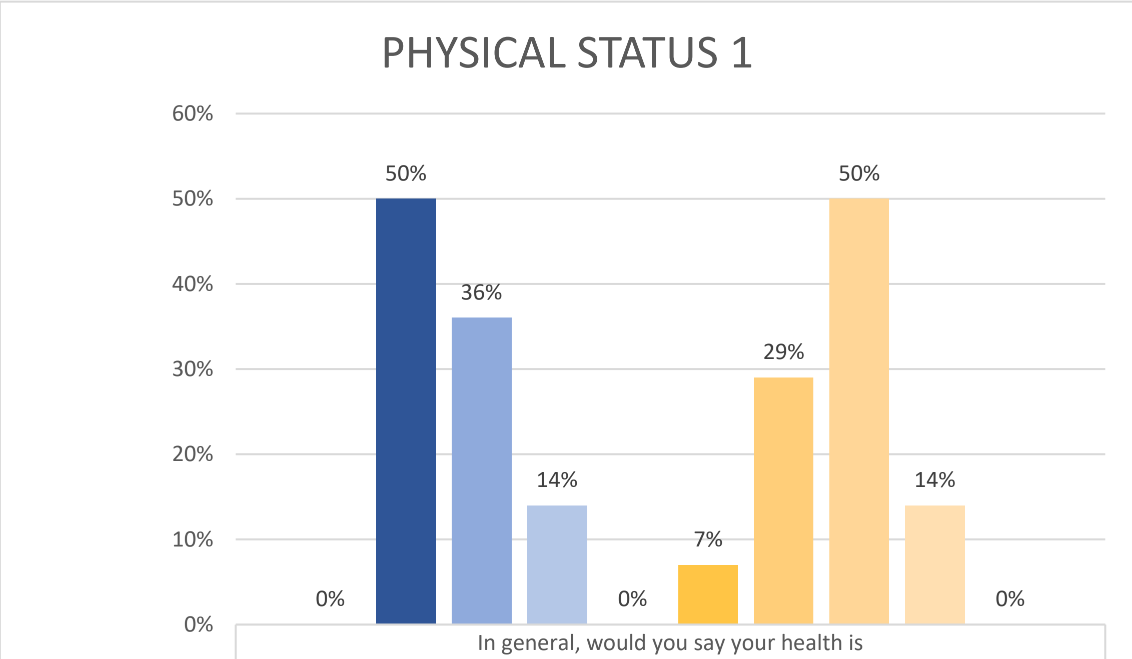
Figure 36: Work life conflict. Office use case

2.3.3.1.3 *Physical status*

To evaluate the physical status of the participants at the beginning and at the end of the pilots, they were asked seven questions related with this subject.

Five graphs will show the detailed data of their answers.

- "In general, would you say your health is…"

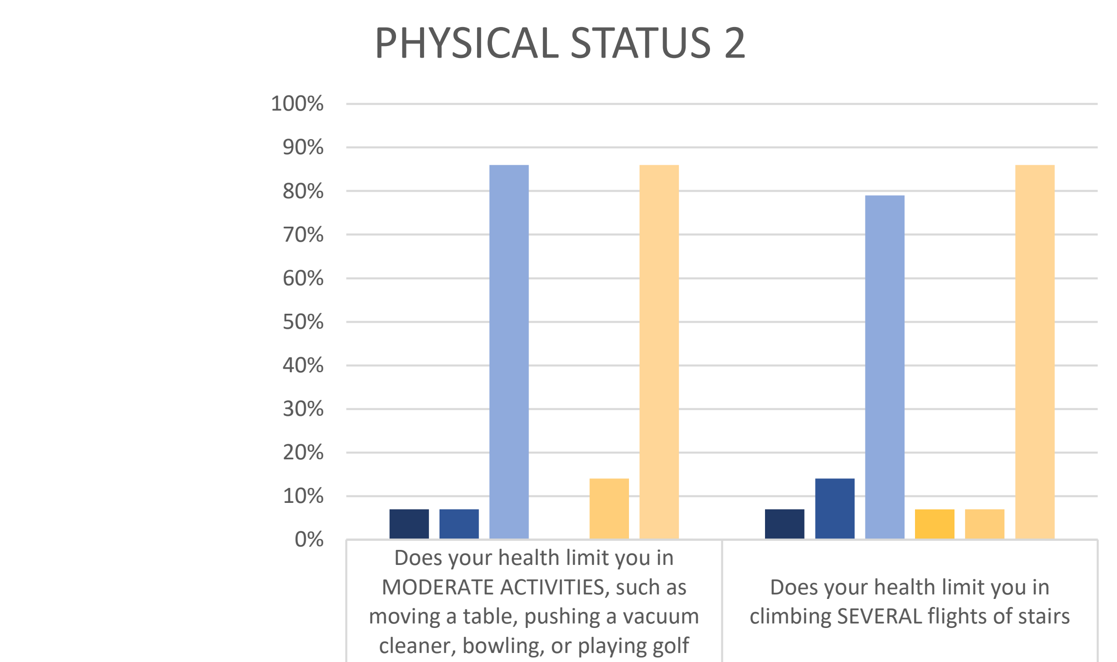


	In general, would you say your health is
PRE Excellent	0%
PRE Very good	50%
PRE Good	36%
PRE Fair	14%
PRE Poor	0%
POST Excellent	7%
POST Very good	29%
POST Good	50%
POST Fair	14%
POST Poor	0%

Figure 37: Physical status1. Office use case.

The next figure shows two questions related with the physical status of the participants at the beginning and at the end of the pilots.

- "Does your health limit you in MODERATE ACTIVITIES, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf?"
- "Does your health limit you in climbing SEVERAL flights of stairs?"



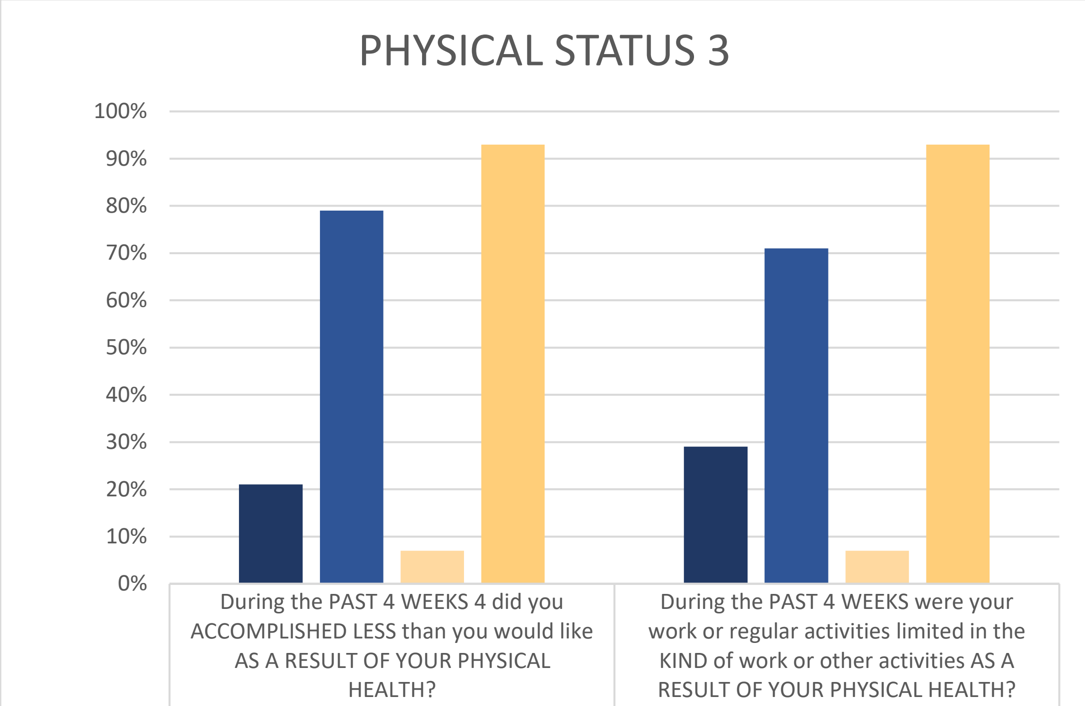
	Does your health limit you in MODERATE ACTIVITIES, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf	Does your health limit you in climbing SEVERAL flights of stairs
PRE Yes, Limited A Lot	7%	7%
PRE Yes, Limited A Little	7%	14%
PRE No, Not Limited At All	86%	79%
POST Yes, Limited A Lot	0%	7%
POST Yes, Limited A Little	14%	7%
POST No, Not Limited At All	86%	86%

Figure 38: Physical status 2. Office use case.

Again, the next figure shows two questions related with the physical status of the participants in a period of four weeks in advance (Figure 39).

- "During the PAST 4 WEEKS did you ACCOMPLISHED LESS than you would like AS A RESULT OF YOUR PHYSICAL HEALTH?"

- "During the PAST 4 WEEKS were your work or regular activities limited in the KIND of work or other activities AS A RESULT OF YOUR PHYSICAL HEALTH?"



	During the PAST 4 WEEKS 4 did you ACCOMPLISHED LESS than you would like AS A RESULT OF YOUR PHYSICAL HEALTH?	During the PAST 4 WEEKS were your work or regular activities limited in the KIND of work or other activities AS A RESULT OF YOUR PHYSICAL HEALTH?
PRE Yes	21%	29%
PRE No	79%	71%
POST Yes	7%	7%
POST No	93%	93%

Figure 39: Physical status 3. Office use case.

The next physical status question can be seen in Figure 40.

- "During the PAST 4 WEEKS, how much did PAIN interfere with your normal work (including both work outside the home and housework)?"

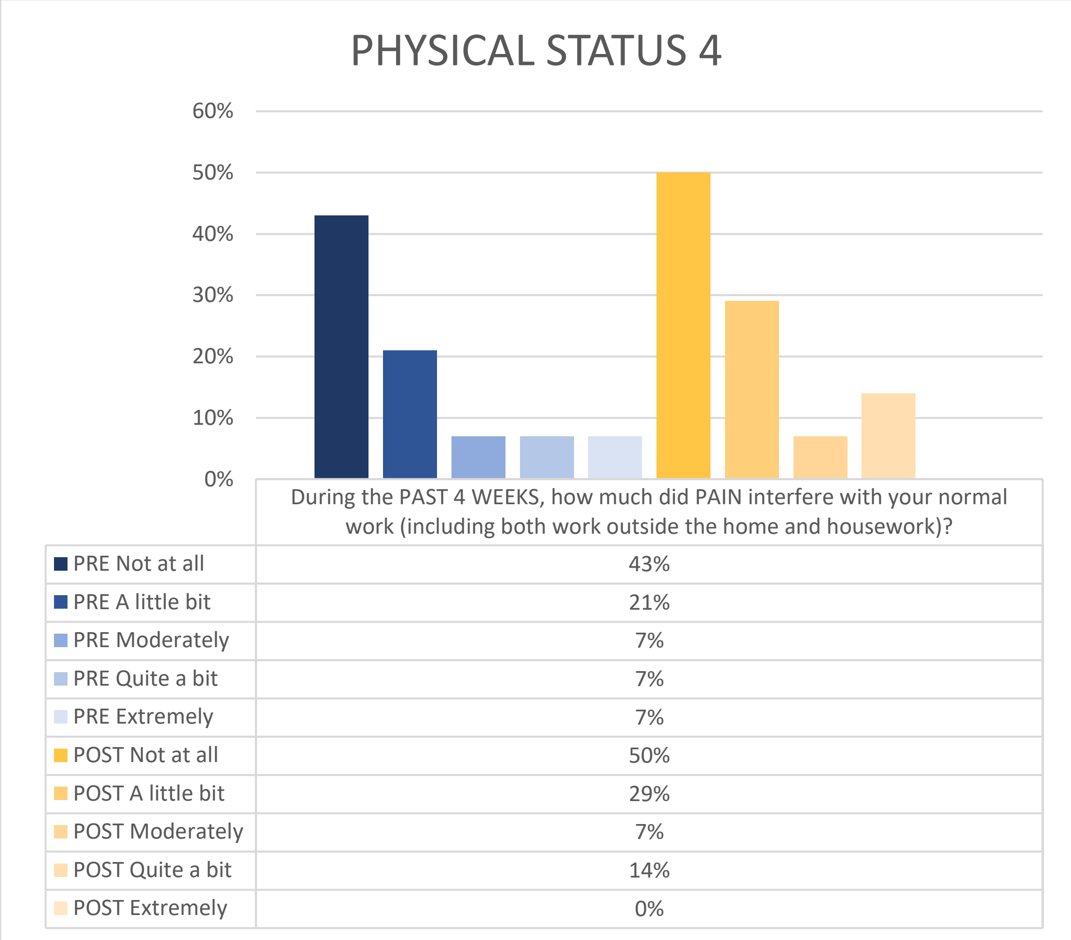
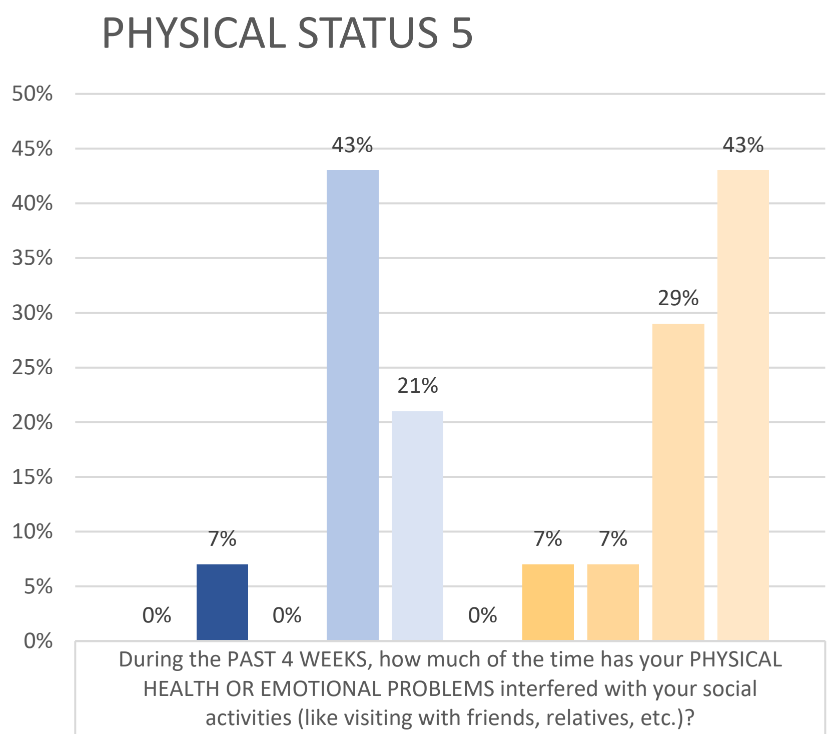


Figure 40: Physical status 4. Office use case.

The last graphic about physical status of the participants is shown below.

- "How much of the time has your PHYSICAL HEALTH OR EMOTIONAL PROBLEMS interfered with your social activities (like visiting with friends, relatives, etc.)?"



	During the PAST 4 WEEKS, how much of the time has your PHYSICAL HEALTH OR EMOTIONAL PROBLEMS interfered with your social activities (like visiting with friends, relatives, etc.)?
PRE All the time	0%
PRE Most of the time	7%
PRE A good bit of the time	0%
PRE A little of the time	43%
PRE None of the time	21%
POST All the time	0%
POST Most of the time	7%
POST A good bit of the time	7%
POST A little of the time	29%
POST None of the time	43%

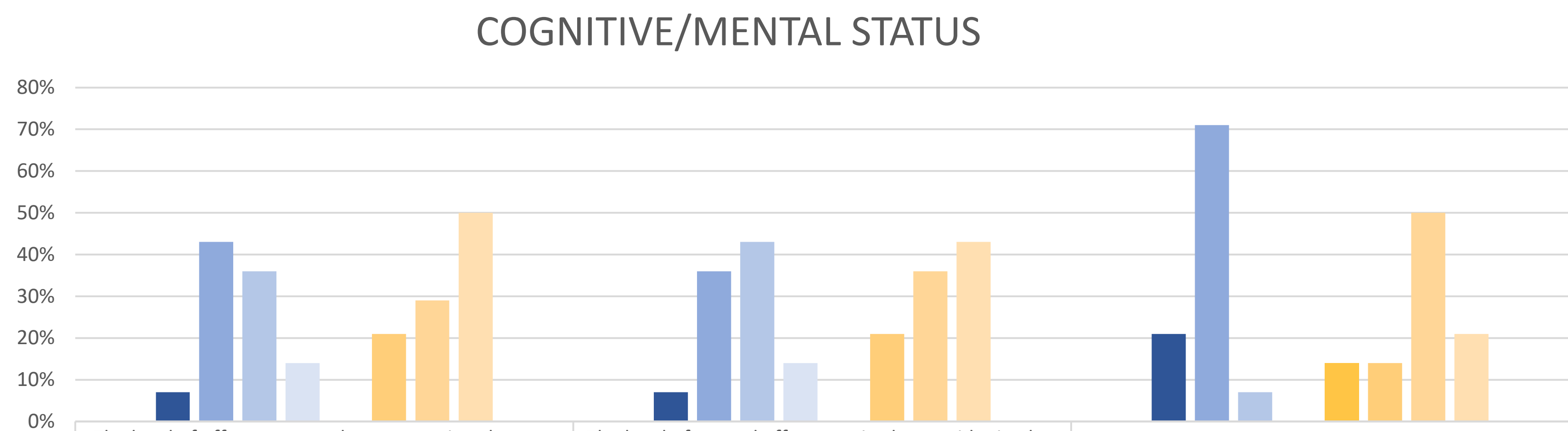
Figure 41: Physical status 5. Office use case.

2.3.3.1.4 Cognitive/Mental status

For the assessment of the cognitive/mental status of the office use case workers, they were asked six questions that address this topic.

The next graphic shows the detailed answers of the participants to the three first questions (Figure 42).

- "The level of effort or mental concentration that my job requires is…"
- "The level of mental effort required to avoid mistakes in my work is…"
- "The fatigue that my work produces is…"

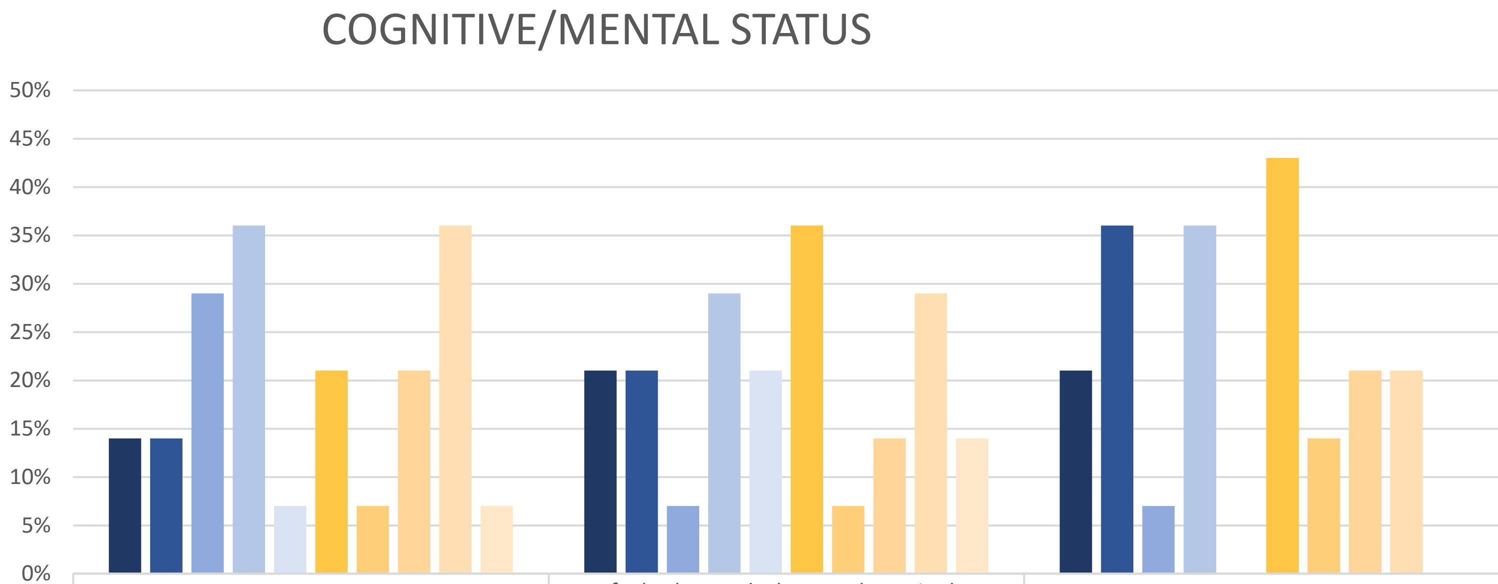


	The level of effort or mental concentration that my job requires is	The level of mental effort required to avoid mistakes in my work is	The fatigue that my work produces is
PRE Very low	0%	0%	0%
PRE Low	7%	7%	21%
PRE Medium	43%	36%	71%
PRE High	36%	43%	7%
PRE Very high	14%	14%	0%
POST Very low	0%	0%	14%
POST Low	21%	21%	14%
POST Medium	29%	36%	50%
POST High	50%	43%	21%
POST Very high	0%	0%	0%

Figure 42: Cognitive/Mental status 1. Office use case.

The next three questions about cognitive/mental status of the volunteers are reflected in the next graph (Figure 43).

- "At the end of the workday I feel exhausted"
- "I feel exhausted when I wake up in the morning and have to face another day of work"
- "I have a hard time relaxing after work"



	At the end of the work day I feel exhausted	I feel exhausted when I wake up in the morning and have to face another day of work	I have a hard time relaxing after work
PRE Strongly disagree	14%	21%	21%
PRE Somewhat disagree	14%	21%	36%
PRE Neither disagree, nor agree	29%	7%	7%
PRE Somewhat agree	36%	29%	36%
PRE Strongly agree	7%	21%	0%
POST Strongly disagree	21%	36%	43%
POST Somewhat disagree	7%	7%	14%
POST Neither disagree, nor agree	21%	14%	21%
POST Somewhat agree	36%	29%	21%
POST Strongly agree	7%	14%	0%

Figure 43: Cognitive/Mental status 2. Office use case.

2.3.3.1.5 *Affective/Emotional status*

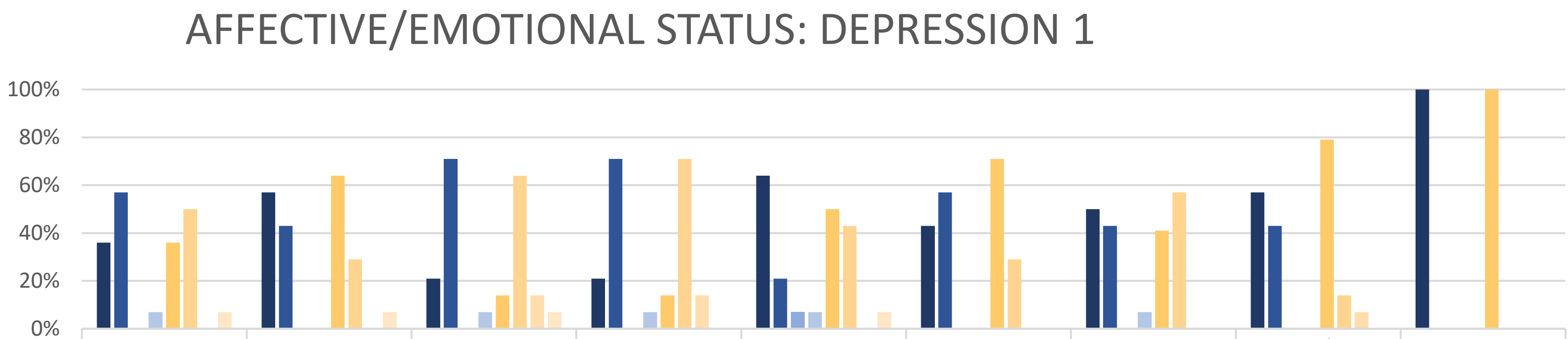
To assess the affective and emotional status of the office participants, ten questions collected users general state with respect to a possible depression. They were asked how often they have been bothered by any of the following problems over the last 2 weeks:

- "Little interest or pleasure in doing things"
- "Feeling down, depressed, or hopeless"
- "Trouble falling or staying asleep, or sleeping too much"
- "Feeling tired or having little energy"
- "Poor appetite or overeating"
- "Feeling bad about yourself — or that you are a failure or have let yourself or your family down"
- "Trouble concentrating on things, such as reading the newspaper or watching television"
- "Moving or speaking so slowly that other people could have noticed. Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual"
- "Thoughts that you would be better off dead or of hurting yourself in some way"

The participants answer to these questions can be seen in the next graph: Depression 1 (Figure 44).

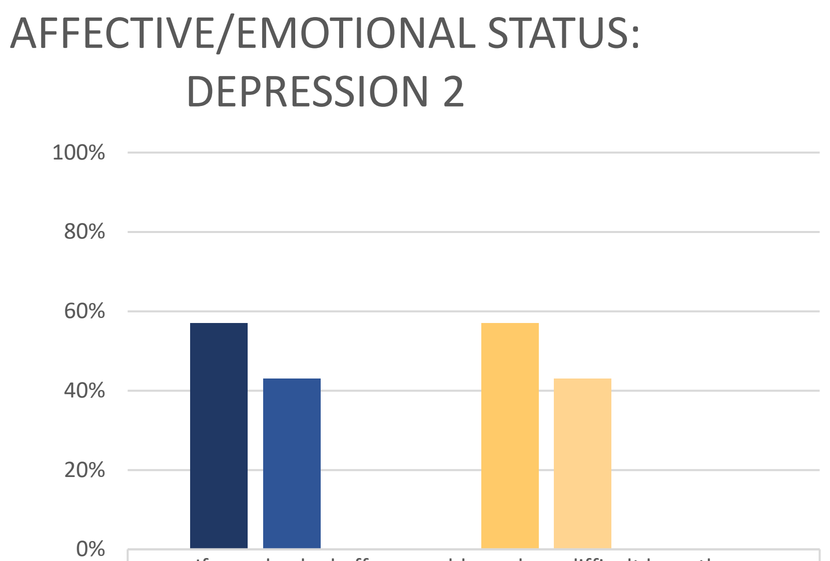
There is one final question reflected in the Affective/Emotional status: Depression 2.

- "If you checked off any problems, how difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?"



	Little interest or pleasure in doing things	Feeling down, depressed, or hopeless	Trouble falling or staying asleep, or sleeping too much	Feeling tired or having little energy	Poor appetite or overeating	Feeling bad about yourself — or that you are a failure or have let yourself or your family down	Trouble concentrating on things, such as reading the newspaper or watching television	Moving or speaking so slowly that other people could have noticed. Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual	Thoughts that you would be better off dead or of hurting yourself in some way
PRE Not at all	36%	57%	21%	21%	64%	43%	50%	57%	100%
PRE Several days	57%	43%	71%	71%	21%	57%	43%	43%	0%
PRE More than half of the days	0%	0%	0%	0%	7%	0%	0%	0%	0%
PRE Nearly every day	7%	0%	7%	7%	7%	0%	7%	0%	0%
POST Not at all	36%	64%	14%	14%	50%	71%	41%	79%	100%
POST Several days	50%	29%	64%	71%	43%	29%	57%	14%	0%
POST More than half of the days	0%	0%	14%	14%	0%	0%	0%	7%	0%
POST Nearly every day	7%	7%	7%	0%	7%	0%	0%	0%	0%

Figure 44: Affective/Emotional status: Depression. Office use case.



	If you checked off any problems, how difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?
PRE Not difficult at all	57%
PRE Somewhat difficult	43%
PRE Very difficult	0%
PRE Extremely difficult	0%
POST Not difficult at all	57%
POST Somewhat difficult	43%
POST Very difficult	0%
POST Extremely difficult	0%

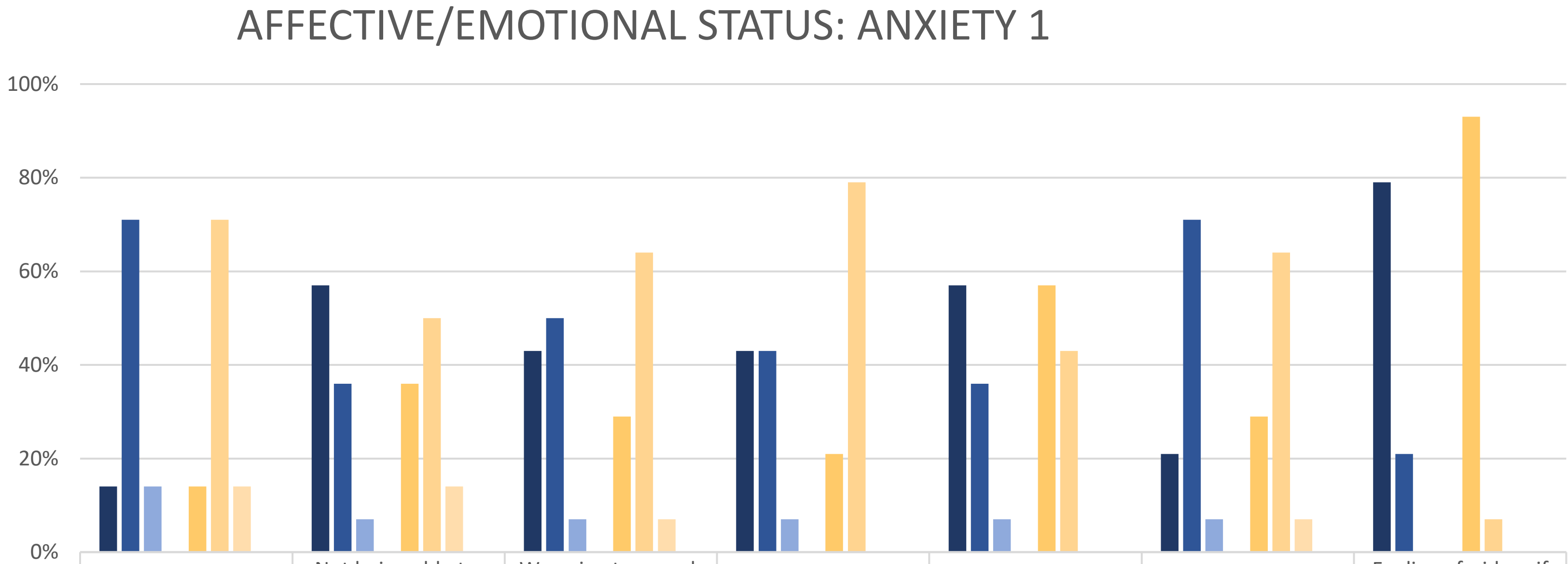
Figure 45: Affective/Emotional status: Depression 2. Office use case.

Next, the participants were asked with eight questions, that collect their general state with respect to a possible anxiety problem. As in the previous section, they were asked if over the last 2 weeks, how often have you been bothered by any of the following problems:

- “Feeling nervous, anxious, or on edge”
- “Not being able to stop or control worrying”
- “Worrying too much about different things”
- “Trouble relaxing”
- “Being so restless that it is hard to sit still”
- “Becoming easily annoyed or irritable”
- “Feeling afraid, as if something awful might happen”

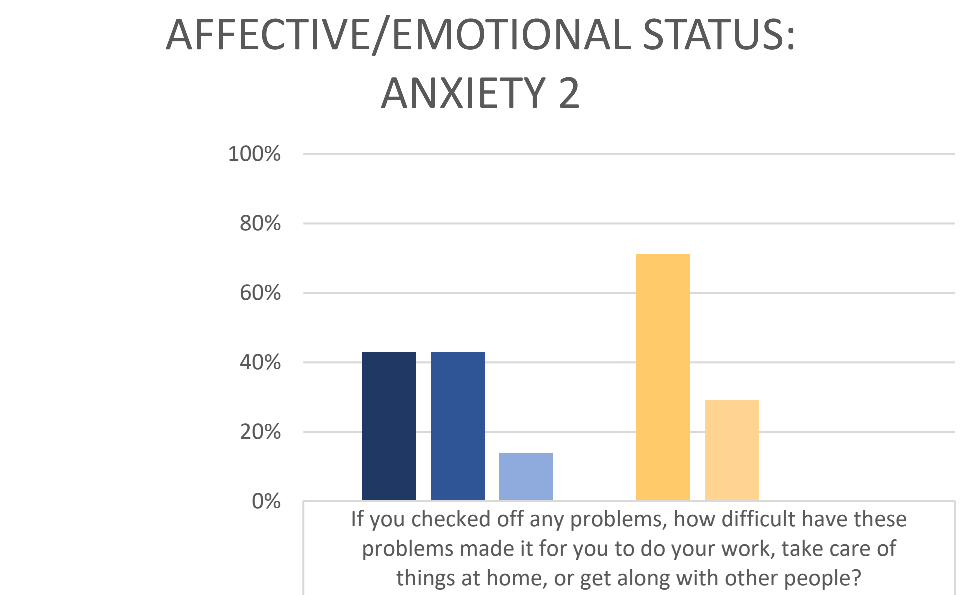
One final question completes this evaluation and is reflected in the Affective/Emotional status: Anxiety 2 (Figure 47).

- “If you checked off any problems, how difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?”



	Feeling nervous, anxious, or on edge	Not being able to stop or control worrying	Worrying too much about different things	Trouble relaxing	Being so restless that it is hard to sit still	Becoming easily annoyed or irritable	Feeling afraid, as if something awful might happen
PRE Not at all	14%	57%	43%	43%	57%	21%	79%
PRE Several days	71%	36%	50%	43%	36%	71%	21%
PRE More than half of the days	14%	7%	7%	7%	7%	7%	0%
PRE Nearly every day	0%	0%	0%	0%	0%	0%	0%
POST Not at all	14%	36%	29%	21%	57%	29%	93%
POST Several days	71%	50%	64%	79%	43%	64%	7%
POST More than half of the days	14%	14%	7%	0%	0%	7%	0%
POST Nearly every day	0%	0%	0%	0%	0%	0%	0%

Figure 46: Affective/Emotional status: Anxiety 1. Office use case.



	If you checked off any problems, how difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?
PRE Not difficult at all	43%
PRE Somewhat difficult	43%
PRE Very difficult	14%
PRE Extremely difficult	0%
POST Not difficult at all	71%
POST Somewhat difficult	29%
POST Very difficult	0%
POST Extremely difficult	0%

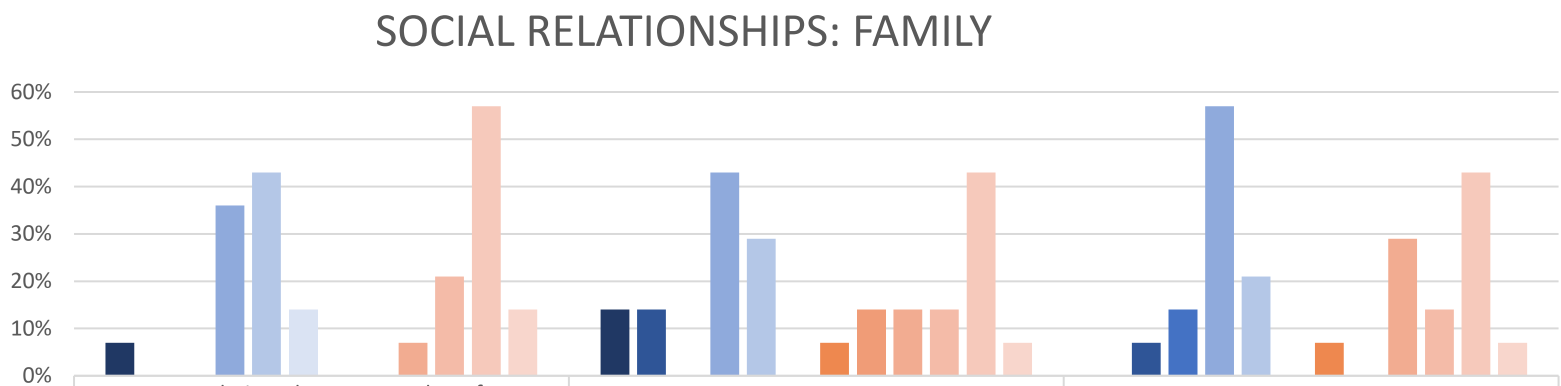
Figure 47: Affective/Emotional status: Anxiety 2. Office use case.

2.3.3.1.6 Social relationships

For the evaluation of the social relationships of the participants, they were asked first about their family and later about their friends.

The next graph shows how is the relationship between the participants and their relatives, with three questions (Figure 48).

- "How many relatives do you see or hear from at least once a month?"
- "How many relatives do you feel at ease with that you can talk about private matters?"
- "How many relatives do you feel close to such that you could call on them for help?"

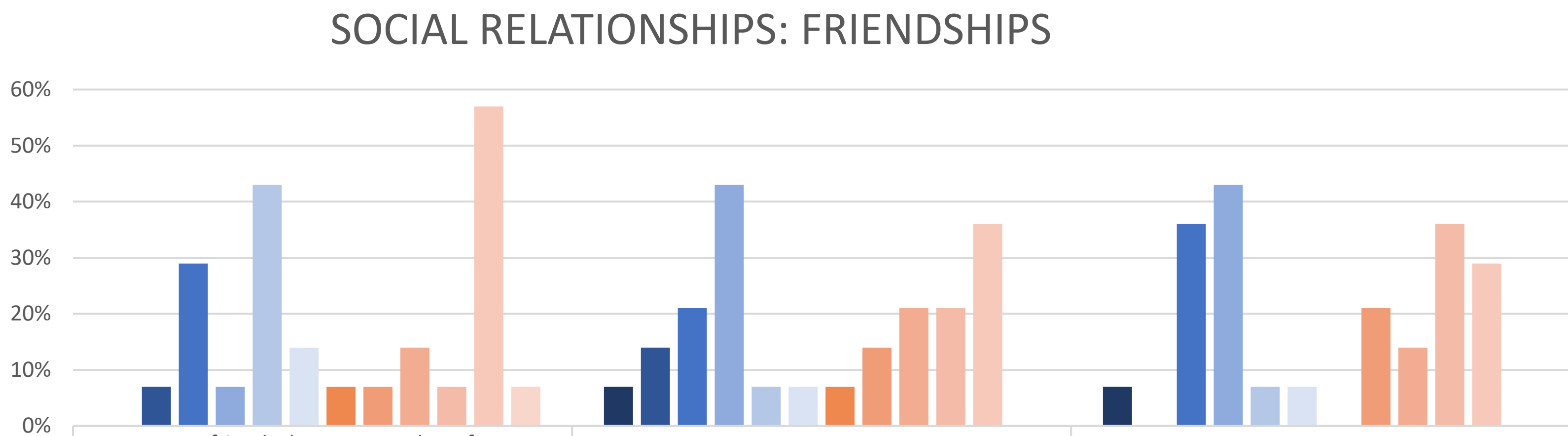


	How many relatives do you see or hear from at least once a month?The level of effort or mental concentration that my job requires is	How many relatives do you feel at ease with that you can talk about private matters?	How many relatives do you feel close to such that you could call on them for help?
PRE NONE	7%	14%	0%
PRE ONE	0%	14%	7%
PRE TWO	0%	0%	14%
PRE THREE OR FOUR	36%	43%	57%
PRE FIVE TO EIGHT	43%	29%	21%
PRE NINE OR MORE	14%	0%	0%
POST NONE	0%	7%	7%
POST ONE	0%	14%	0%
POST TWO	7%	14%	29%
POST THREE OR FOUR	21%	14%	14%
POST FIVE TO EIGHT	57%	43%	43%
POST NINE OR MORE	14%	7%	7%

Figure 48: Social relationships. Family. Office use case.

The next graph shows how is the relationship between the participants and their friends, with the same three questions used for their relatives (Figure 49).

- "How many friends do you see or hear from at least once a month?"
- "How many friends do you feel at ease with that you can talk about private matters?"
- "How many friends do you feel close to such that you could call on them for help?"



	How many friends do you see or hear from at least once a month?The level of effort or mental concentration that my job requires is	How many friends do you feel at ease with that you can talk about private matters?	How many friends do you feel close to such that you could call on them for help?
PRE NONE	0%	7%	7%
PRE ONE	7%	14%	0%
PRE TWO	29%	21%	36%
PRE THREE OR FOUR	7%	43%	43%
PRE FIVE TO EIGHT	43%	7%	7%
PRE NINE OR MORE	14%	7%	7%
POST NONE	7%	7%	0%
POST ONE	7%	14%	21%
POST TWO	14%	21%	14%
POST THREE OR FOUR	7%	21%	36%
POST FIVE TO EIGHT	57%	36%	29%
POST NINE OR MORE	7%	0%	0%

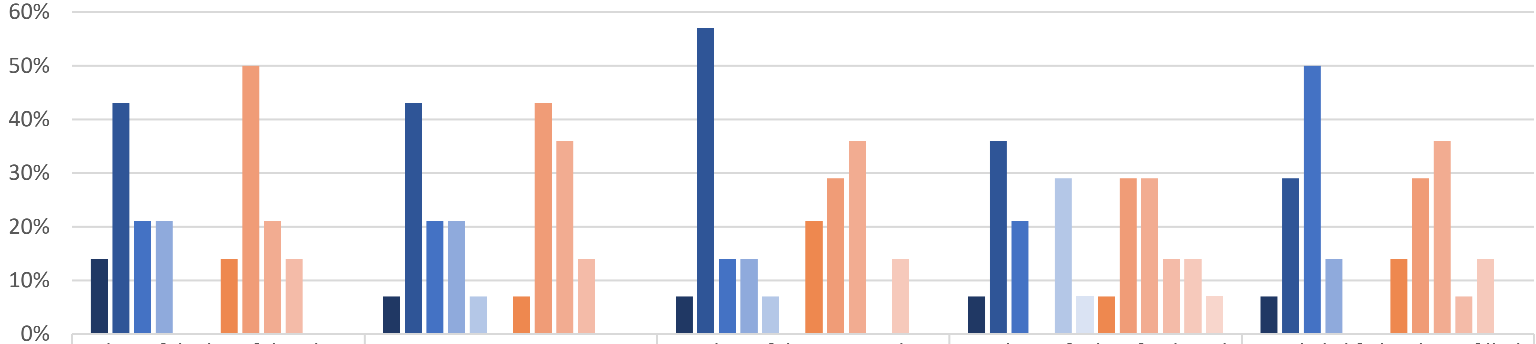
Figure 49: Social relationships. Friendships. Office use case.

2.3.3.1.7 *General wellbeing & Quality of life*

The volunteers were asked about their subjective wellbeing and quality of life, firstly with five questions that can be seen in detail in the next graph.

- “I have felt cheerful and in good spirits”
- “I have felt calm and relaxed”
- “I have felt active and vigorous”
- “I woke up feeling fresh and rested”
- “My daily life has been filled with things that interest me”

WELLBEING

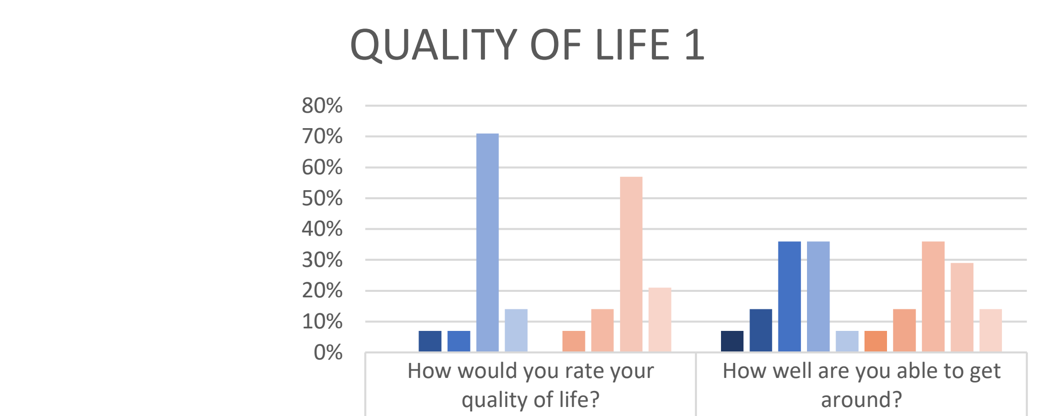


	I have felt cheerful and in good spirits	I have felt calm and relaxed	I have felt active and vigorous	I woke up feeling fresh and rested	My daily life has been filled with things that interest me
PRE All the time	14%	7%	7%	7%	7%
PRE Most of the time	43%	43%	57%	36%	29%
PRE More than half the time	21%	21%	14%	21%	50%
PRE Less than half the time	21%	21%	14%	0%	14%
PRE Some of the time	0%	7%	7%	29%	0%
PRE At no time	0%	0%	0%	7%	0%
POST All the time	14%	7%	21%	7%	14%
POST Most of the time	50%	43%	29%	29%	29%
POST More than half the time	21%	36%	36%	29%	36%
POST Less than half the time	14%	14%	0%	14%	7%
POST Some of the time	0%	0%	14%	14%	14%
POST At no time	0%	0%	0%	7%	0%

Figure 50: Wellbeing. Office use case

The next questions assess the quality of life of the participants and can be seen in detail in the next six graphs.

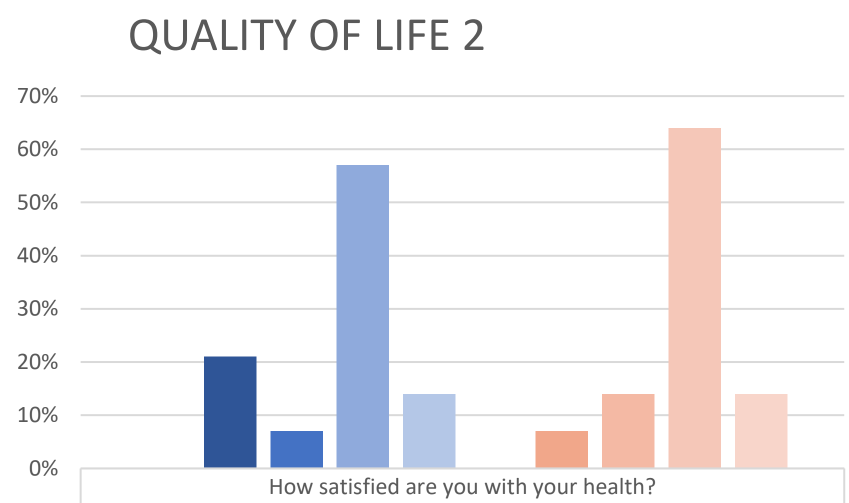
- "How would you rate your quality of life?"
- "How well are you able to get around?"



	How would you rate your quality of life?	How well are you able to get around?
PRE Very poor	0%	7%
PRE Poor	7%	14%
PRE Neither poor nor good	7%	36%
PRE Good	71%	36%
PRE Very good	14%	7%
POST Very poor	0%	7%
POST Poor	7%	14%
POST Neither poor nor good	14%	36%
POST Good	57%	29%
POST Very good	21%	14%

Figure 51: Quality of life 1. Office use case.

- "How satisfied are you with your health?"

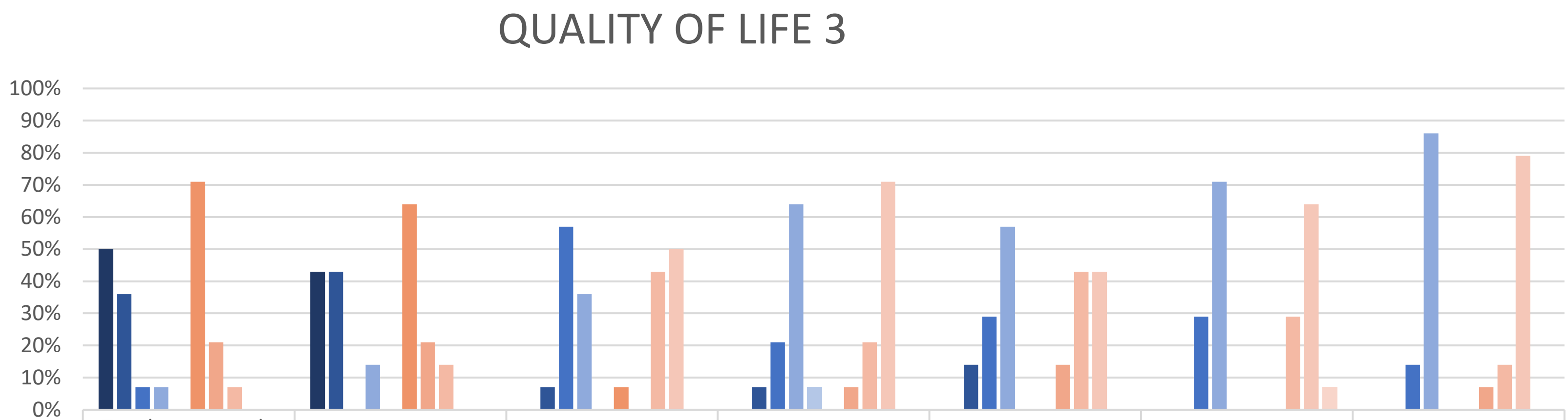


	How satisfied are you with your health?
PRE Very dissatisfied	0%
PRE Dissatisfied	21%
PRE Neither satisfied nor dissatisfied	7%
PRE Satisfied	57%
PRE Very satisfied	14%
POST Very dissatisfied	0%
POST Dissatisfied	7%
POST Neither satisfied nor dissatisfied	14%
POST Satisfied	64%
POST Very satisfied	14%

Figure 52: Quality of life 2. Office use case.

Next, you will see seven more questions that assess the quality of life of the WAOW tool users. The details are shown in the following image.

- "To what extent do you feel that physical pain prevents you from doing what you need to do?"
- "How much do you need any medical treatment to function in your daily life?"
- "How much do you enjoy life?"
- "To what extent do you feel your life to be meaningful?"
- "How well are you able to concentrate?"
- "How safe do you feel in your daily life?"
- "How healthy is your physical environment?"

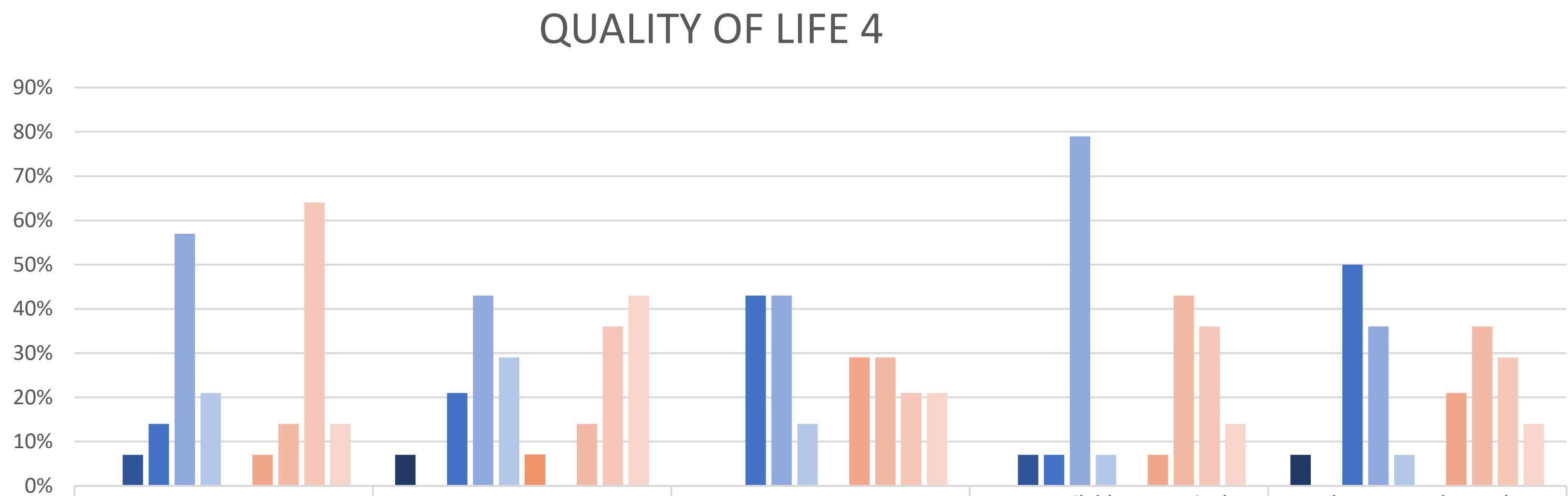


	To what extent do you feel that physical pain prevents you from doing what you need to do?	How much do you need any medical treatment to function in your daily life?	How much do you enjoy life?	To what extent do you feel your life to be meaningful?	How well are you able to concentrate?	How safe do you feel in your daily life?	How healthy is your physical environment?
PRE Not at all	50%	43%	0%	0%	0%	0%	0%
PRE A little	36%	43%	7%	7%	14%	0%	0%
PRE A moderate amount	7%	0%	57%	21%	29%	29%	14%
PRE Very much	7%	14%	36%	64%	57%	71%	86%
PRE An extreme amount	0%	0%	0%	7%	0%	0%	0%
POST Not at all	71%	64%	7%	0%	0%	0%	0%
POST A little	21%	21%	0%	7%	14%	0%	7%
POST A moderate amount	7%	14%	43%	21%	43%	29%	14%
POST Very much	0%	0%	50%	71%	43%	64%	79%
POST An extreme amount	0%	0%	0%	0%	0%	7%	0%

Figure 53: Quality of life 3. Office use case.

Another five questions that evaluate the quality of life are shown in the next graph: “Quality of life 4”.

- “Do you have enough energy for everyday life?”
- “Are you able to accept your bodily appearance?”
- “Have you enough money to meet your needs?”
- “How available to you is the information that you need in your day-to-day life?”
- “To what extent do you have the opportunity for leisure activities?”

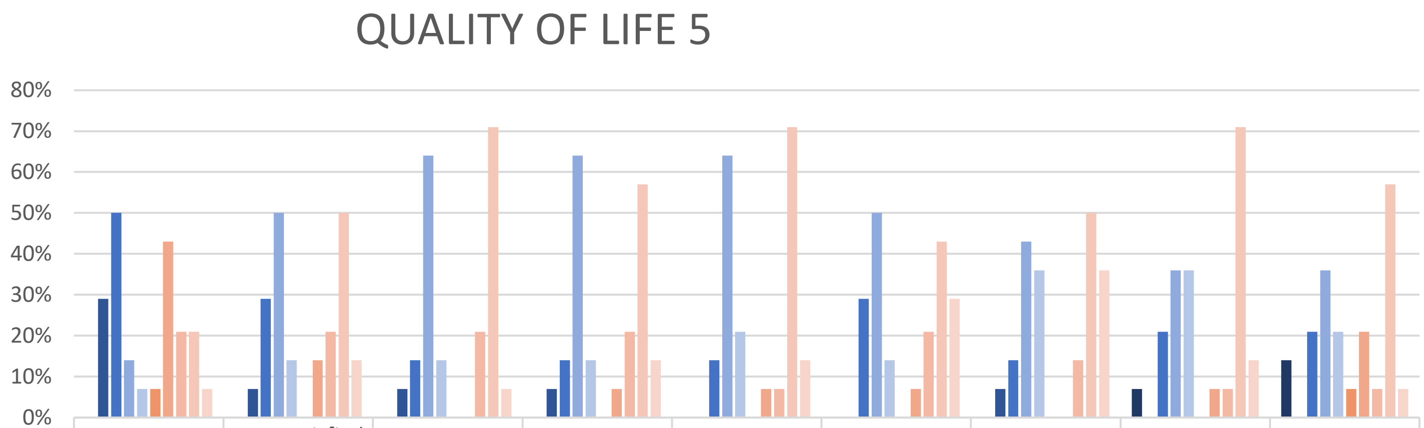


	Do you have enough energy for everyday life?	Are you able to accept your bodily appearance?	Have you enough money to meet your needs?	How available to you is the information that you need in your day-to-day life?	To what extent do you have the opportunity for leisure activities?
PRE Not at all	0%	7%	0%	0%	7%
PRE A little	7%	0%	0%	7%	0%
PRE Moderately	14%	21%	43%	7%	50%
PRE Mostly	57%	43%	43%	79%	36%
PRE Completely	21%	29%	14%	7%	7%
POST Not at all	0%	7%	0%	0%	0%
POST A little	7%	0%	29%	7%	21%
POST Moderately	14%	14%	29%	43%	36%
POST Mostly	64%	36%	21%	36%	29%
POST Completely	14%	43%	21%	14%	14%

Figure 54: Quality of life 4. Office use case.

Continuing assessing the quality of life, the participants answered the next nine questions. The details can be seen in the next graph: “Quality of life 5”.

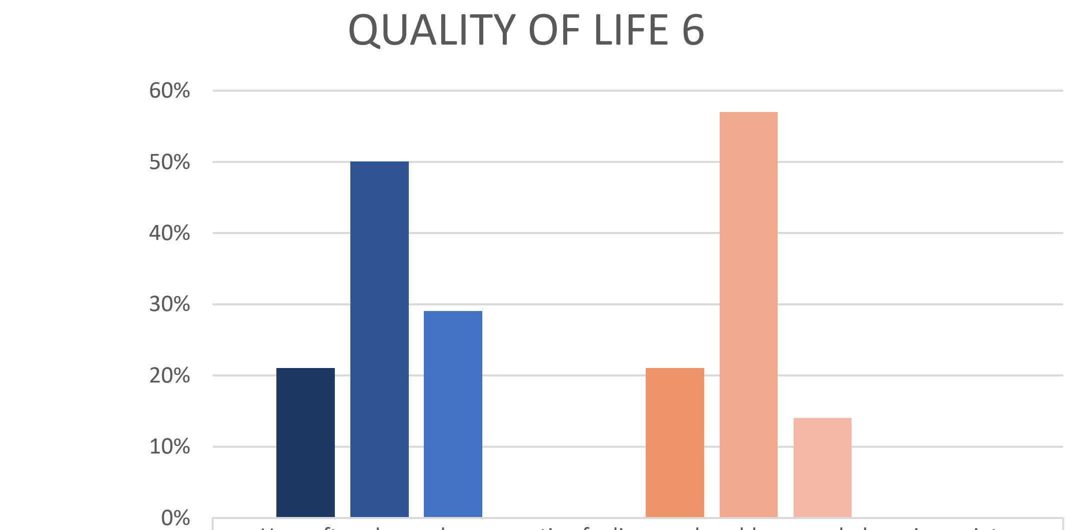
- “How satisfied are you with your sleep?”
- “How satisfied are you with your ability to perform your daily living activities?”
- “How satisfied are you with your capacity for work?”
- “How satisfied are you with yourself?”
- “How satisfied are you with your personal relationships?”
- “How satisfied are you with the support you get from your friends?”
- “How satisfied are you with the conditions of your living place?”
- “How satisfied are you with your access to health services?”
- “How satisfied are you with your transport?”



	How satisfied are you with your sleep?	How satisfied are you with your ability to perform your daily living activities?	How satisfied are you with your capacity for work?	How satisfied are you with yourself?	How satisfied are you with your personal relationships?	How satisfied are you with the support you get from your friends?	How satisfied are you with the conditions of your living place?	How satisfied are you with your access to health services?	How satisfied are you with your transport?
PRE Very dissatisfied	0%	0%	0%	0%	0%	0%	0%	7%	14%
PRE Dissatisfied	29%	7%	7%	7%	0%	0%	7%	0%	0%
PRE Neither satisfied nor dissatisfied	50%	29%	14%	14%	14%	29%	14%	21%	21%
PRE Satisfied	14%	50%	64%	64%	64%	50%	43%	36%	36%
PRE Very satisfied	7%	14%	14%	14%	21%	14%	36%	36%	21%
POST Very dissatisfied	7%	0%	0%	0%	0%	0%	0%	0%	7%
POST Dissatisfied	43%	14%	0%	7%	7%	7%	0%	7%	21%
POST Neither satisfied nor dissatisfied	21%	21%	21%	21%	7%	21%	14%	7%	7%
POST Satisfied	21%	50%	71%	57%	71%	43%	50%	71%	57%
POST Very satisfied	7%	14%	7%	14%	14%	29%	36%	14%	7%

Figure 55: Quality of life 5. Office use case.

The last question to assess the quality of life of the participants was "How often do you have negative feelings such as blue mood, despair, anxiety, depression?" The participants selected the following frequency:



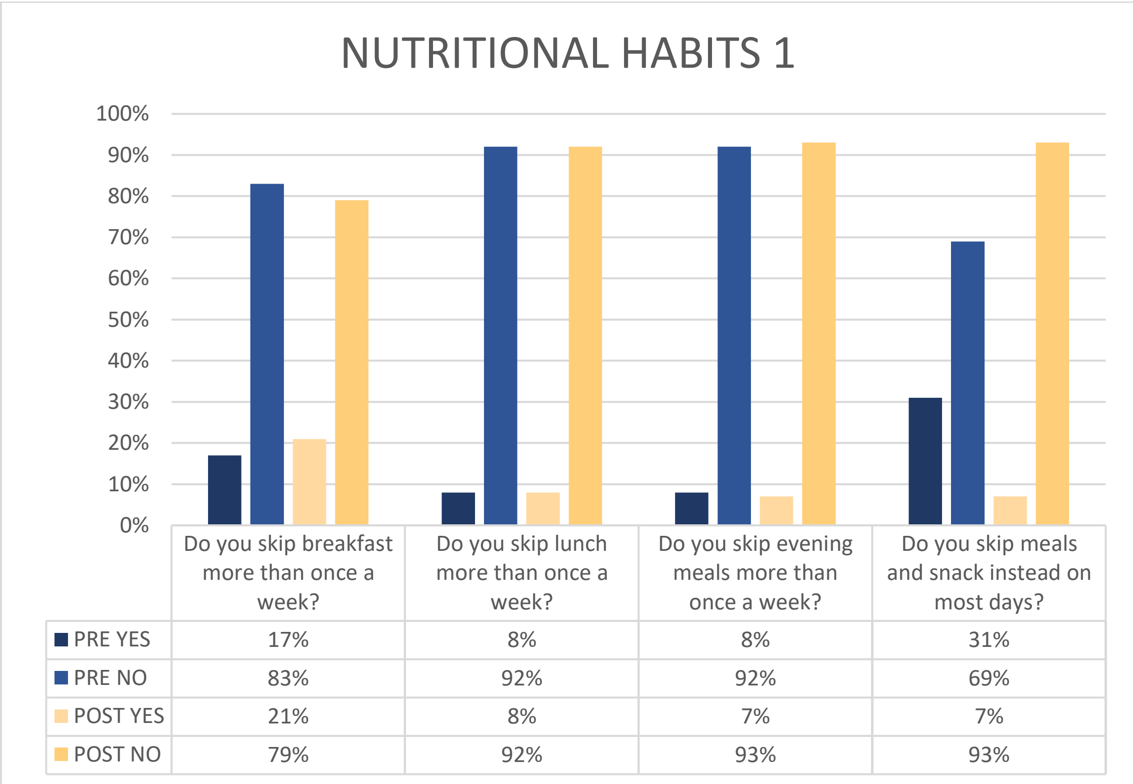
	How often do you have negative feelings such as blue mood, despair, anxiety, depression?
PRE Never	21%
PRE Seldom	50%
PRE Quite often	29%
PRE Very often	0%
PRE Always	0%
POST Never	21%
POST Seldom	57%
POST Quite often	14%
POST Very often	0%
POST Always	0%

Figure 56: Quality of life 6. Office use case.

For assessing the nutritional habits, very important in the general wellbeing of the participants, they were asked 30 questions that they had to answer with a simple yes or no.

The first four questions are related with the frequency they skip different meals during the week. A positive answer will mean a bad nutritional habit, while a negative will mean a good one. See the graph "Nutritional habits 1" below.

- "Do you skip breakfast more than once a week?"
- "Do you skip lunch more than once a week?"
- "Do you skip evening meals more than once a week?"
- "Do you skip meals and snack instead on most days?"

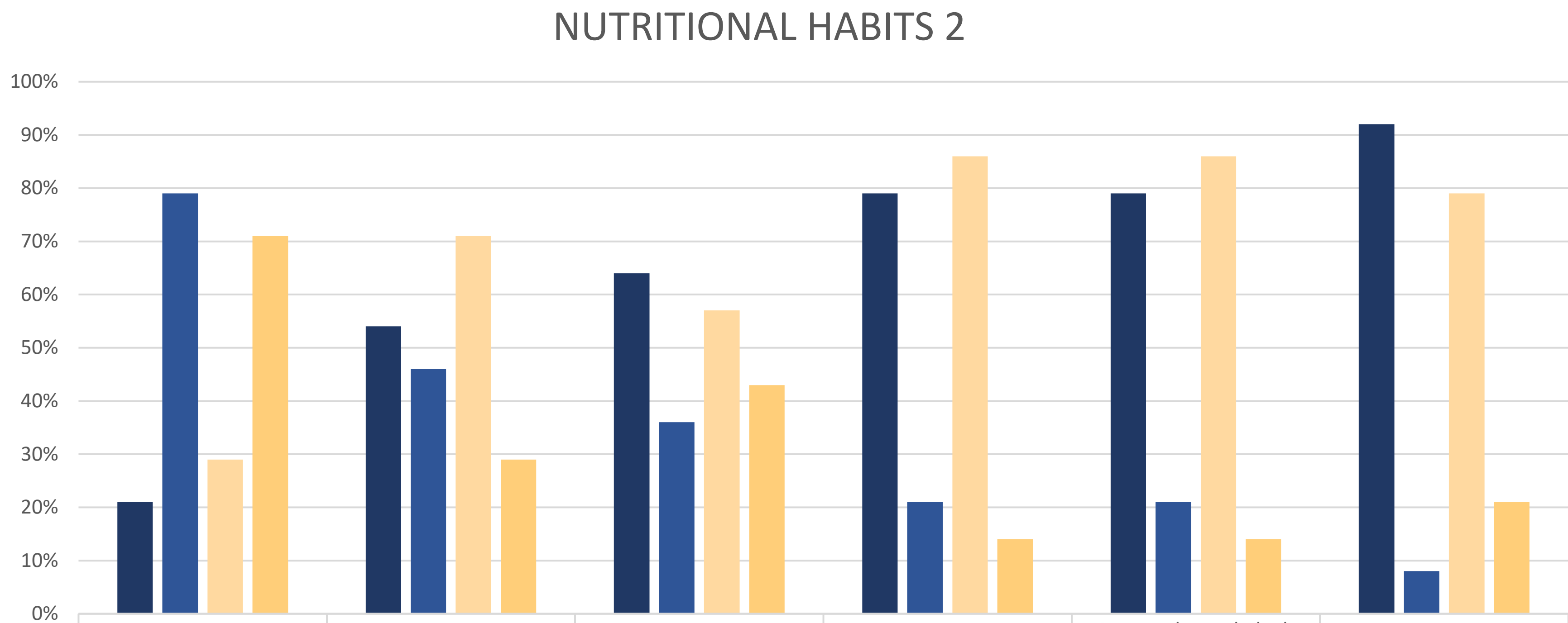


	Do you skip breakfast more than once a week?	Do you skip lunch more than once a week?	Do you skip evening meals more than once a week?	Do you skip meals and snack instead on most days?
PRE YES	17%	8%	8%	31%
PRE NO	83%	92%	92%	69%
POST YES	21%	8%	7%	7%
POST NO	79%	92%	93%	93%

Figure 57: Nutritional habits 1. Office use case

The next six questions are about the choice of the volunteers when they select their food or ingredients. A positive answer will mean a good nutritional habit, while a negative will mean a bad one. See the graph "Nutritional habits 2" below.

- "Do you eat more than 5 portions of fruit and/or vegetables every day?"
- "Do you eat more than 4 different varieties of fruit each week?"
- "Do you eat more than 4 different varieties of vegetables each week?"
- "Do you choose low-fat products when available?"
- "Do you choose baked, steamed, or grilled options when available, rather than fried foods (such as crisps and snacks, or fish and chips)?"
- "Do you opt for lean cuts of meat or remove visible fat – for example, removing the skin on chicken or the rind on bacon?"

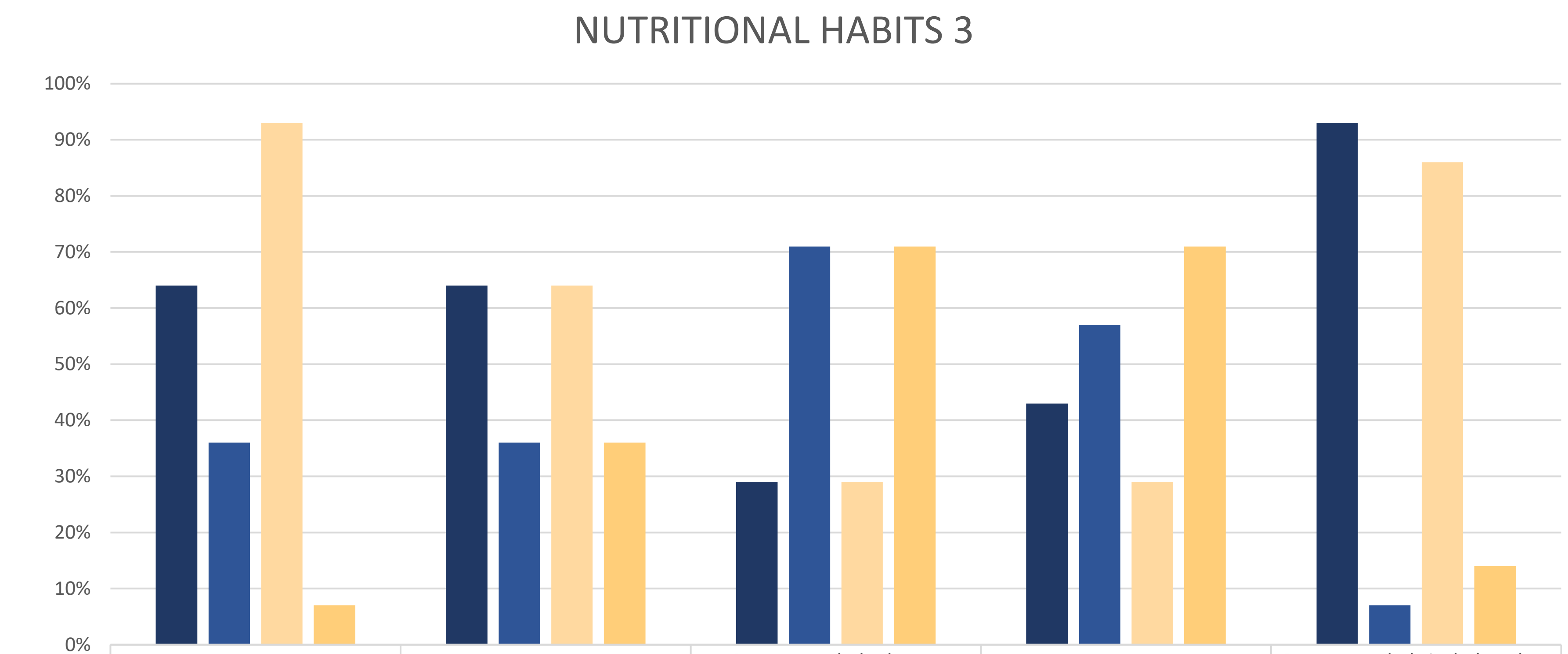


	Do you eat more than 5 portions of fruit and/or vegetables every day?	Do you eat more than 4 different varieties of fruit each week?	Do you eat more than 4 different varieties of vegetables each week?	Do you choose low-fat products when available?	Do you choose baked, steamed or grilled options when available, rather than fried foods (such as crisps and snacks, or fish and chips)?	Do you opt for lean cuts of meat or remove visible fat – for example, removing the skin on chicken or the rind on bacon?
PRE YES	21%	54%	64%	79%	79%	92%
PRE NO	79%	46%	36%	21%	21%	8%
POST YES	29%	71%	57%	86%	86%	79%
POST NO	71%	29%	43%	14%	14%	21%

Figure 58: Nutritional habits 2. Office use case

The next five questions are about the diet habits of the volunteers. A positive answer will mean a good nutritional habit, while a negative will mean a bad one. The details are shown in the graph "Nutritional habits 3" below.

- "Did you eat any oily fish last week?"
- "Do you include some unsalted nuts and seeds in your diet?"
- "Do you regularly choose wholemeal bread or rolls rather than white?"
- "Do you regularly eat wholegrain cereals, with no added sugar?"
- "Do you regularly include pulses in your diet? For example, beans and lentils"

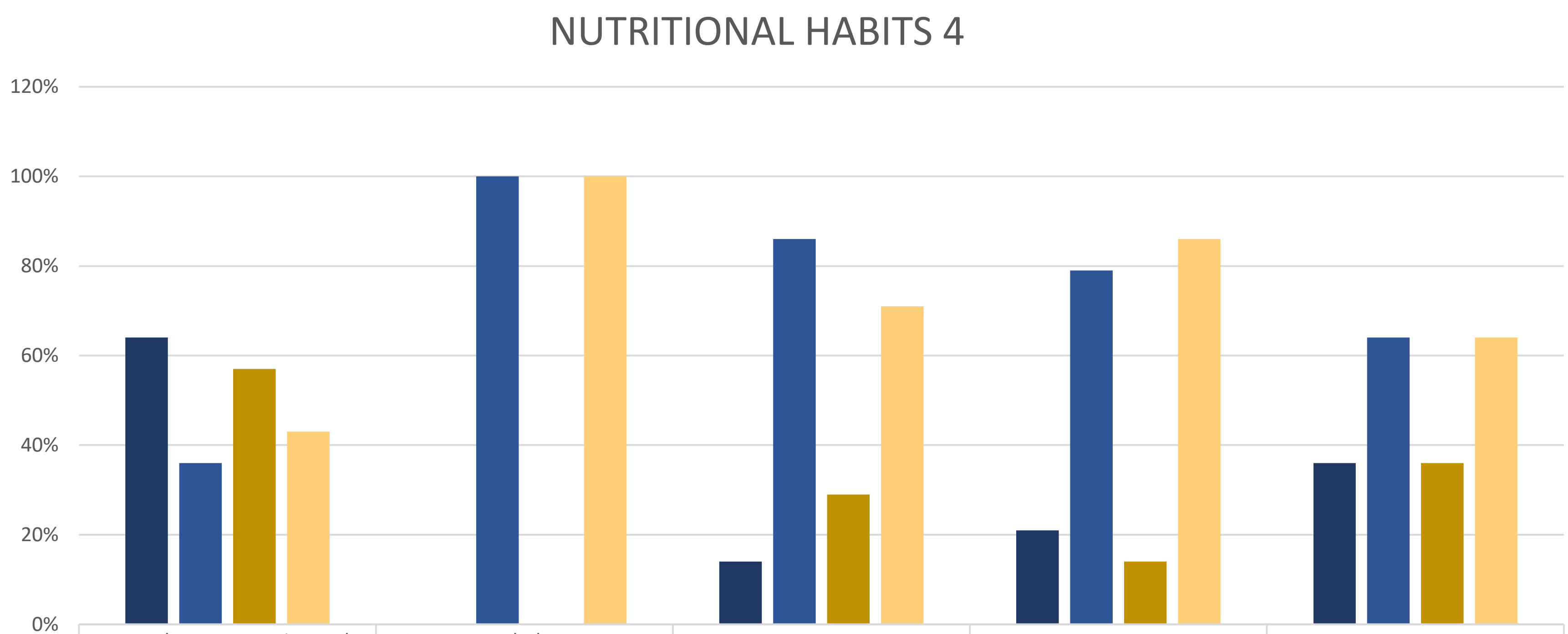


	Did you eat any oily fish last week?	Do you include some unsalted nuts and seeds in your diet?	Do you regularly choose wholemeal bread or rolls rather than white?	Do you regularly eat wholegrain cereals, with no added sugar?	Do you regularly include pulses in your diet? For example, beans and lentils.
PRE YES	64%	64%	29%	43%	93%
PRE NO	36%	36%	71%	57%	7%
POST YES	93%	64%	29%	29%	86%
POST NO	7%	36%	71%	71%	14%

Figure 59: Nutritional habits 3. Office use case

The next five questions are about starchy foods, sugar, and sweets. A positive answer will mean a bad nutritional habit, while a negative will mean a good one. The details are shown in the graph "Nutritional habits 4" below.

- "Do you base your main meals around starchy foods? For example, potatoes, pasta, rice, or bread"
- "Do you regularly eat sugar-coated breakfast cereals or add sugar to your breakfast cereals?"
- "Do you add sugar to your drinks?"
- "Do you regularly drink sweet fizzy drinks?"
- "Do you regularly eat cakes, sweets, chocolate, or biscuits at work?"

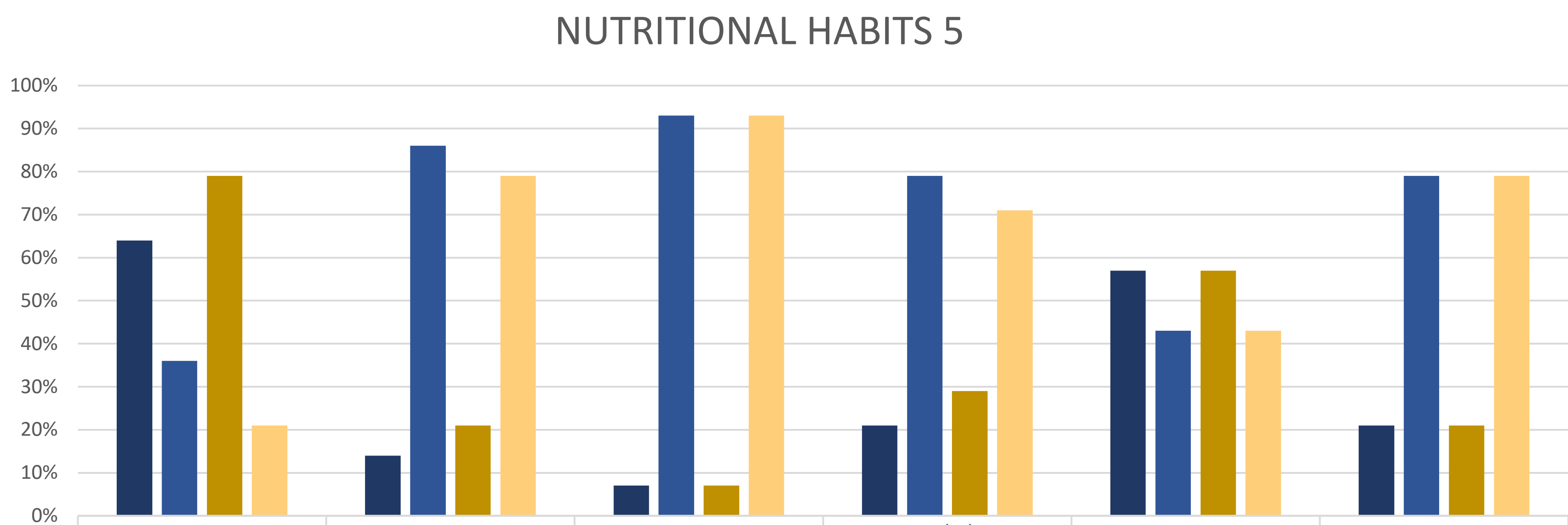


	Do you base your main meals around starchy foods? For example, potatoes, pasta, rice or bread.	Do you regularly eat sugar-coated breakfast cereals or add sugar to your breakfast cereals?	Do you add sugar to your drinks?	Do you regularly drink sweet fizzy drinks?	Do you regularly eat cakes, sweets, chocolate or biscuits at work?
PRE YES	64%	0%	14%	21%	36%
PRE NO	36%	100%	86%	79%	64%
POST YES	57%	0%	29%	14%	36%
POST NO	43%	100%	71%	86%	64%

Figure 60: Nutritional habits 4. Office use case

The next six questions are about bad nutritional habits and high blood pressure. A positive answer will mean a bad nutritional habit, while a negative will mean a good one. The details are shown in the graph "Nutritional habits 5" below.

- "Do you regularly add salt to food during cooking?"
- "Do you regularly add salt to meals at the table?"
- "Do you regularly eat savoury snacks at work? For example, crisps or salted nuts?"
- "Do you regularly eat pre-prepared meals? For example, pre-prepared sandwiches, ready meals or canned soups?"
- "Do you regularly eat processed meats such as ham or bacon, or smoked fish?"
- "Has your GP advised you that you have high blood pressure?"



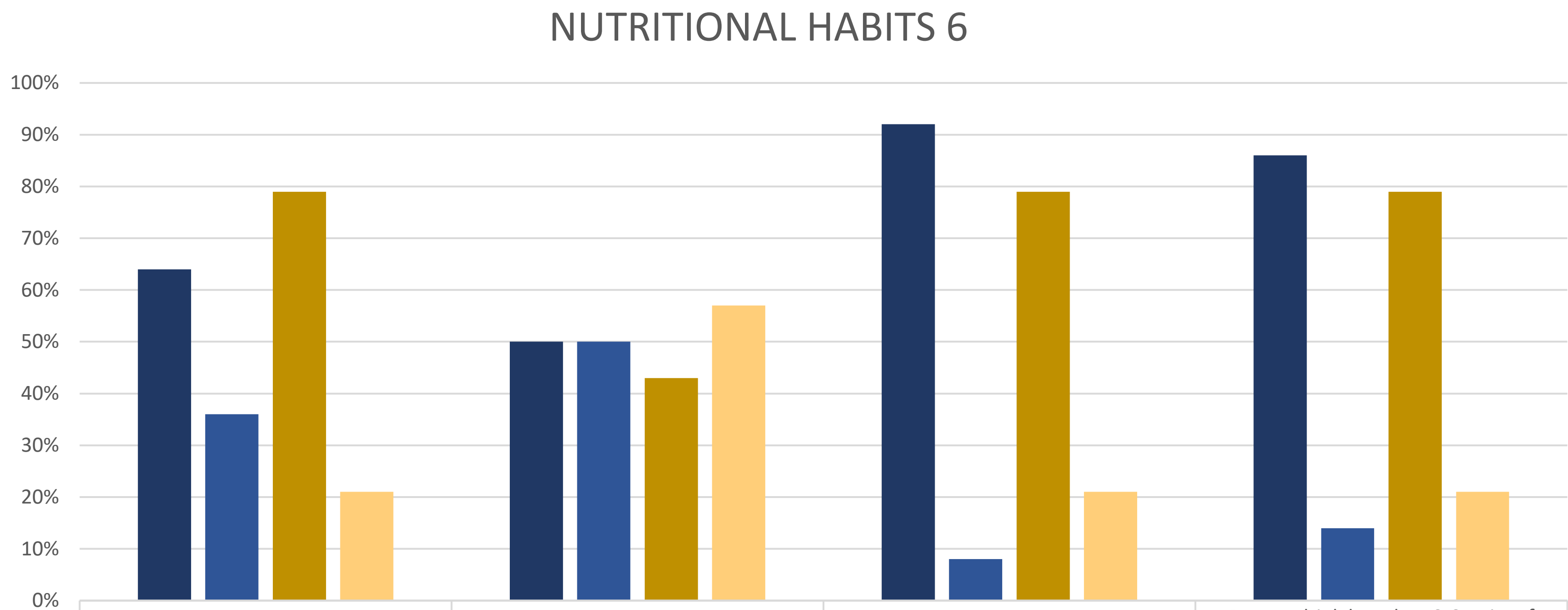
	Do you regularly add salt to food during cooking?	Do you regularly add salt to meals at the table?	Do you regularly eat savoury snacks at work? For example, crisps or salted nuts.	Do you regularly eat pre-prepared meals? For example, pre-prepared sandwiches, ready meals or canned soups.	Do you regularly eat processed meats such as ham or bacon, or smoked fish?	Has your GP advised you that you have high blood pressure?
PRE YES	64%	14%	7%	21%	57%	21%
PRE NO	36%	86%	93%	79%	43%	79%
POST YES	79%	21%	7%	29%	57%	21%
POST NO	21%	79%	93%	71%	43%	79%

PRE YES PRE NO POST YES POST NO

Figure 61: Nutritional habits 5. Office use case

The last four questions are about drinks choice and drinking frequency. A positive answer will mean a good nutritional habit, while a negative will mean a bad one. The details are shown in the graph "Nutritional habits 6" below.

- "Do you drink plenty of fluids at regular intervals during the working day?"
- "Do you opt for a variety of different drinks, including water, at work?"
- "Do you avoid sugary fizzy drinks?"
- "Do you drink less than 2-3 units of alcohol a day if you're a woman, or less than 3-4 units of alcohol a day if you're a man? 1 unit of alcohol is equivalent to 100ml of 10% ABV (alcohol by volume)"



	Do you drink plenty of fluids at regular intervals during the working day?	Do you opt for a variety of different drinks, including water, at work?	Do you avoid sugary fizzy drinks?	Do you drink less than 2-3 units of alcohol a day if you're a woman, or less than 3-4 units of alcohol a day if you're a man? 1 unit of alcohol is equivalent to 100ml of 10% ABV (alcohol by volume)
PRE YES	64%	50%	92%	86%
PRE NO	36%	50%	8%	14%
POST YES	79%	43%	79%	79%
POST NO	21%	57%	21%	21%

Figure 62: Nutritional habits 6. Office use case

2.3.3.1.8 *Health-related physical activity*

To finish evaluating the health of the participants with respect to their physical activity, we have asked them 7 questions, about vigorous and moderate physical activities, walking and sitting time.

We consider them one by one in detail and with the corresponding graph.

- "During the last 7 days, on how many days did you do vigorous physical activities like heavy lifting, digging, aerobics, or fast bicycling?"

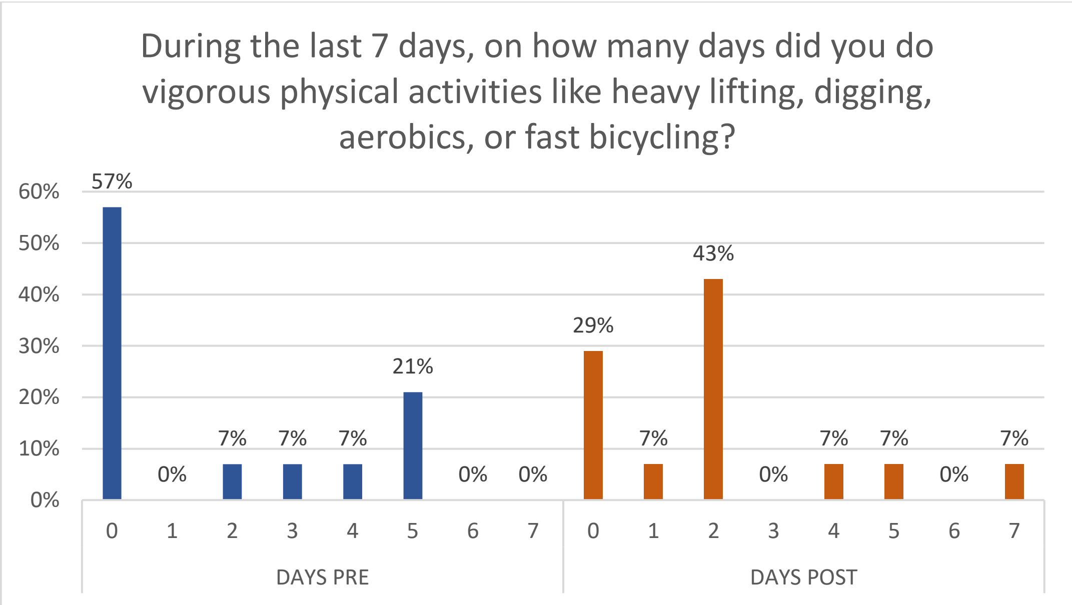


Figure 63: Health related vigorous physical activity 1. Office use case

- For those participants who answered yes to the previous question, we made an additional question: "How much time did you usually spend doing vigorous physical activities on one of those days?"

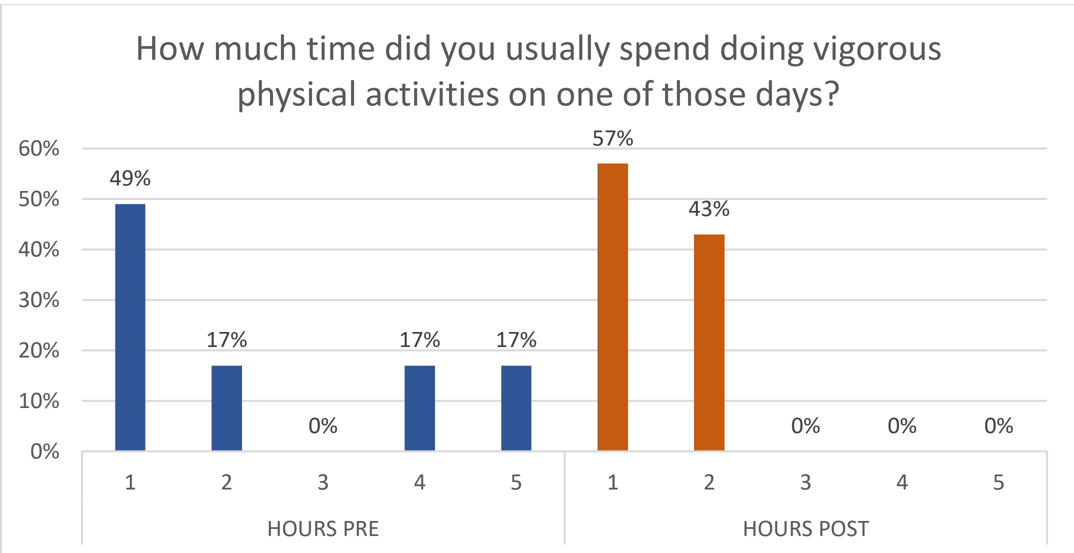


Figure 64: Health related vigorous physical activity 2. Office use case

- "During the last 7 days, on how many days did you do moderate physical activities like carrying light loads, bicycling at a regular pace, or doubles tennis?"

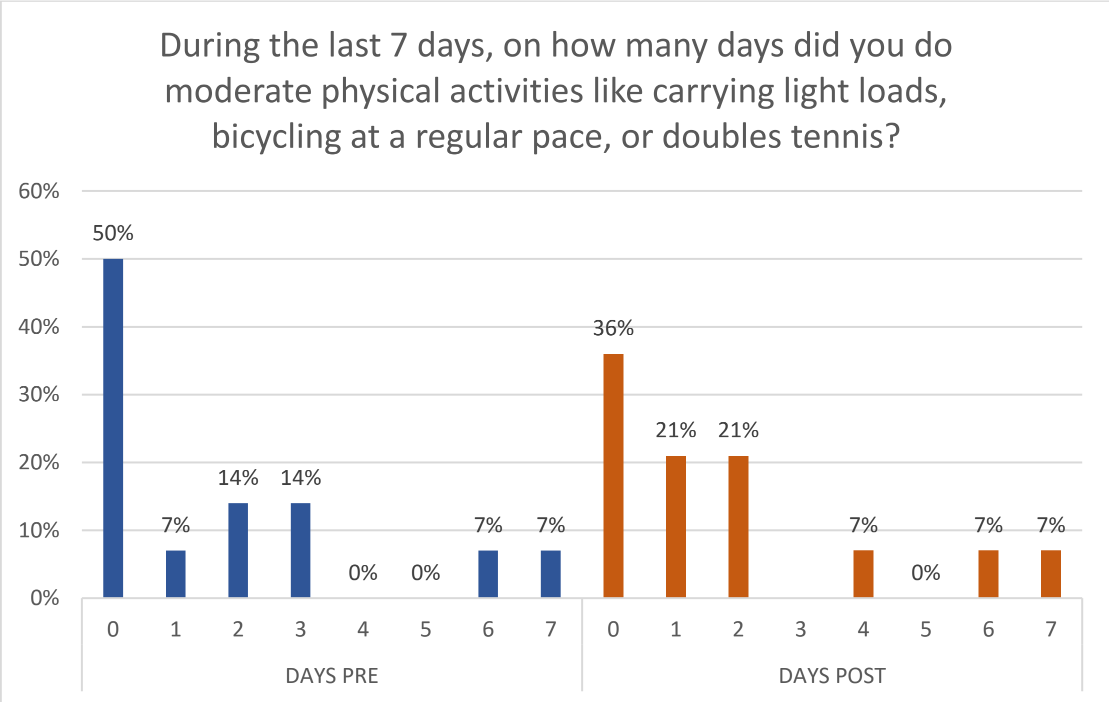


Figure 65: Health related moderate physical activity 1. Office use case

- For those volunteers who answered yes to the previous question, were asked: "How much time did you usually spend doing moderate physical activities on one of those days?". The results can be seen in Figure 66.

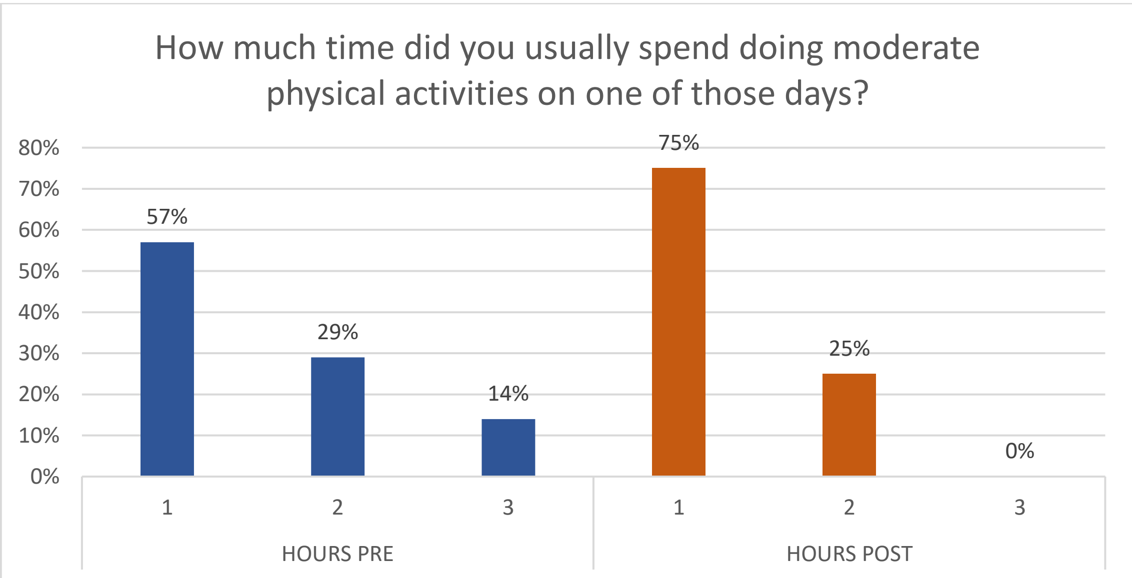


Figure 66: Health related moderate physical activity 2. Office use case

- "During the last 7 days, on how many days did you walk for at least 10 minutes at a time?". Participants answers can be seen in (Figure 67).

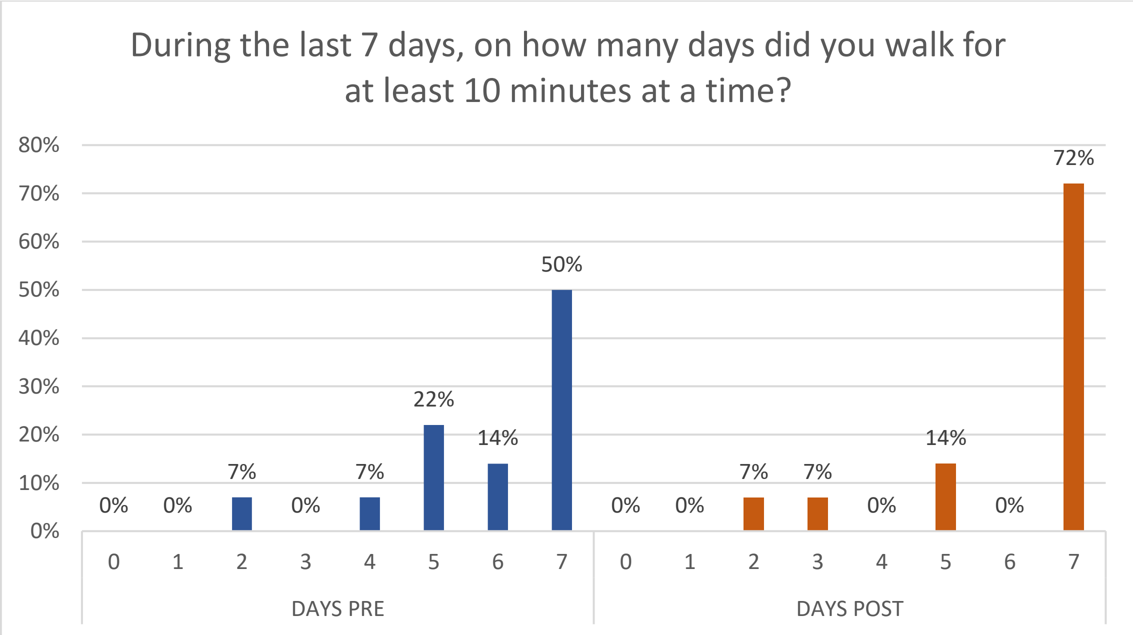


Figure 67: Health related walking physical activity 1. Office use case

- For those volunteers who answered yes to the previous question about walking, were asked about "How much time did you usually spend walking on one of those days?" (Figure 68).

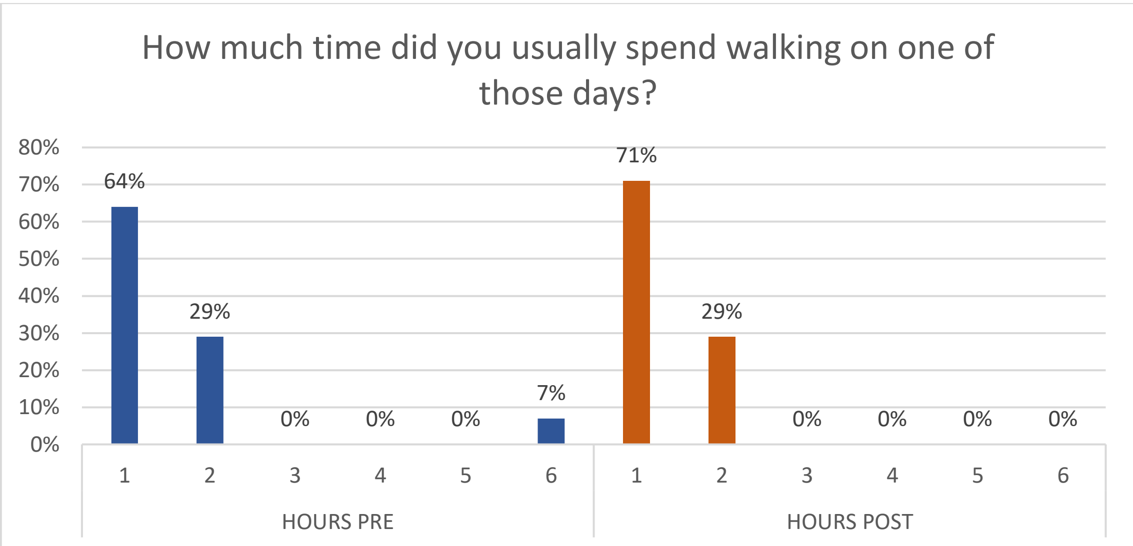


Figure 68: Health related walking physical activity 2. Office use case

- "During the last 7 days, how much time did you spend sitting on a weekday?"

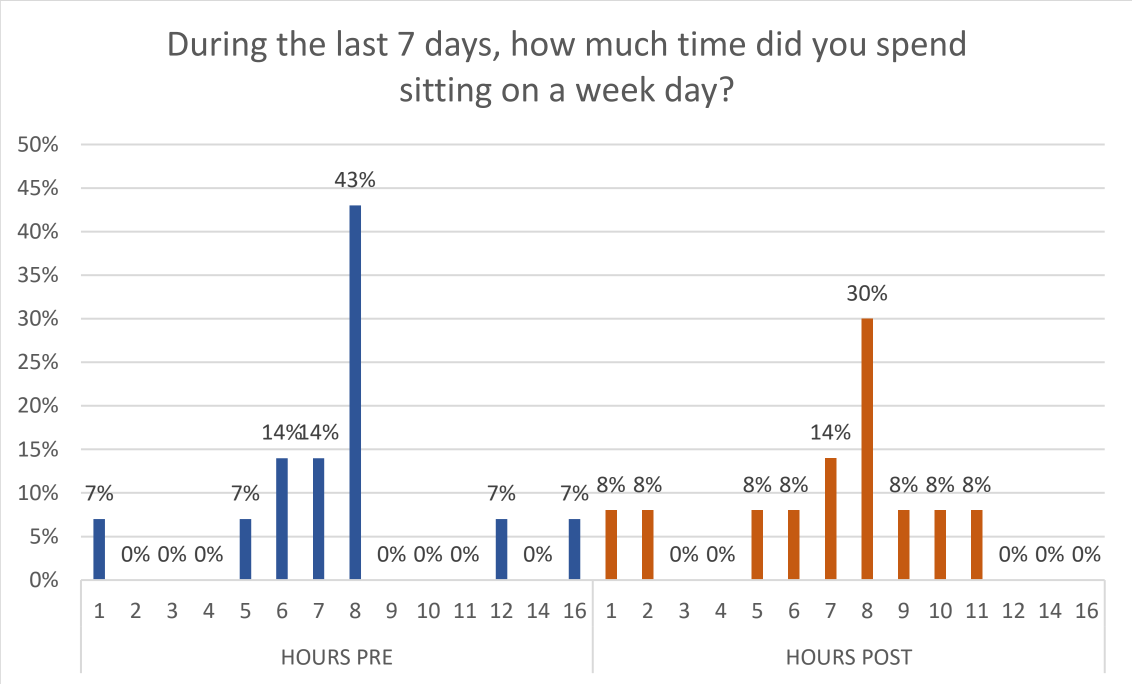


Figure 69: Health related sitting physical activity. Office use case

2.3.3.1.9 *Life-changing*

At the end of the post-questionnaire, there were 2 additional questions. The first one, asked the participants if they have had any significant situation that could have affected their lives. Only 4 of them answered this question in the affirmative (29%). These volunteers could have suffered some changes in their life, but they may not be due to the WAOW tool usage. None of them wanted to explain what kind of situation was the one affecting their lives.

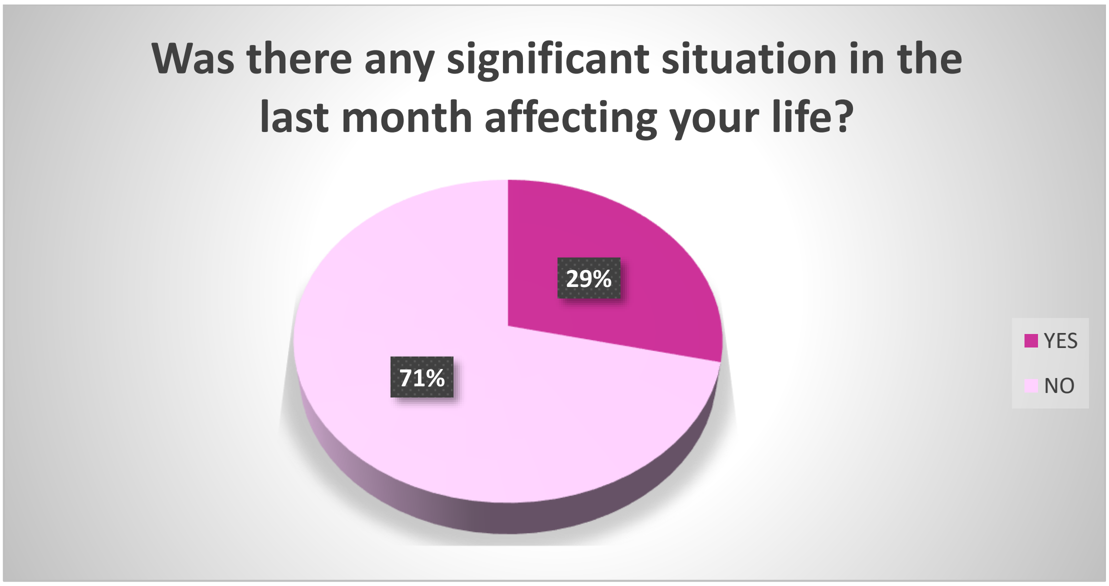


Figure 70: Life changing

2.3.3.1.10 *User experience assessment*

For the evaluation of the degree of difficulty experienced by the user in understanding, learning, and using the WAOW tool, as well as for the assessment of the attractiveness of the technology, the participants were asked through the UEQ questionnaire, 26 items grouped in 6 dimensions: 1) Attractiveness, 2) Perspicuity, 3) Efficiency, 4) Dependability, 5) Stimulation, and 6) Novelty.

The questionnaire consists of pairs of contrasting attributes that may apply to the WAOW tool. The numbers from 1 to 7 between the attributes represent gradations between the opposites. The volunteers can express their agreement with the attributes by selecting the number that most closely reflects their impression.

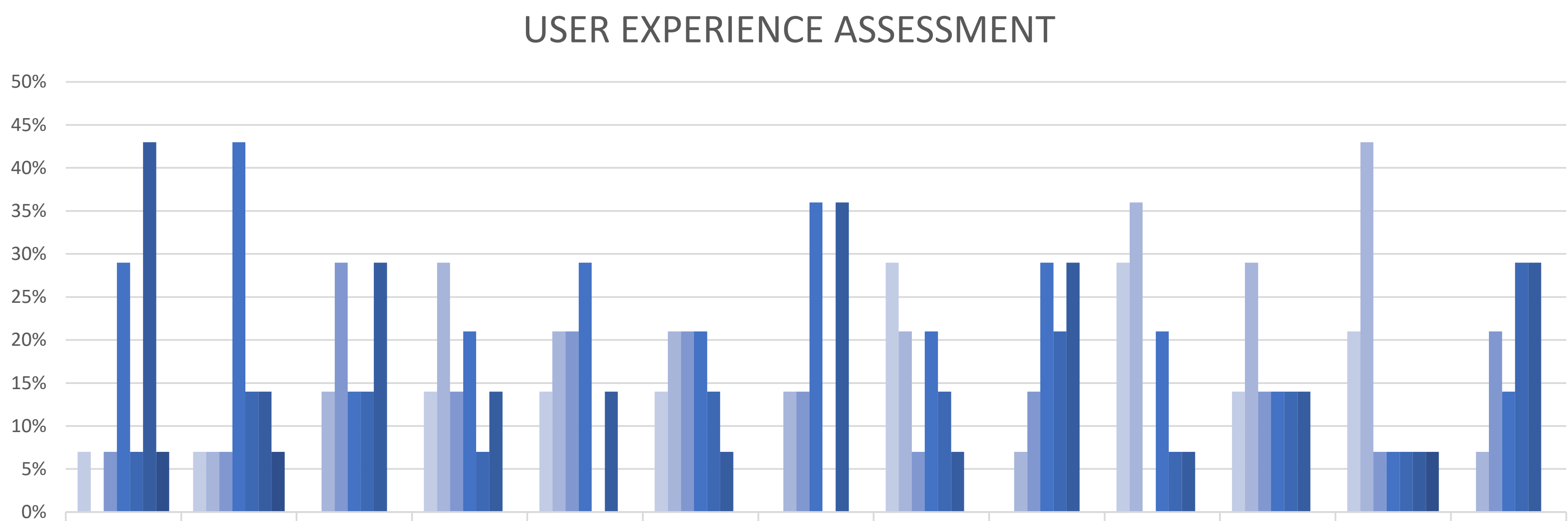
There are positive and negative items.

- Positive items (Positive evaluation when the score is high): 1, 2, 6, 7, 8, 11, 13, 14, 15, 16, 20, 22,26.
- Negative items (Positive evaluation when the score is low): 3, 4, 5, 9, 10, 12, 17, 18, 19, 21, 23, 24, 25.

The graph is divided in two parts, to improve the viewing of the data.

	1 Annoying - 7 Enjoyable	1 Not understandable - 7 Understandable	1 Creative - 7 Dull	1 Easy to learn - 7 Difficult to learn	1 Valuable - 7 Inferior	1 Boring - 7 Exciting	1 Not interesting - 7 Interesting	1 Unpredictable - 7 Predictable	1 Fast - 7 Slow	1 Inventive - 7 Conventional	1 Obstructive - 7 Supportive	1 Good - 7 Bad	1 Complicated - 7 Easy
1	0%	7%	7%	29%	14%	0%	0%	0%	0%	7%	0%	21%	0%
2	7%	7%	21%	21%	29%	14%	7%	14%	7%	29%	14%	21%	14%
3	14%	21%	14%	7%	21%	14%	7%	7%	0%	7%	7%	0%	0%
4	21%	14%	14%	14%	14%	36%	21%	43%	21%	29%	29%	21%	14%
5	43%	14%	36%	14%	21%	7%	29%	14%	36%	21%	21%	29%	29%
6	7%	29%	7%	14%	0%	21%	29%	21%	21%	7%	29%	7%	43%
7	7%	7%	0%	0%	0%	7%	7%	0%	14%	0%	0%	0%	0%

Figure 71: User Experience Assessment 1. Office use case



	1 Unlikeable - 7 Pleasing	1 Usual - 7 Leading edge	1 Unpleasant - 7 Pleasant	1 Secure - 7 Not secure	1 Motivating - 7 Demotivating	1 Meets expectations - 7 Does not meet expectations	1 Inefficient - 7 Efficient	1 Clear - 7 Confusing	1 Impractical - 7 Practical	1 Organized - 7 Cluttered	1 Attractive - 7 Unattractive	1 Friendly - 7 Unfriendly	1 Conservative - 7 Innovative
1	7%	7%	0%	14%	14%	14%	0%	29%	0%	29%	14%	21%	0%
2	0%	7%	14%	29%	21%	21%	14%	21%	7%	36%	29%	43%	7%
3	7%	7%	29%	14%	21%	21%	14%	7%	14%	0%	14%	7%	21%
4	29%	43%	14%	21%	29%	21%	36%	21%	29%	21%	14%	7%	14%
5	7%	14%	14%	7%	0%	14%	0%	14%	21%	7%	14%	7%	29%
6	43%	14%	29%	14%	14%	7%	36%	7%	29%	7%	14%	7%	29%
7	7%	7%	0%	0%	0%	0%	0%	0%	0%	0%	0%	7%	0%

Figure 72: User Experience Assessment 2. Office use case

2.3.3.2 Use Case: Production (Grupo Antolín)

2.3.3.2.1 Physical Working Conditions

To the question "Does your job ever require that you wear personal protective equipment?", 85% of the participants answered affirmatively and only a 15% answered in a negative way, as it is showed in the next figure.

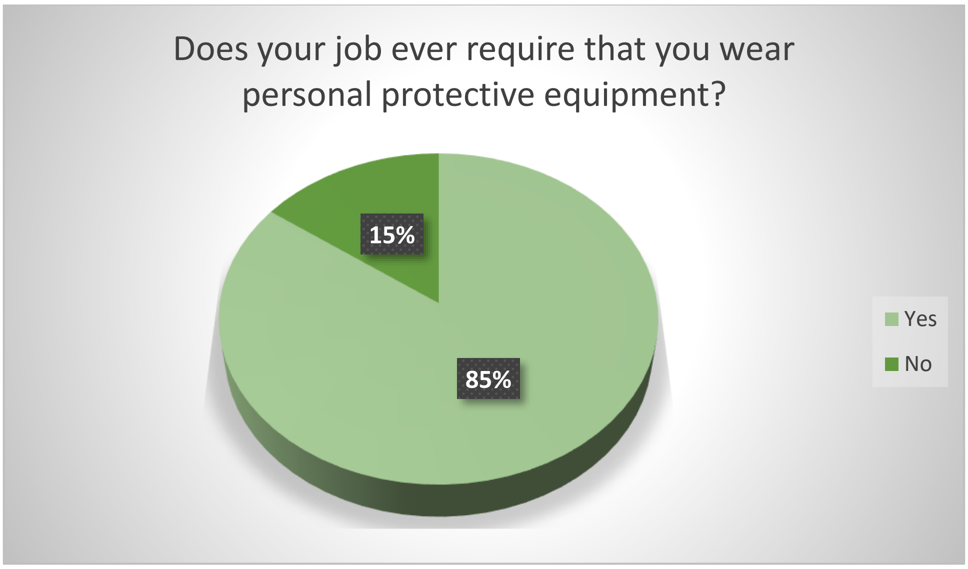


Figure 73: Protective equipment. Production use case.

91% of the workers who must wear this protective equipment, always use it. The remaining 9% don't always use it.

Next are the answers of the production use case workers comparing the pre and post.

When the volunteers were asked if they health or safety were at risk because of their work, they gave the following answers:

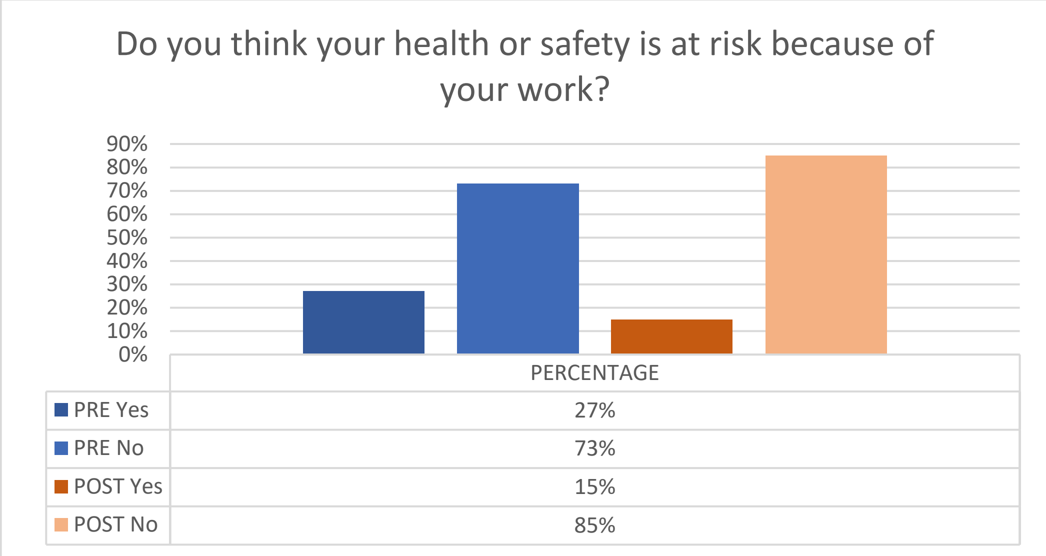


Figure 74: Risk because of work. Production use case.

The production volunteers were asked if their work affected their health. This is shown in the next figure.

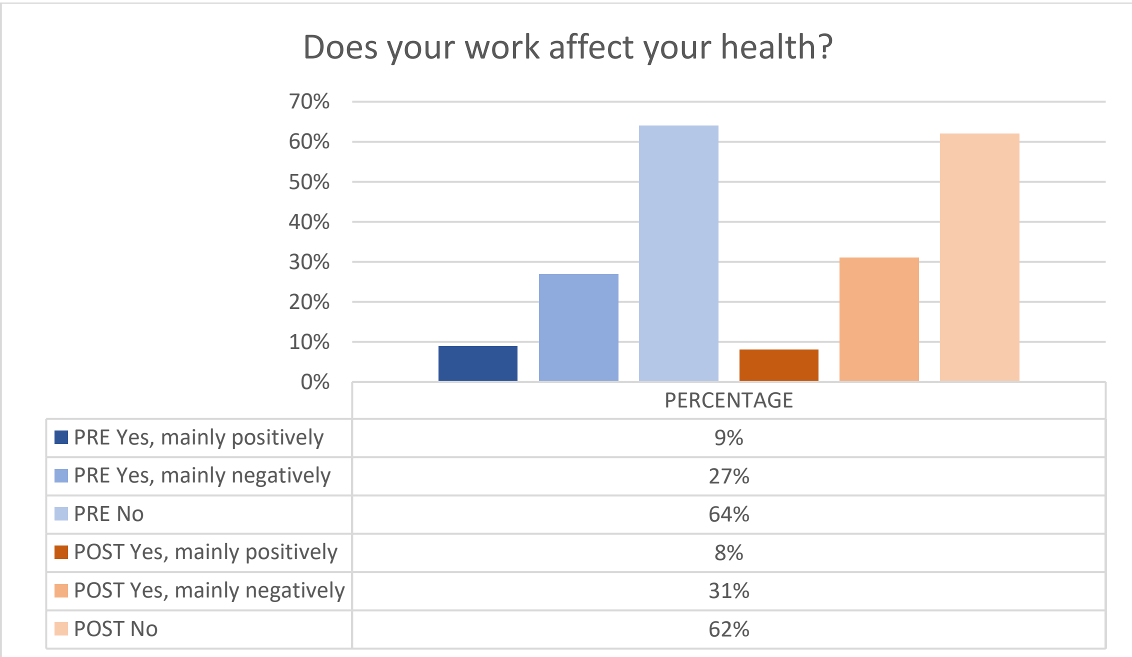
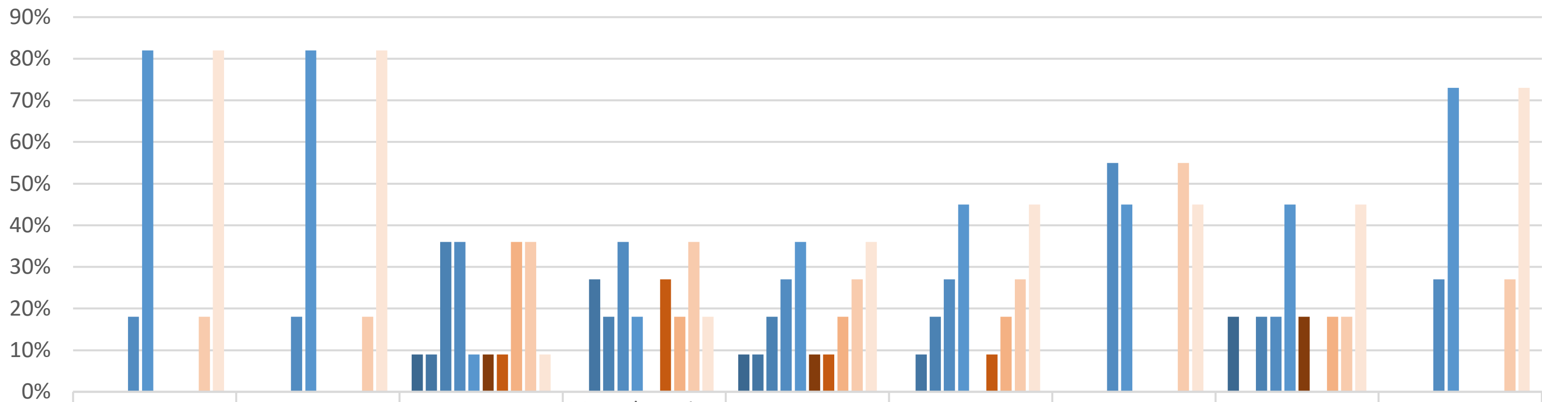


Figure 75: Health affected by the work. Production use case.

The next graph compares different health problems that the participants suffered at the beginning and at the end of the pilots.

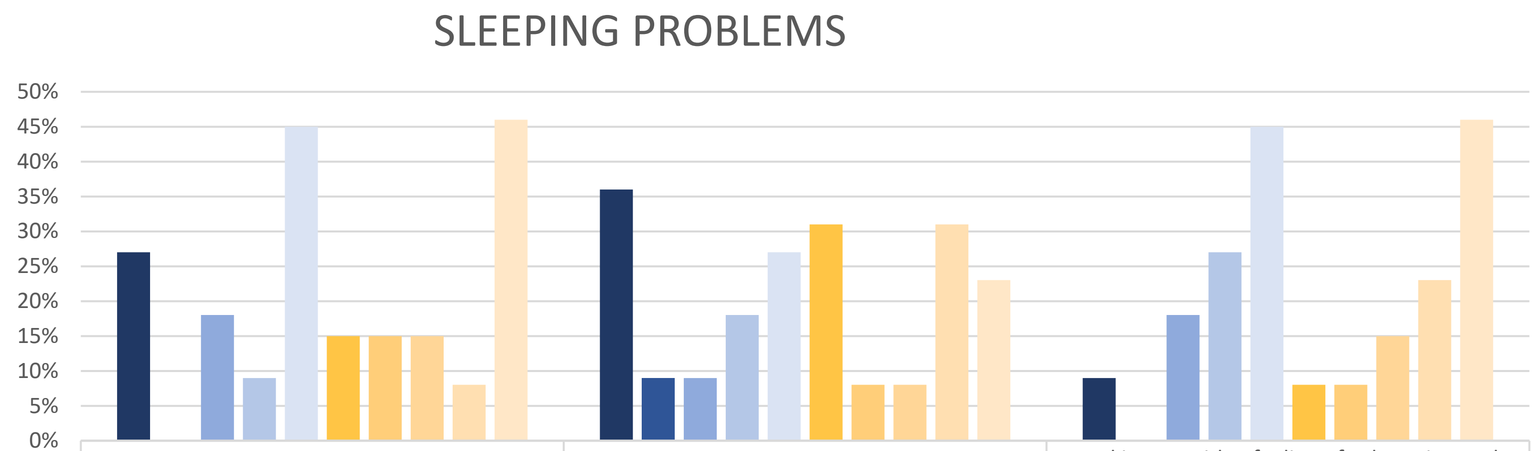
Over the last 6 months, did you have any of the following health problems?



	Hearing problems	Skin problems	Backache	Muscular pains in shoulders, neck and/or upper limbs	Muscular pains in lower limbs	Headaches, eyestrain	Injury	Overall fatigue	Other (Spontaneus)
PRE All the time	0%	0%	9%	0%	9%	0%	0%	18%	0%
PRE A large part of the time	0%	0%	9%	27%	9%	9%	0%	0%	0%
PRE Part of the time	0%	0%	36%	18%	18%	18%	0%	18%	0%
PRE A small part of the time	18%	18%	36%	36%	27%	27%	55%	18%	27%
PRE Not at all	82%	82%	9%	18%	36%	45%	45%	45%	73%
POST All the time	0%	0%	9%	0%	9%	0%	0%	18%	0%
POST A large part of the time	0%	0%	9%	27%	9%	9%	0%	0%	0%
POST Part of the time	0%	0%	36%	18%	18%	18%	0%	18%	0%
POST A small part of the time	18%	18%	36%	36%	27%	27%	55%	18%	27%
POST Not at all	82%	82%	9%	18%	36%	45%	45%	45%	73%

Figure 76: Health problems. Production use case.

The next question is "Over the last 6 months, how often did you have any of the following sleep related problems?". The volunteers were asked about 3 different aspects about sleeping problems:



	Difficulty falling asleep	Waking up repeatedly during the sleep	Waking up with a feeling of exhaustion and fatigue
PRE Daily	27%	36%	9%
PRE Several times a week	0%	9%	0%
PRE Several times a month	18%	9%	18%
PRE Less often	9%	18%	27%
PRE Never	45%	27%	45%
POST Daily	15%	31%	8%
POST Several times a week	15%	8%	8%
POST Several times a month	15%	8%	15%
POST Less often	8%	31%	23%
POST Never	46%	23%	46%

Figure 77: Sleeping problems. Production use case.

The next questions are about times when the participants might have been absent from work.

- Over the past 6 months how many days in total were you absent from work due to sick leave or health-related leave? (Figure 78)

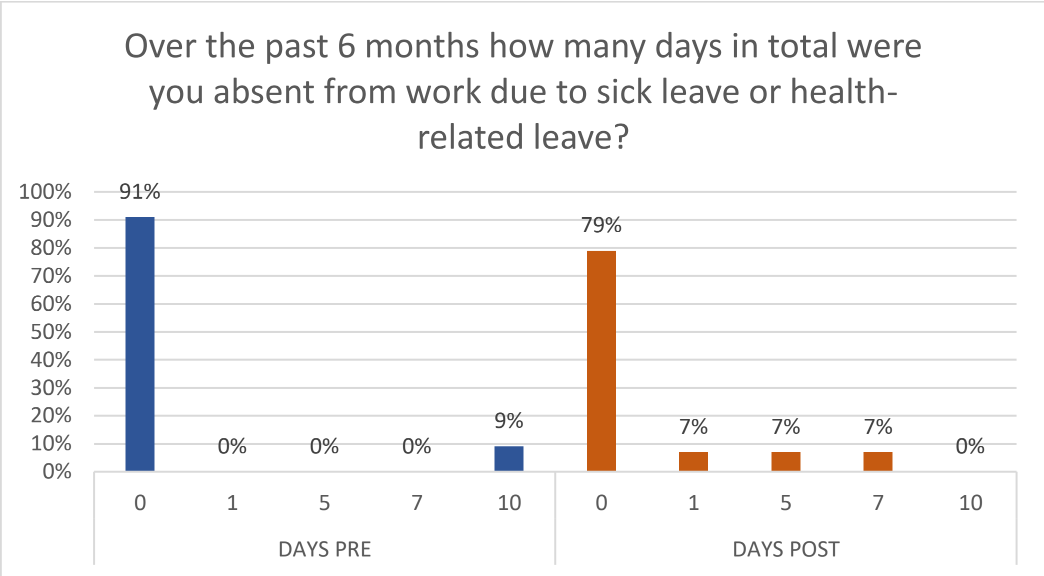


Figure 78: Absences from work. Production use case.

- Asked if these days were caused by accidents at work, the participants, answered the following:

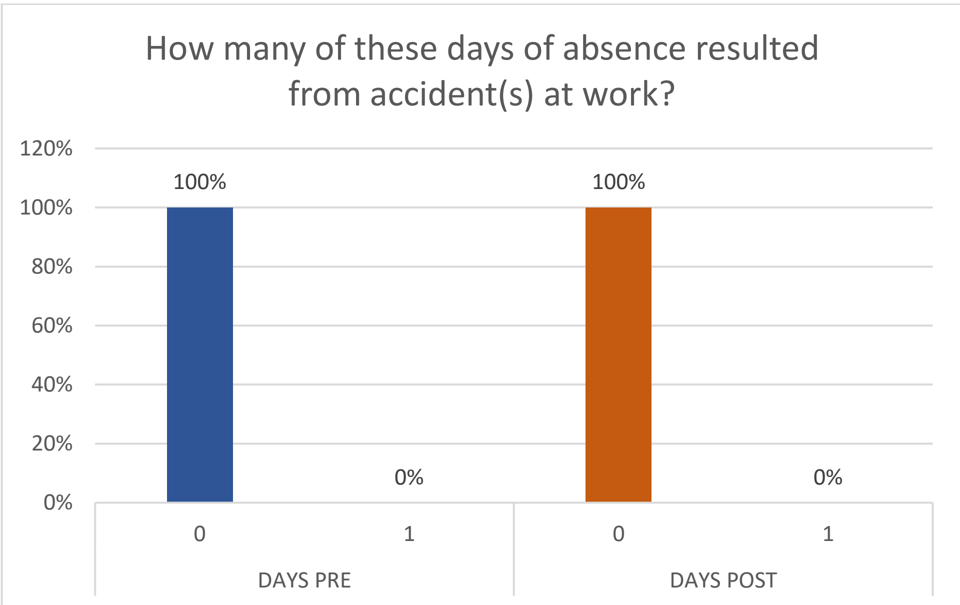


Figure 79: Accidents at work. Production use case.

- To the question if these days of absence resulted from health problems caused or made worse by their work (excluding accidents), the participants answered this way:

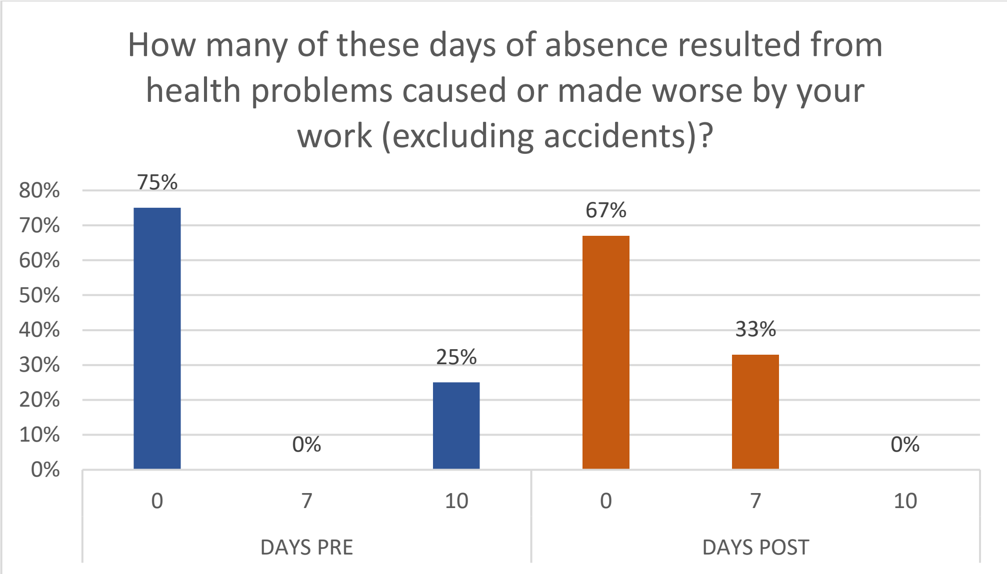


Figure 80: Health problems caused or made worse by work. Production use case.

- The volunteers were asked if they worked when they were sick, and the number of days. To have a more graphic view of the data, you can see the graphs below.

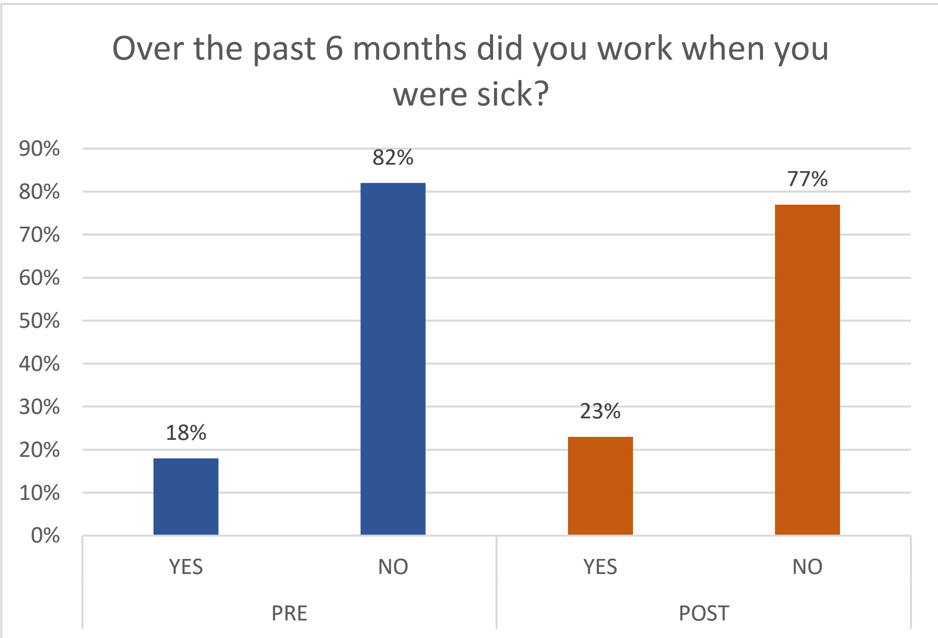


Figure 81: Working when sick. Production use case.

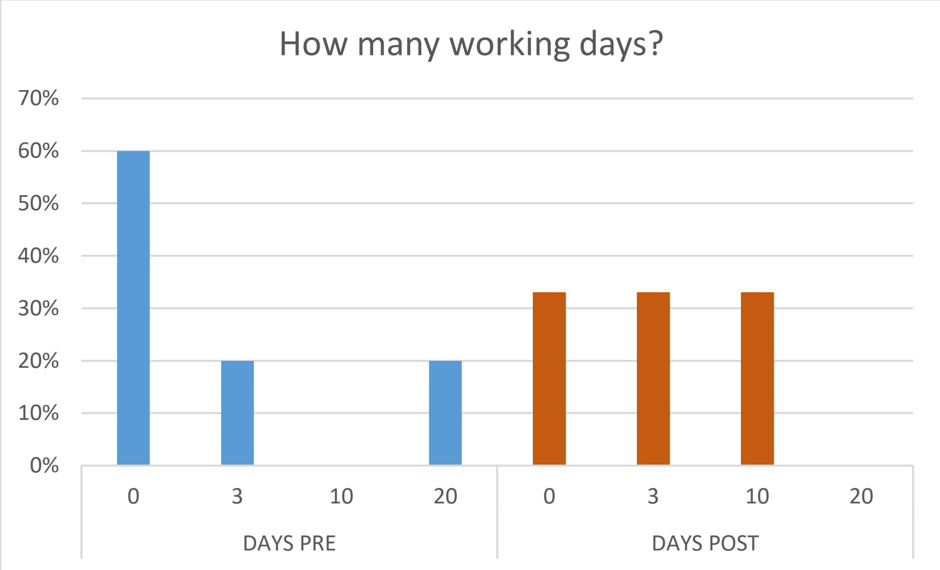
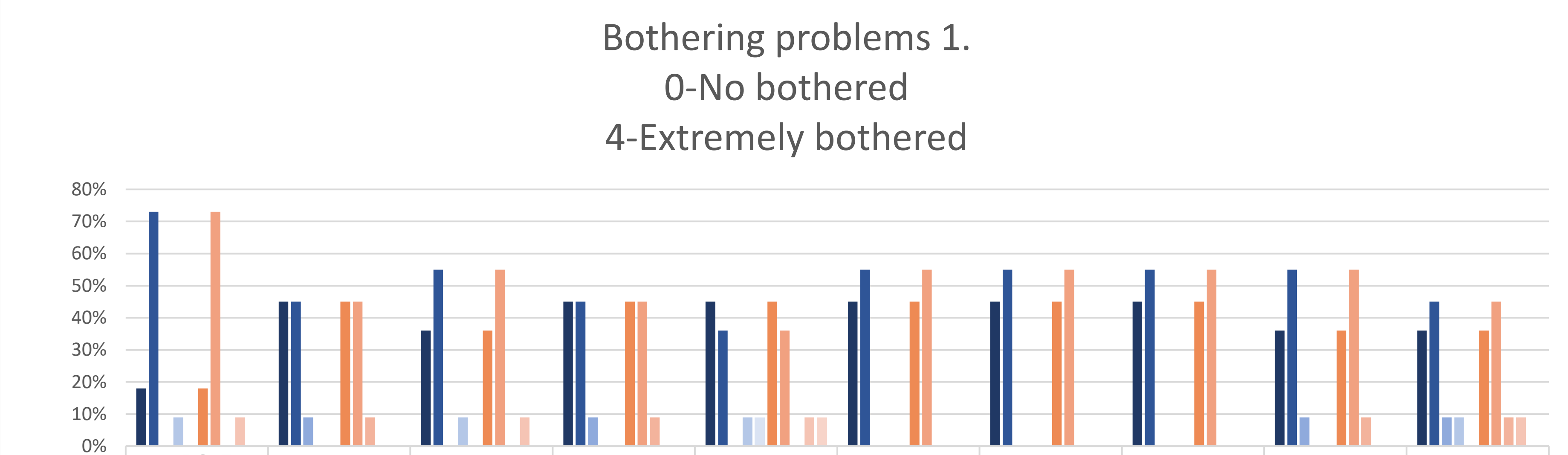


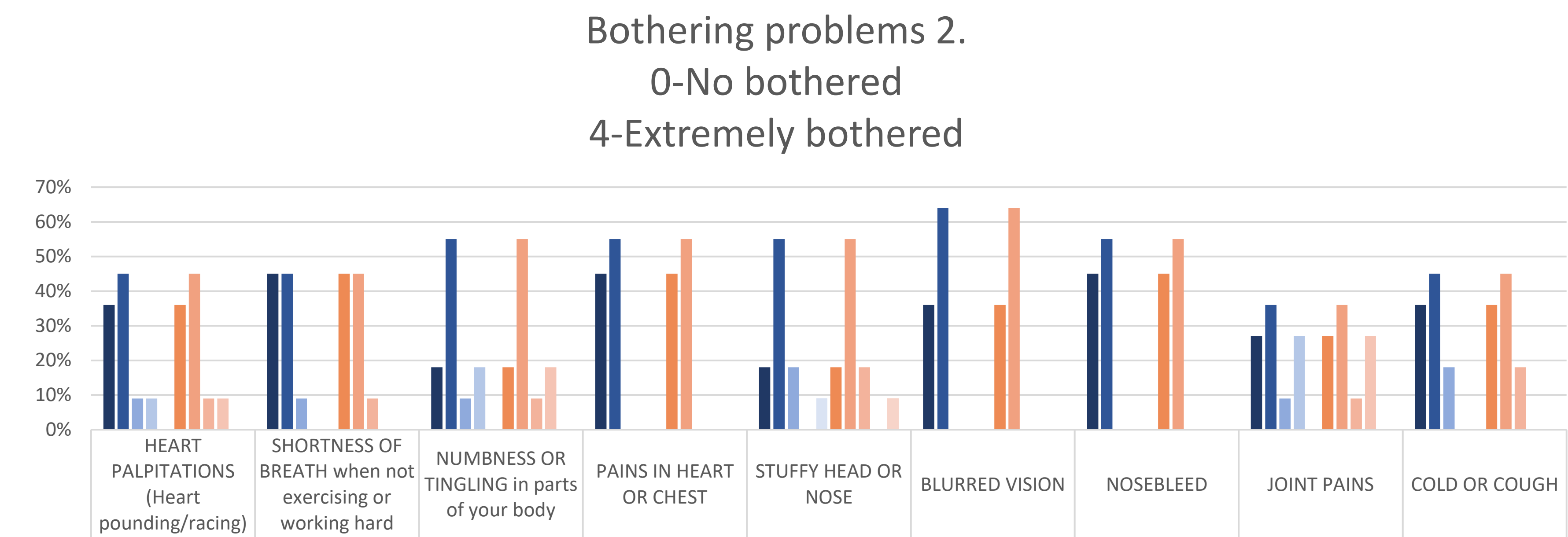
Figure 82: Working days when sick. Production use case.

To end with the questions about physical working conditions, the volunteers were asked about the physical symptoms they had suffered during the last few weeks and how much discomfort these problems caused them. The answering options are divided from 0 to 4, where 0 means that the participant has not been bothered by the problem, and 4 means that the problem has been an extreme bother. The graph is divided in two, due to the large number of symptoms related.



	WEIGHT CHANGE (gain or loss of 5 libs. or more)	CONSTIPATION	DIZZINES	DIARRHEA	FAINTNESS	NAUSEA AND/OR VOMITING	ACID STOMACH OR INDIGESTION	STOMACH PAIN (E.G., CRAMPS)	HOT OR COLD SPELLS	HANDS TREMBLING
PRE 0	18%	45%	36%	45%	45%	45%	45%	45%	36%	36%
PRE 1	73%	45%	55%	45%	36%	55%	55%	55%	55%	45%
PRE 2	0%	9%	0%	9%	0%	0%	0%	0%	9%	9%
PRE 3	9%	0%	9%	0%	9%	0%	0%	0%	0%	9%
PRE 4	0%	0%	0%	0%	9%	0%	0%	0%	0%	0%
POST 0	18%	45%	36%	45%	45%	45%	45%	45%	36%	36%
POST 1	73%	45%	55%	45%	36%	55%	55%	55%	55%	45%
POST 2	0%	9%	0%	9%	0%	0%	0%	0%	9%	9%
POST 3	9%	0%	9%	0%	9%	0%	0%	0%	0%	9%
POST 4	0%	0%	0%	0%	9%	0%	0%	0%	0%	0%

Figure 83: Bothering problems in the last 2 weeks 1. Production use case.



	HEART PALPITATIONS (Heart pounding/racing)	SHORTNESS OF BREATH when not exercising or working hard	NUMBNESS OR TINGLING in parts of your body	PAINS IN HEART OR CHEST	STUFFY HEAD OR NOSE	BLURRED VISION	NOSEBLEED	JOINT PAINS	COLD OR COUGH
PRE 0	36%	45%	18%	45%	18%	36%	45%	27%	36%
PRE 1	45%	45%	55%	55%	55%	64%	55%	36%	45%
PRE 2	9%	9%	9%	0%	18%	0%	0%	9%	18%
PRE 3	9%	0%	18%	0%	0%	0%	0%	27%	0%
PRE 4	0%	0%	0%	0%	9%	0%	0%	0%	0%
POST 0	36%	45%	18%	45%	18%	36%	45%	27%	36%
POST 1	45%	45%	55%	55%	55%	64%	55%	36%	45%
POST 2	9%	9%	9%	0%	18%	0%	0%	9%	18%
POST 3	9%	0%	18%	0%	0%	0%	0%	27%	0%
POST 4	0%	0%	0%	0%	9%	0%	0%	0%	0%

Figure 84: Bothering problems in the last 2 weeks 1. Production use case.

2.3.3.2.2 *Psychosocial working conditions*

To evaluate the psychological working conditions, the volunteers were asked the following questions in the pre and post questionnaires:

- "Is your work meaningful?"

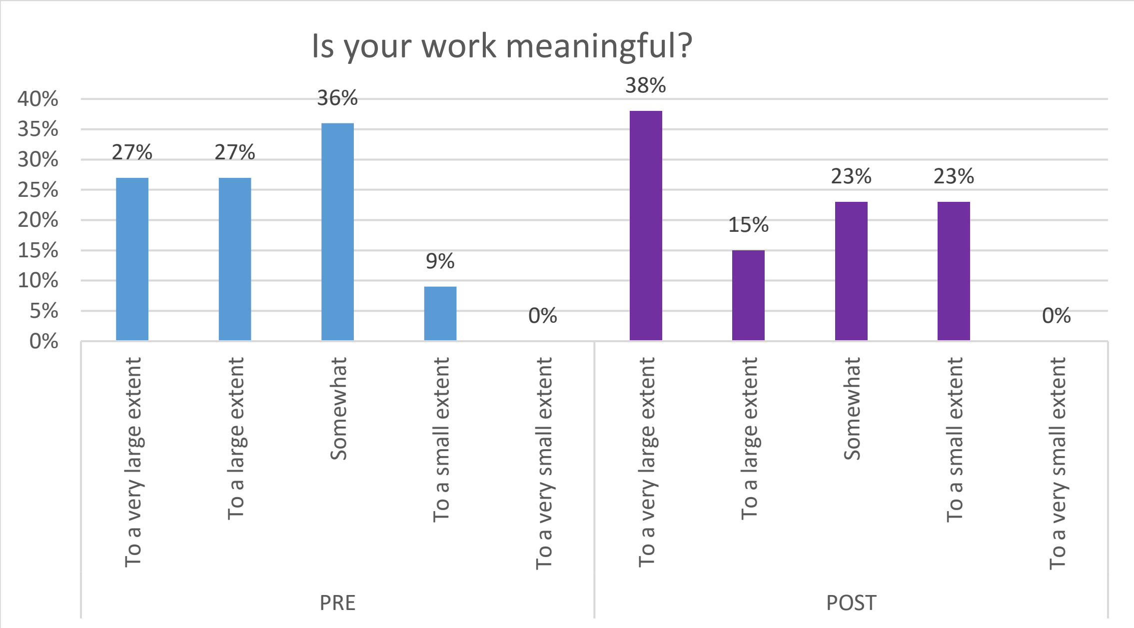


Figure 85: Meaning of work. Production use case.

- "How pleased are you with your job as a whole?"

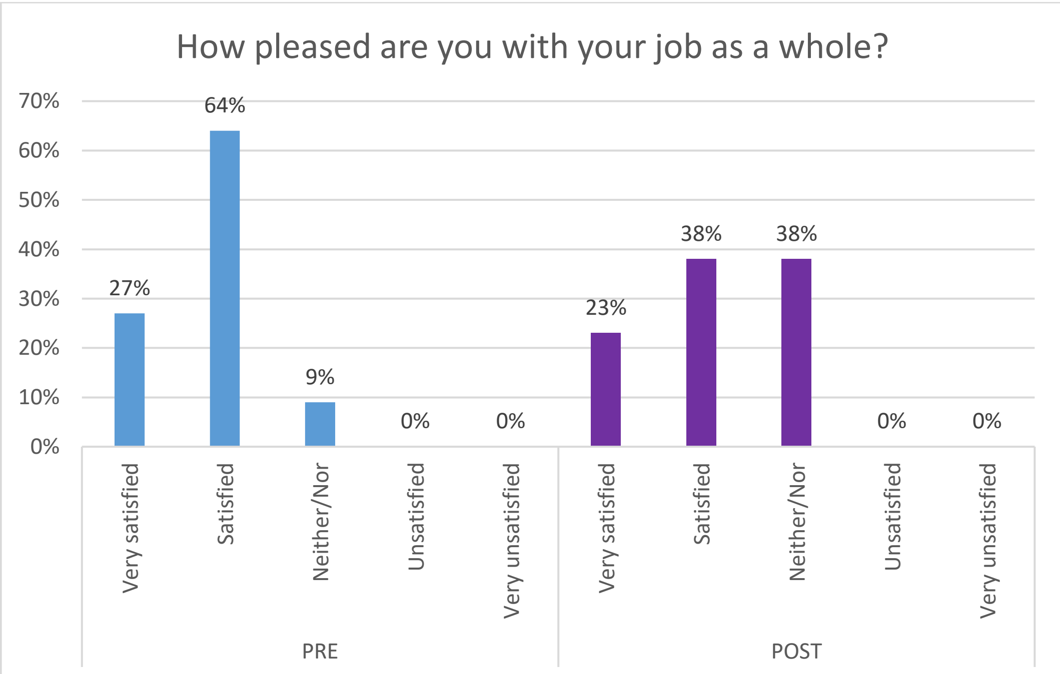


Figure 86: Job satisfaction. Production use case.

- "How often have you felt worn out?"

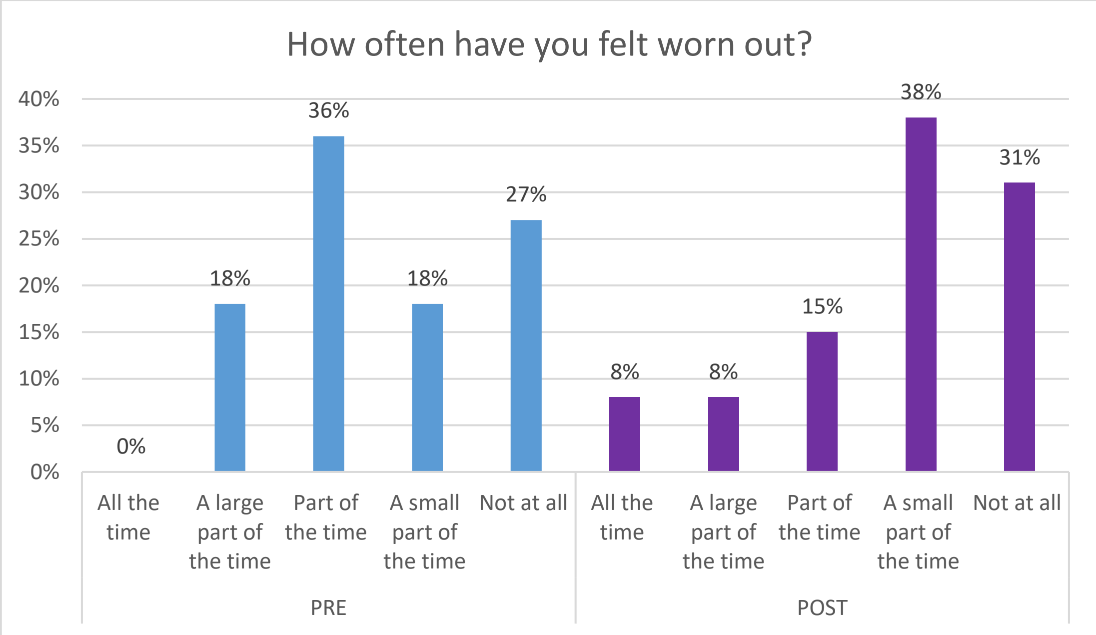


Figure 87: Burn out and stress 1. Production use case.

- "How often have you been physically exhausted?"

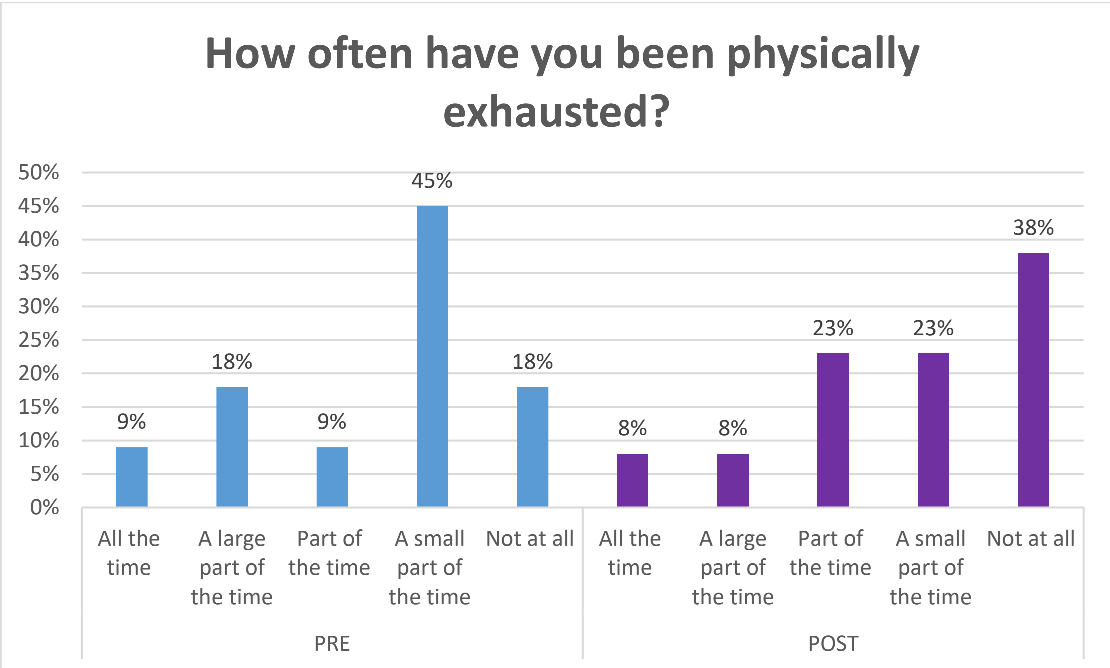


Figure 88: Burn out and stress 2. Production use case.

- "How often have you been emotionally exhausted?"

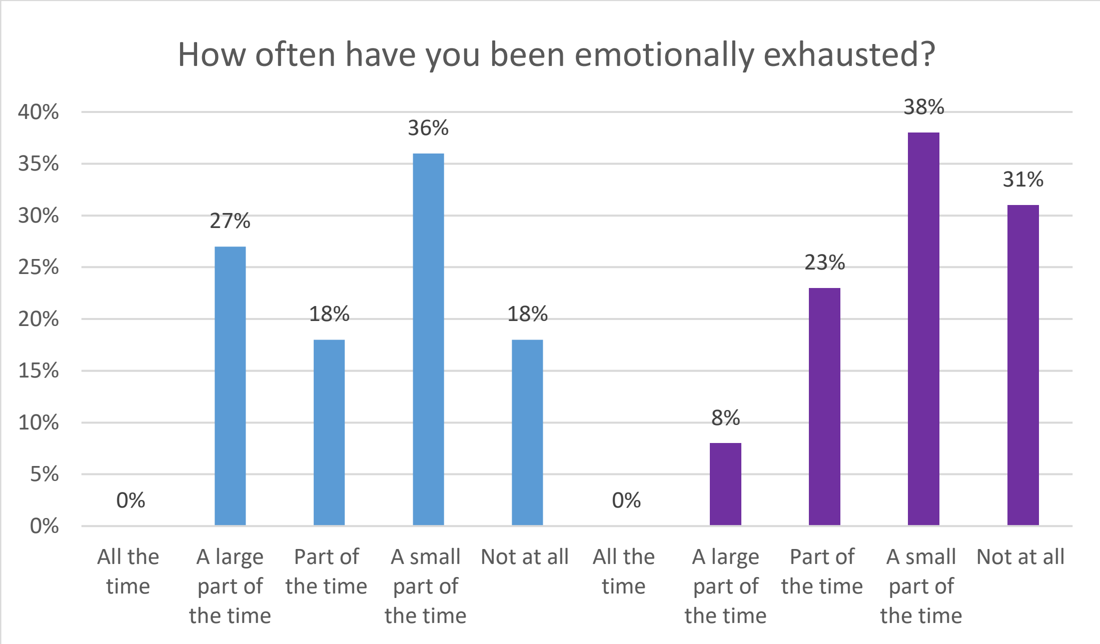


Figure 89: Burn out and stress 3. Production use case.

- "How often have you been tense?"

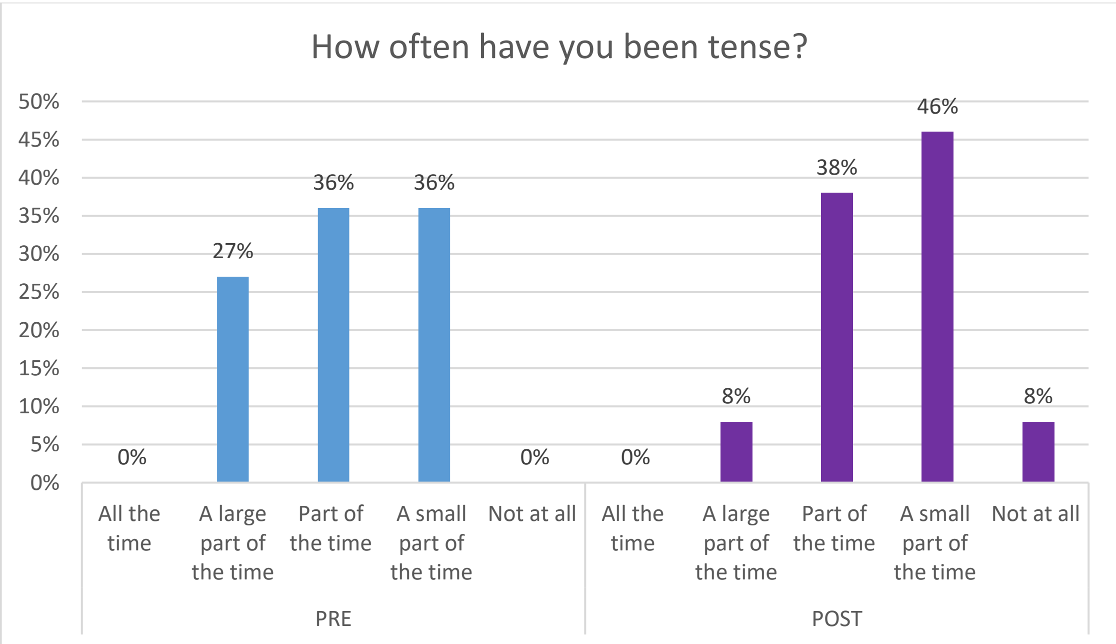


Figure 90: Burn out and stress 4. Production use case.

- "Is your work recognized and appreciated by the management?"

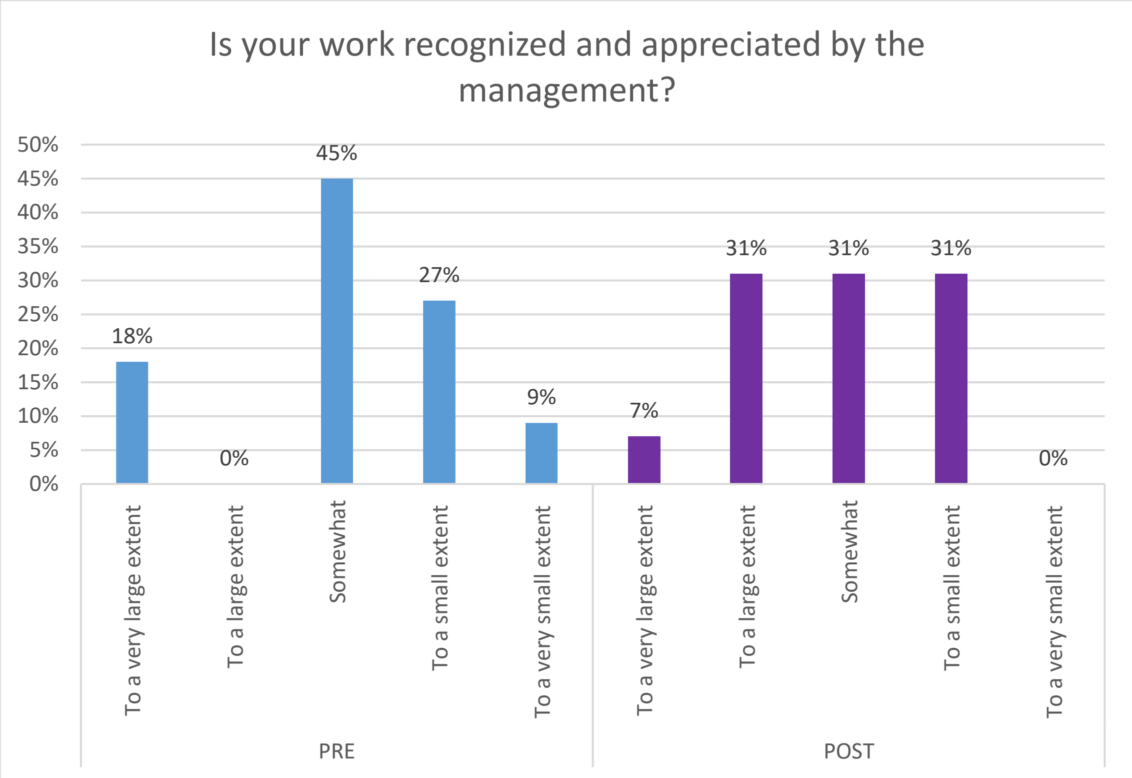
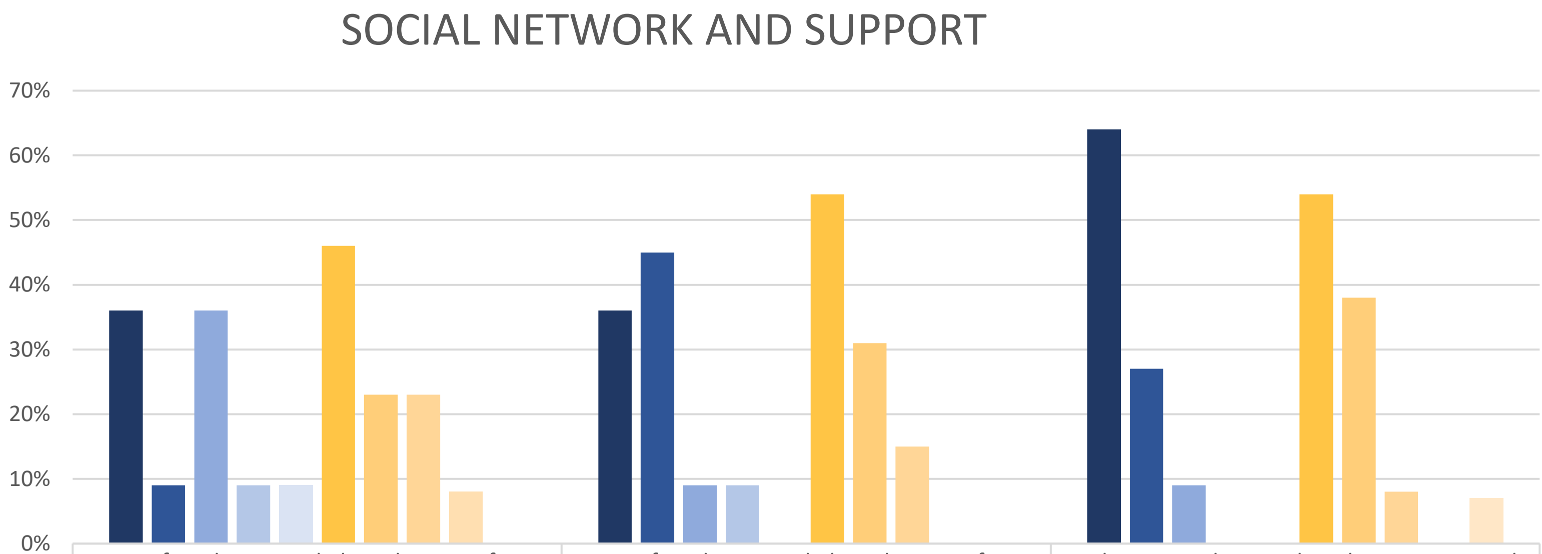


Figure 91: Social network and Support 1. Production use case

The next figure shows the social network and support felt by the volunteers in the pre and post questionnaire. To evaluate this aspect, three questions were asked to the participants.

- "How often do you get help and support from your immediate superior, if needed?"
- "How often do you get help and support from your colleagues, if needed?"
- "Is there a good atmosphere between you and your colleagues?"

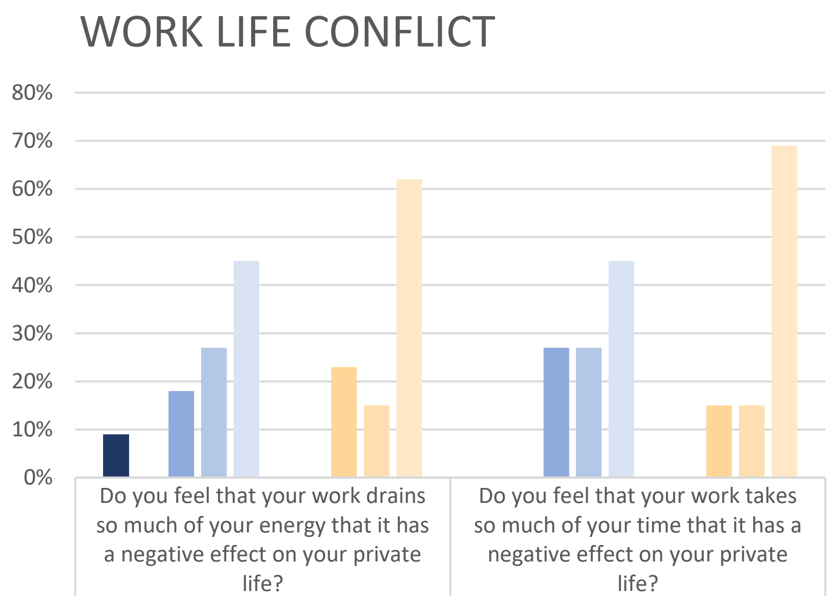


	How often do you get help and support from your immediate superior, if needed?	How often do you get help and support from your colleagues, if needed?	Is there a good atmosphere between you and your colleagues?
PRE Always	36%	36%	64%
PRE Often	9%	45%	27%
PRE Sometimes	36%	9%	9%
PRE Seldom	9%	9%	0%
PRE Never/hardly ever	9%	0%	0%
POST Always	46%	54%	54%
POST Often	23%	31%	38%
POST Sometimes	23%	15%	8%
POST Seldom	8%	0%	0%
POST Never/hardly ever	0%	0%	7%

Figure 92: Social network and support 2. Production use case.

We continue with two questions related to work life conflict. They can be seen in detail in the next graphic.

- "Do you feel that your work drains so much of your energy that it has a negative effect on your private life?"
- "Do you feel that your work takes so much of your time that it has a negative effect on your private life?"



	Do you feel that your work drains so much of your energy that it has a negative effect on your private life?	Do you feel that your work takes so much of your time that it has a negative effect on your private life?
PRE To a very large extent	9%	0%
PRE To a large extent	0%	0%
PRE Somewhat	18%	27%
PRE To a small extent	27%	27%
PRE To a very small extent	45%	45%
POST To a very large extent	0%	0%
POST To a large extent	0%	0%
POST Somewhat	23%	15%
POST To a small extent	15%	15%
POST To a very small extent	62%	69%

Figure 93: Work life conflict. Production use case

2.3.3.2.3 *Physical status*

To evaluate the physical status of the participants at the beginning and at the end of the pilots, they were asked seven questions related with this subject.

Five graphs will show the detailed data of their answers.

- "In general, would you say your health is…"

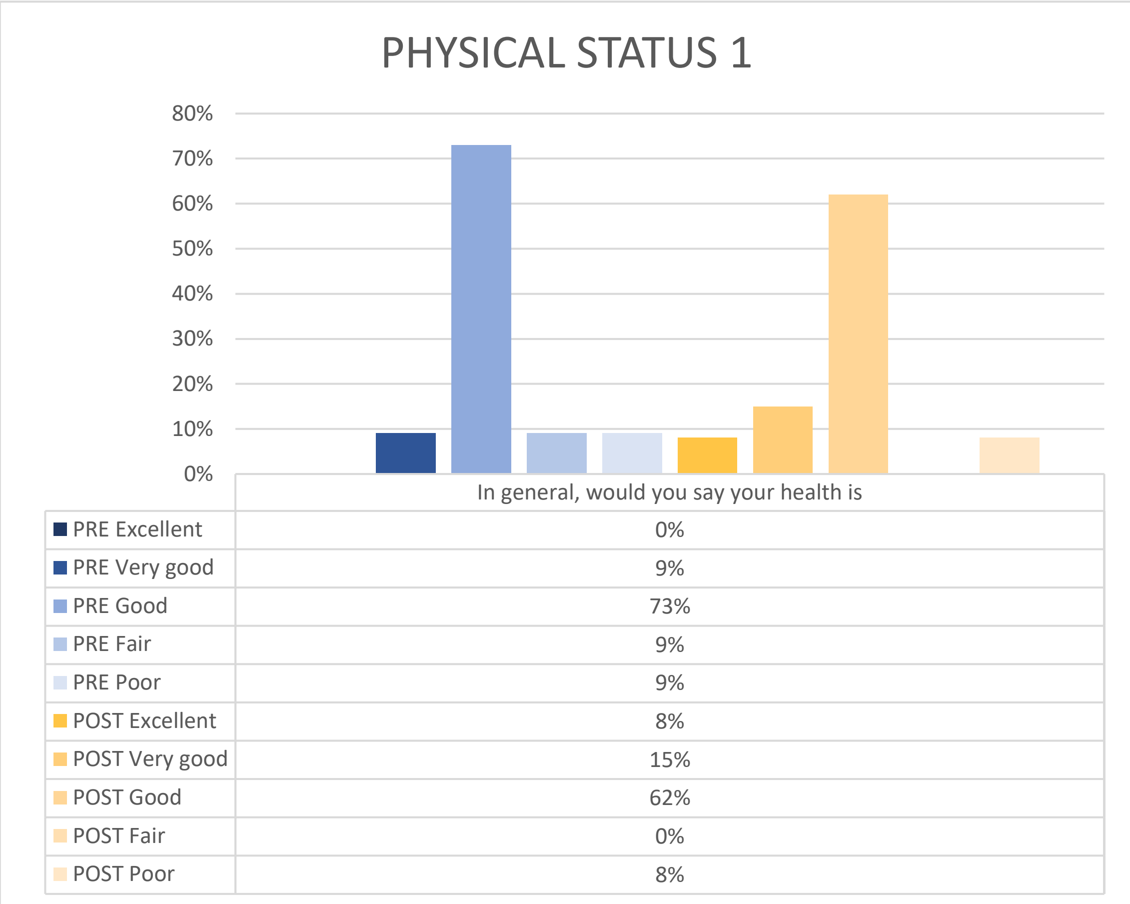
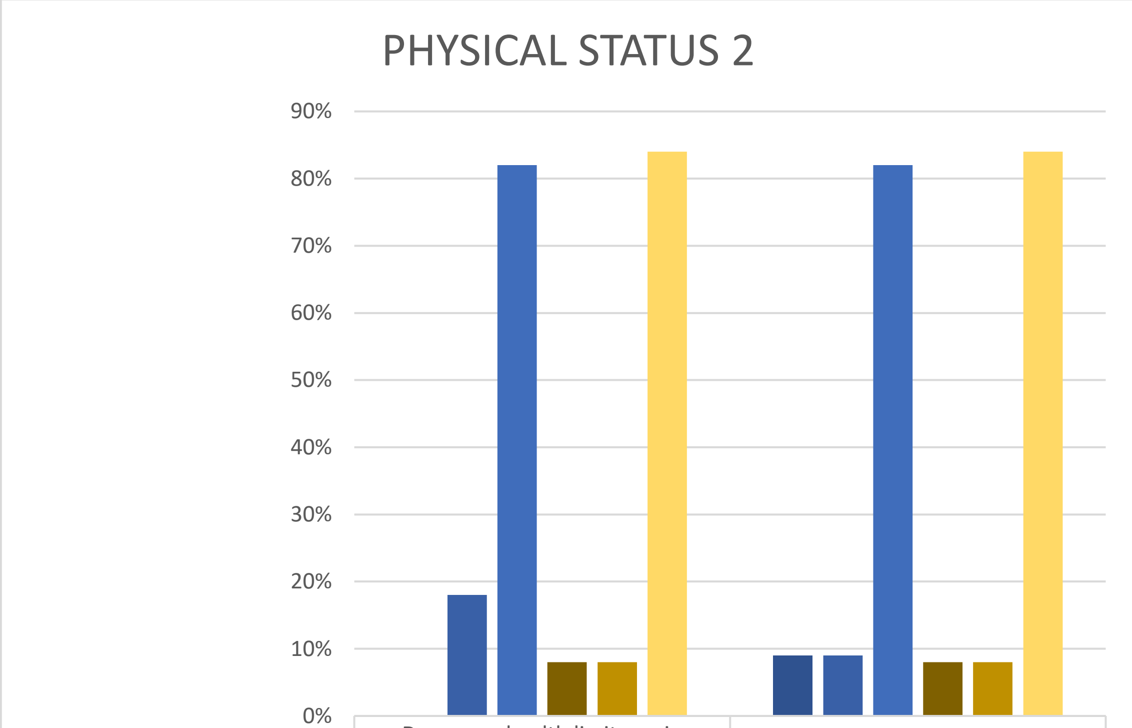


Figure 94: Physical status1. Production use case.

The next figure shows two questions related with the physical status of the participants at the beginning and at the end of the pilots.

- "Does your health limit you in MODERATE ACTIVITIES, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf?"
- "Does your health limit you in climbing SEVERAL flights of stairs?"



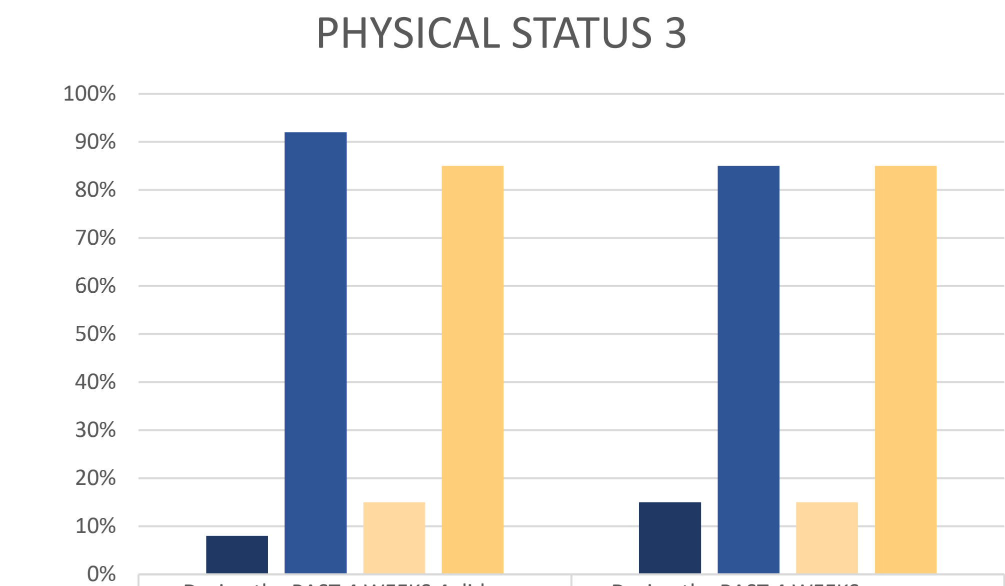
	Does your health limit you in MODERATE ACTIVITIES, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf	Does your health limit you in climbing SEVERAL flights of stairs
PRE Yes, Limited A Lot	0%	9%
PRE Yes, Limited A Little	18%	9%
PRE No, Not Limited At All	82%	82%
POST Yes, Limited A Lot	8%	8%
POST Yes, Limited A Little	8%	8%
POST No, Not Limited At All	84%	84%

Figure 95: Physical status 2. Production use case.

Again, the next figure shows two questions related with the physical status of the participants in a period of four weeks in advance.

- "During the PAST 4 WEEKS 4 did you ACCOMPLISHED LESS than you would like AS A RESULT OF YOUR PHYSICAL HEALTH?"

- "During the PAST 4 WEEKS were your work or regular activities limited in the KIND of work or other activities AS A RESULT OF YOUR PHYSICAL HEALTH?"

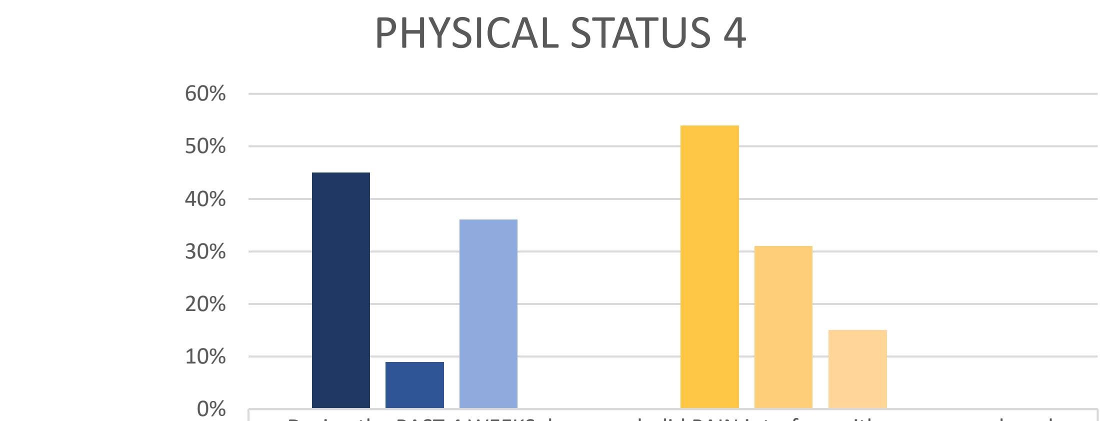


	During the PAST 4 WEEKS 4 did you ACCOMPLISHED LESS than you would like AS A RESULT OF YOUR PHYSICAL HEALTH?	During the PAST 4 WEEKS were your work or regular activities limited in the KIND of work or other activities AS A RESULT OF YOUR PHYSICAL HEALTH?
PRE Yes	8%	15%
PRE No	92%	85%
POST Yes	15%	15%
POST No	85%	85%

Figure 96: Physical status 3. Production use case.

The next physical status question is reflected in the graphic located below.

- "During the PAST 4 WEEKS, how much did PAIN interfere with your normal work (including both work outside the home and housework)?"

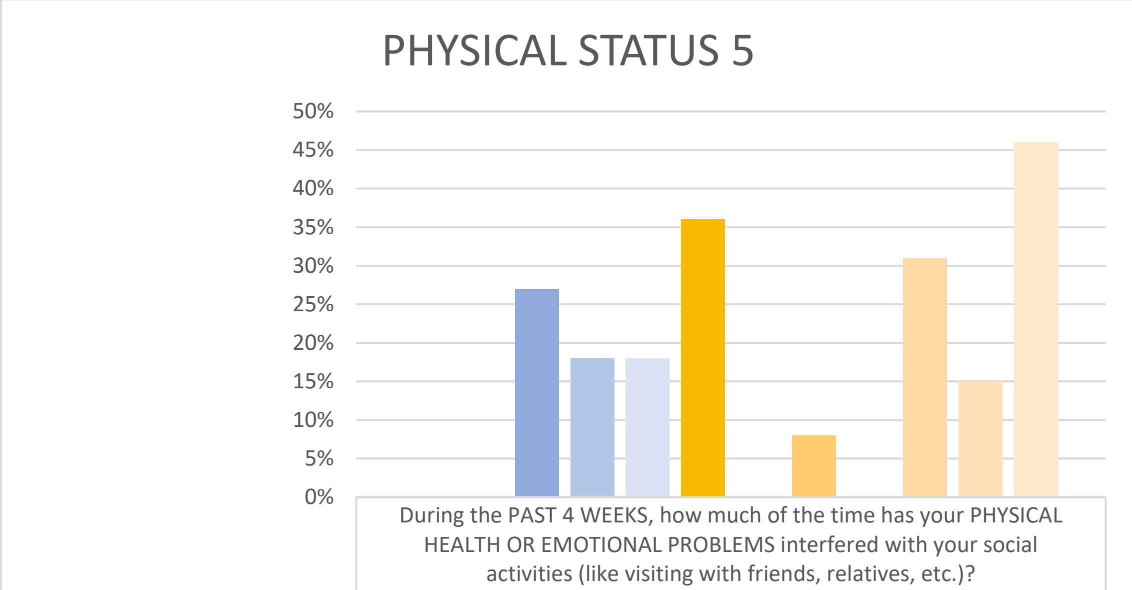


	During the PAST 4 WEEKS, how much did PAIN interfere with your normal work (including both work outside the home and housework)?
PRE Not at all	45%
PRE A little bit	9%
PRE Moderately	36%
PRE Quite a bit	0%
PRE Extremely	0%
POST Not at all	54%
POST A little bit	31%
POST Moderately	15%
POST Quite a bit	0%
POST Extremely	0%

Figure 97: Physical status 4. Production use case.

The last graphic about physical status of the participants is shown below.

- "How much of the time has your PHYSICAL HEALTH OR EMOTIONAL PROBLEMS interfered with your social activities (like visiting with friends, relatives, etc.)?"



	During the PAST 4 WEEKS, how much of the time has your PHYSICAL HEALTH OR EMOTIONAL PROBLEMS interfered with your social activities (like visiting with friends, relatives, etc.)?
PRE All the time	0%
PRE Most of the time	0%
PRE A good bit of the time	27%
PRE Some of the time	18%
PRE A little of the time	18%
POST None of the time	36%
POST All the time	0%
POST Most of the time	8%
POST A good bit of the time	0%
POST Some of the time	31%
POST A little of the time	15%
POST None of the time	46%

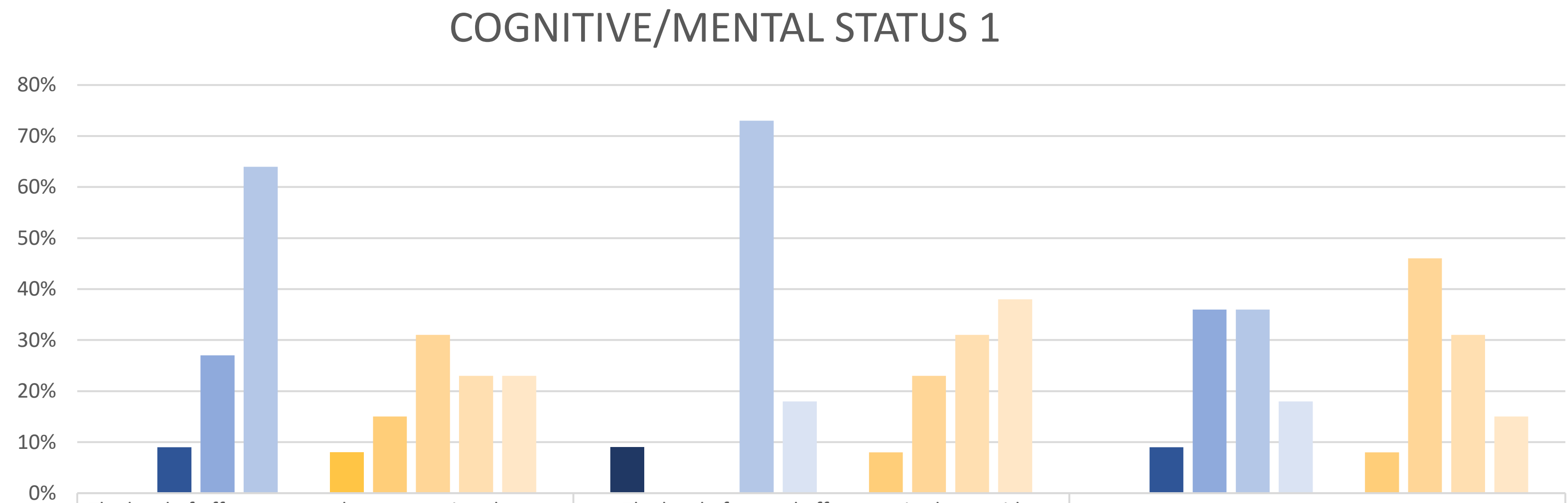
Figure 98: Physical status 5. Production use case.

2.3.3.2.4 Cognitive/Mental status

For the assessment of the cognitive/mental status of the production use case workers, they were asked six questions that address this topic.

The next graphic shows the detailed answers of the participants to the three first questions.

- "The level of effort or mental concentration that my job requires is…"
- "The level of mental effort required to avoid mistakes in my work is…"
- "The fatigue that my work produces is…"

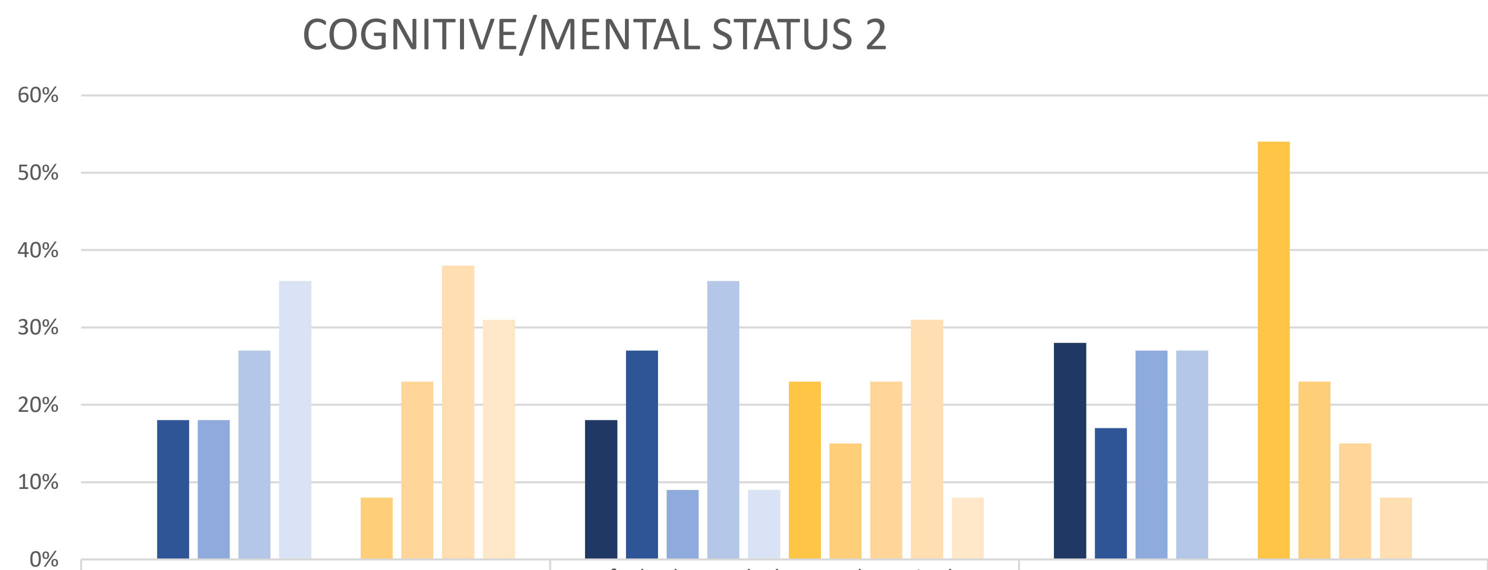


	The level of effort or mental concentration that my job requires is	The level of mental effort required to avoid mistakes in my work is	The fatigue that my work produces is
PRE Very low	0%	9%	0%
PRE Low	9%	0%	9%
PRE Medium	27%	0%	36%
PRE High	64%	73%	36%
PRE Very high	0%	18%	18%
POST Very low	8%	0%	0%
POST Low	15%	8%	8%
POST Medium	31%	23%	46%
POST High	23%	31%	31%
POST Very high	23%	38%	15%

Figure 99: Cognitive/Mental status 1. Production use case.

The next three questions about cognitive/mental status of the volunteers are reflected in the next graph.

- **"At the end of the workday I feel exhausted"**
- **"I feel exhausted when I wake up in the morning and have to face another day of work"**
- **"I have a hard time relaxing after work"**



	At the end of the work day I feel exhausted	I feel exhausted when I wake up in the morning and have to face another day of work	I have a hard time relaxing after work
PRE Strongly disagree	0%	18%	28%
PRE Somewhat disagree	18%	27%	17%
PRE Neither disagree, nor agree	18%	9%	27%
PRE Somewhat agree	27%	36%	27%
PRE Strongly agree	36%	9%	0%
POST Strongly disagree	0%	23%	54%
POST Somewhat disagree	8%	15%	23%
POST Neither disagree, nor agree	23%	23%	15%
POST Somewhat agree	38%	31%	8%
POST Strongly agree	31%	8%	0%

Figure 100: Cognitive/Mental status 2. Production use case.

2.3.3.2.5 *Affective/Emotional status*

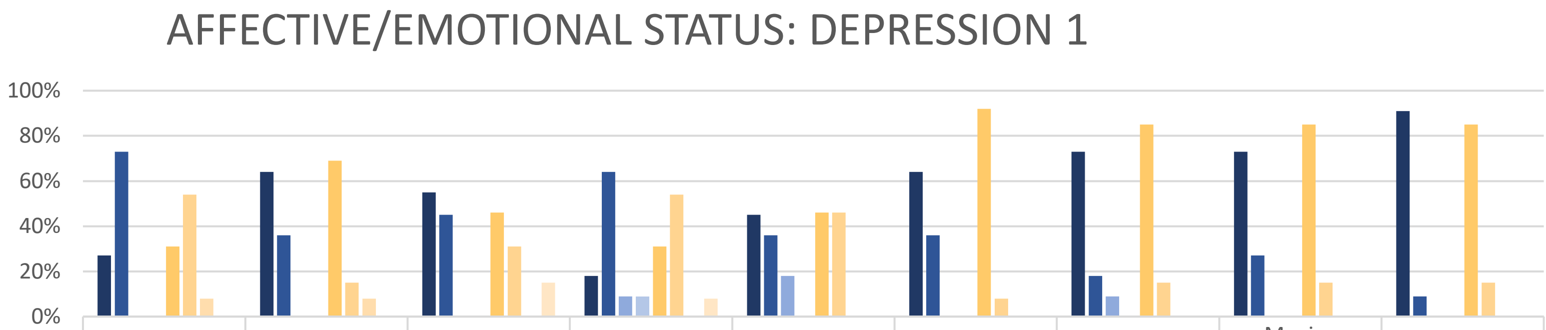
To assess the affective and emotional status of the production participants, they were asked first with ten questions, that collect their general state with respect to a possible depression. They were asked if over the last 2 weeks, how often have you been bothered by any of the following problems:

- "Little interest or pleasure in doing things"
- "Feeling down, depressed, or hopeless"
- "Trouble falling or staying asleep, or sleeping too much"
- "Feeling tired or having little energy"
- "Poor appetite or overeating"
- "Feeling bad about yourself — or that you are a failure or have let yourself or your family down"
- "Trouble concentrating on things, such as reading the newspaper or watching television"
- "Moving or speaking so slowly that other people could have noticed. Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual"
- "Thoughts that you would be better off dead or of hurting yourself in some way"

The participants answer to these questions can be seen in the next graph: Depression 1.

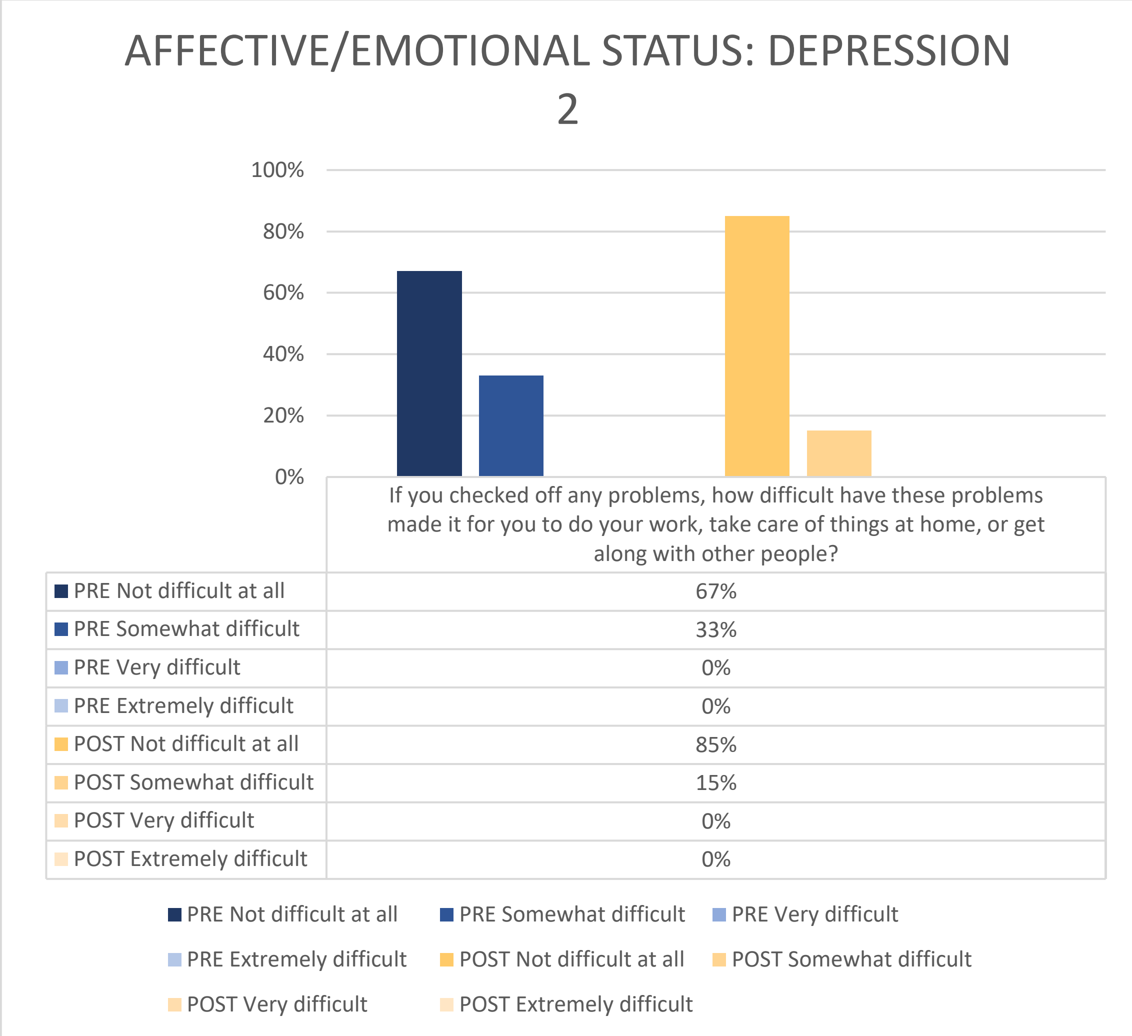
There is one final question reflected in the Affective/Emotional status: Depression 2.

- "If you checked off any problems, how difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?"



	Little interest or pleasure in doing things	Feeling down, depressed, or hopeless	Trouble falling or staying asleep, or sleeping too much	Feeling tired or having little energy	Poor appetite or overeating	Feeling bad about yourself — or that you are a failure or have let yourself or your family down	Trouble concentrating on things, such as reading the newspaper or watching television	Moving or speaking so slowly that other people could have noticed. Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual	Thoughts that you would be better off dead or of hurting yourself in some way
PRE Not at all	27%	64%	55%	18%	45%	64%	73%	73%	91%
PRE Several days	73%	36%	45%	64%	36%	36%	18%	27%	9%
PRE More than half of the days	0%	0%	0%	9%	18%	0%	9%	0%	0%
PRE Nearly every day	0%	0%	0%	9%	0%	0%	0%	0%	0%
POST Not at all	31%	69%	46%	31%	46%	92%	85%	85%	85%
POST Several days	54%	15%	31%	54%	46%	8%	15%	15%	15%
POST More than half of the days	8%	8%	0%	0%	0%	0%	0%	0%	0%
POST Nearly every day	0%	0%	15%	8%	0%	0%	0%	0%	0%

Figure 101: Affective/Emotional status: Depression. Production use case.



	If you checked off any problems, how difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?
PRE Not difficult at all	67%
PRE Somewhat difficult	33%
PRE Very difficult	0%
PRE Extremely difficult	0%
POST Not difficult at all	85%
POST Somewhat difficult	15%
POST Very difficult	0%
POST Extremely difficult	0%

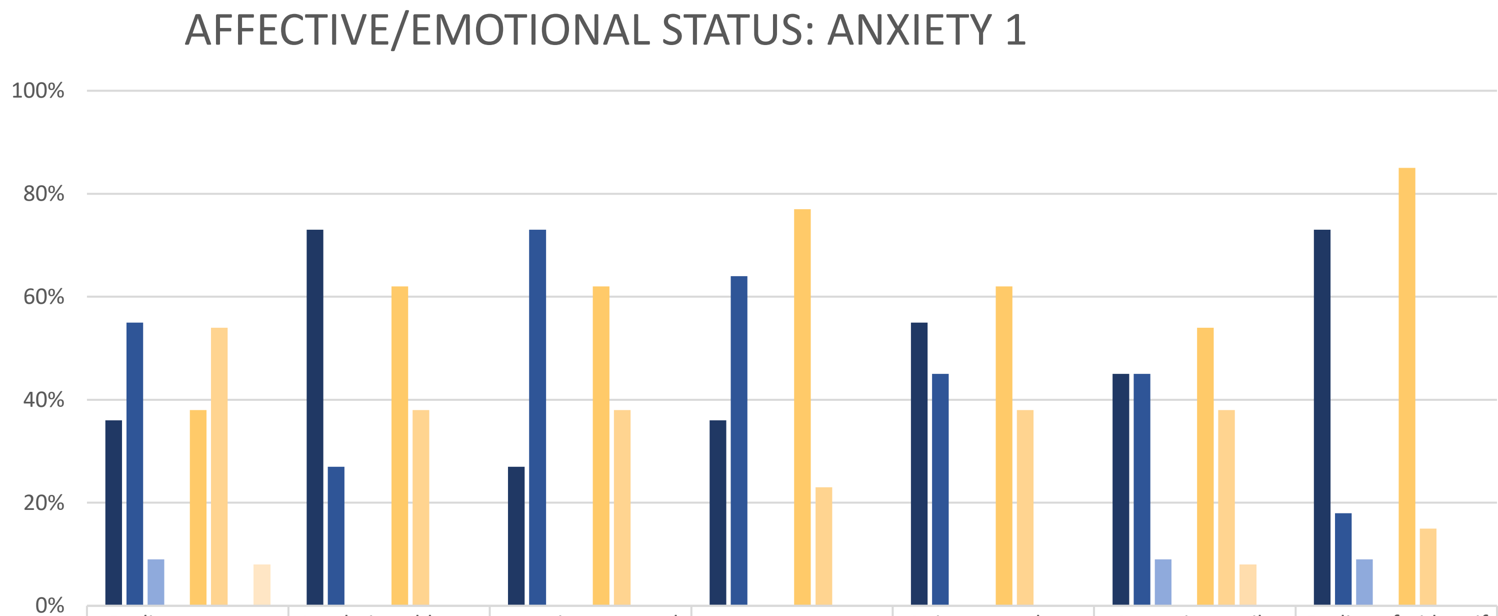
Figure 102: Affective/Emotional status: Depression 2. Production use case.

Continuing with the affective/emotional status assessment, the participants were asked with eight questions, that collect their general state with respect to a possible anxiety problem. As in the previous section, they were asked if over the last 2 weeks, how often have you been bothered by any of the following problems, and this is reflected in the next graph:

- “Feeling nervous, anxious, or on edge”
- “Not being able to stop or control worrying”
- “Worrying too much about different things”
- “Trouble relaxing”
- “Being so restless that it is hard to sit still”
- “Becoming easily annoyed or irritable”
- “Feeling afraid, as if something awful might happen”

One final question completes this evaluation and is reflected in the Affective/Emotional status: Anxiety 2.

- “If you checked off any problems, how difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?”



	Feeling nervous, anxious, or on edge	Not being able to stop or control worrying	Worrying too much about different things	Trouble relaxing	Being so restless that it is hard to sit still	Becoming easily annoyed or irritable	Feeling afraid, as if something awful might happen
PRE Not at all	36%	73%	27%	36%	55%	45%	73%
PRE Several days	55%	27%	73%	64%	45%	45%	18%
PRE More than half of the days	9%	0%	0%	0%	0%	9%	9%
PRE Nearly every day	0%	0%	0%	0%	0%	0%	0%
POST Not at all	38%	62%	62%	77%	62%	54%	85%
POST Several days	54%	38%	38%	23%	38%	38%	15%
POST More than half of the days	0%	0%	0%	0%	0%	8%	0%
POST Nearly every day	8%	0%	0%	0%	0%	0%	0%

Figure 103: Affective/Emotional status: Anxiety 1. Production use case.

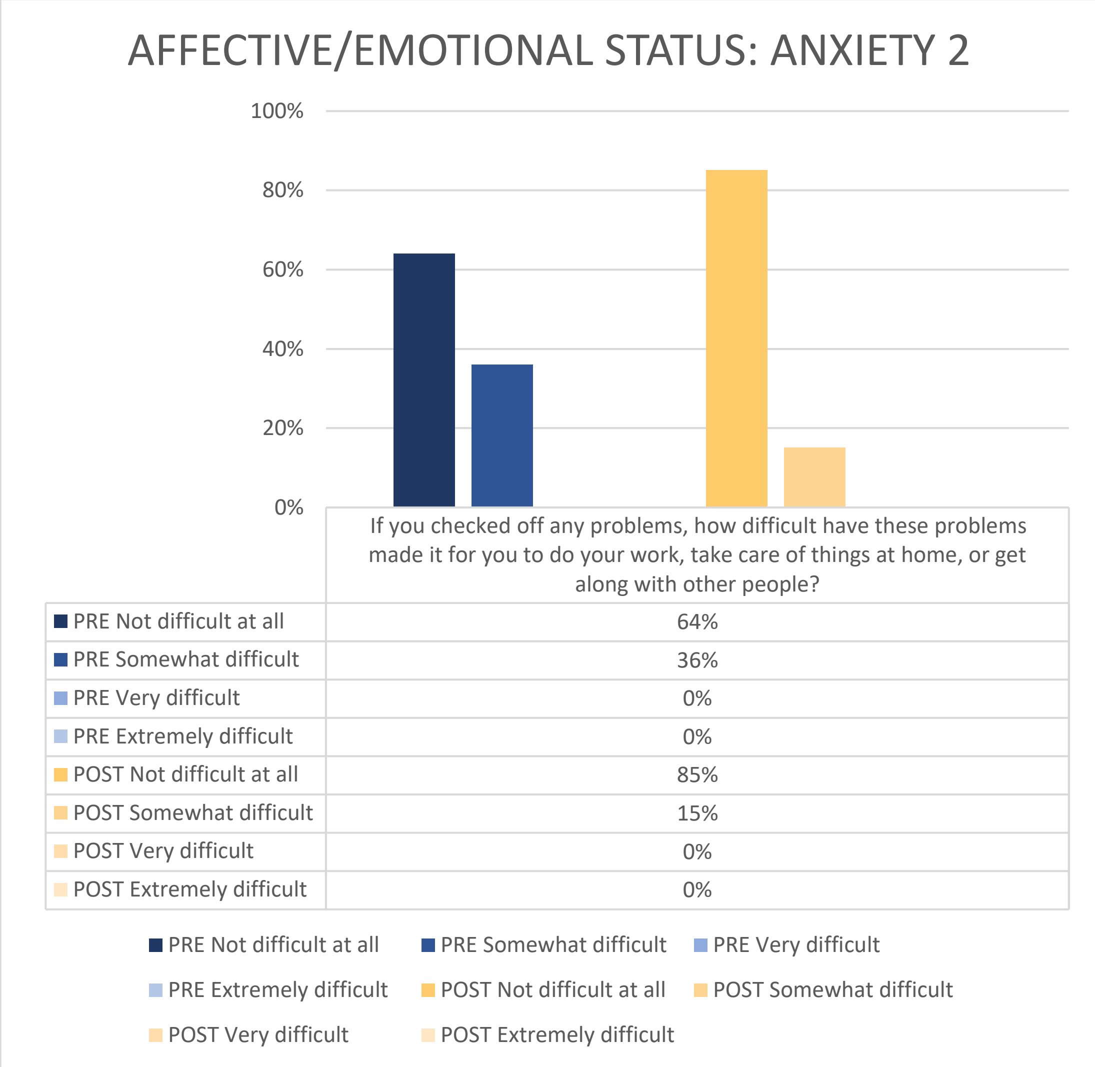


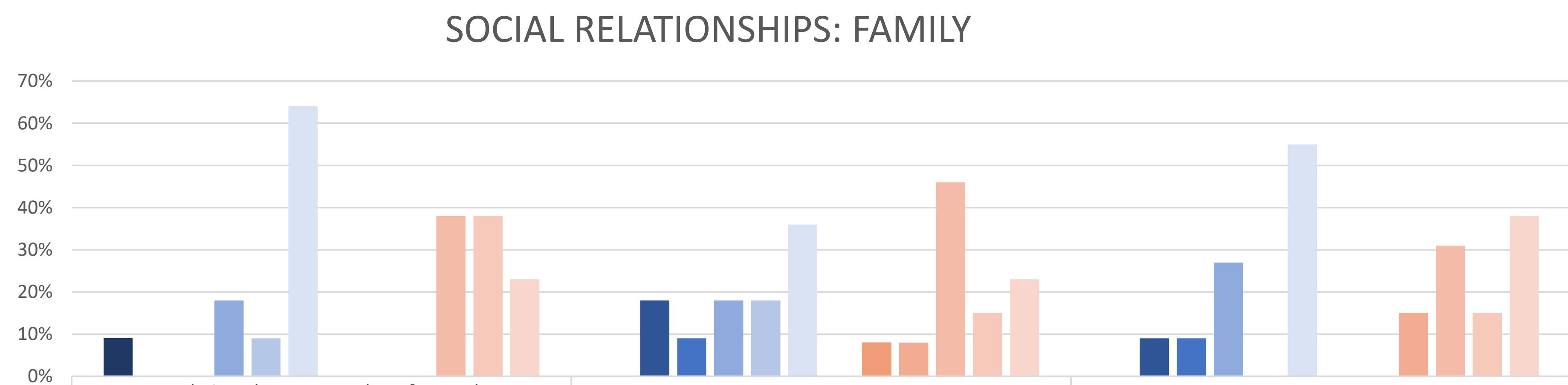
Figure 104: Affective/Emotional status: Anxiety 2. Production use case.

2.3.3.2.6 Social relationships

For the evaluation of the social relationships of the participants, they were asked first about their family and later about their friends.

The next graph shows how is the relationship between the participants and their relatives, with three questions.

- “How many relatives do you see or hear from at least once a month?”
- “How many relatives do you feel at ease with that you can talk about private matters?”
- “How many relatives do you feel close to such that you could call on them for help?”

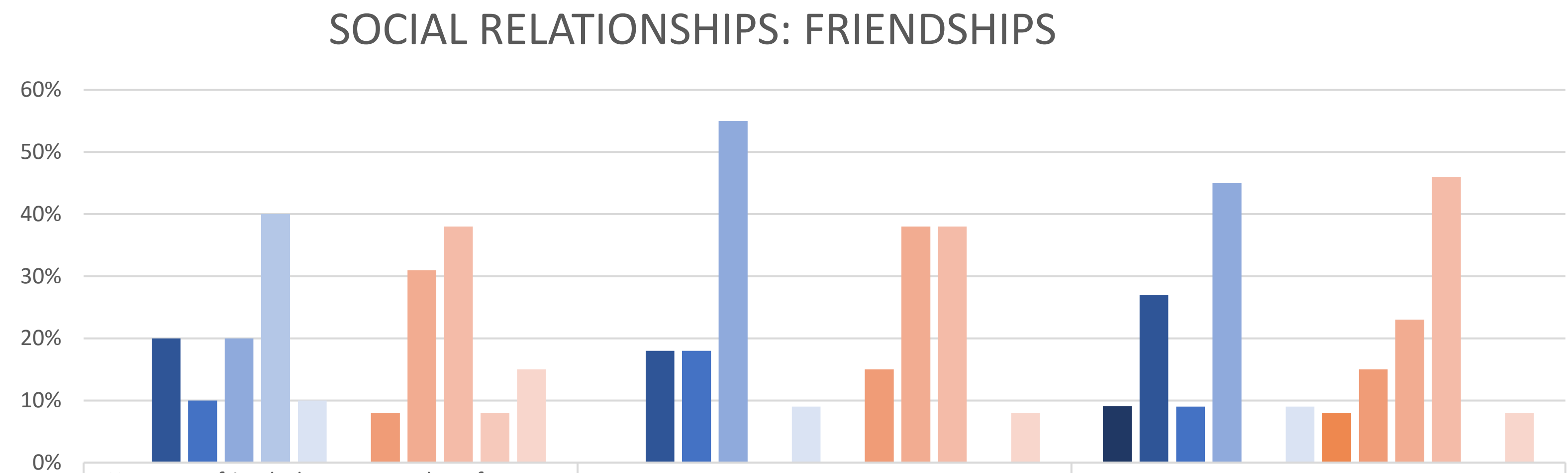


	How many relatives do you see or hear from at least once a month?The level of effort or mental concentration that my job requires is	How many relatives do you feel at ease with that you can talk about private matters?	How many relatives do you feel close to such that you could call on them for help?
PRE NONE	9%	0%	0%
PRE ONE	0%	18%	9%
PRE TWO	0%	9%	9%
PRE THREE OR FOUR	18%	18%	27%
PRE FIVE TO EIGHT	9%	18%	0%
PRE NINE OR MORE	64%	36%	55%
POST NONE	0%	0%	0%
POST ONE	0%	8%	0%
POST TWO	0%	8%	15%
POST THREE OR FOUR	38%	46%	31%
POST FIVE TO EIGHT	38%	15%	15%
POST NINE OR MORE	23%	23%	38%

Figure 105: Social relationships. Family. Production use case.

The next graph shows how is the relationship between the participants and their friends, with the same three questions used for their relatives.

- "How many friends do you see or hear from at least once a month?"
- "How many friends do you feel at ease with that you can talk about private matters?"
- "How many friends do you feel close to such that you could call on them for help?"



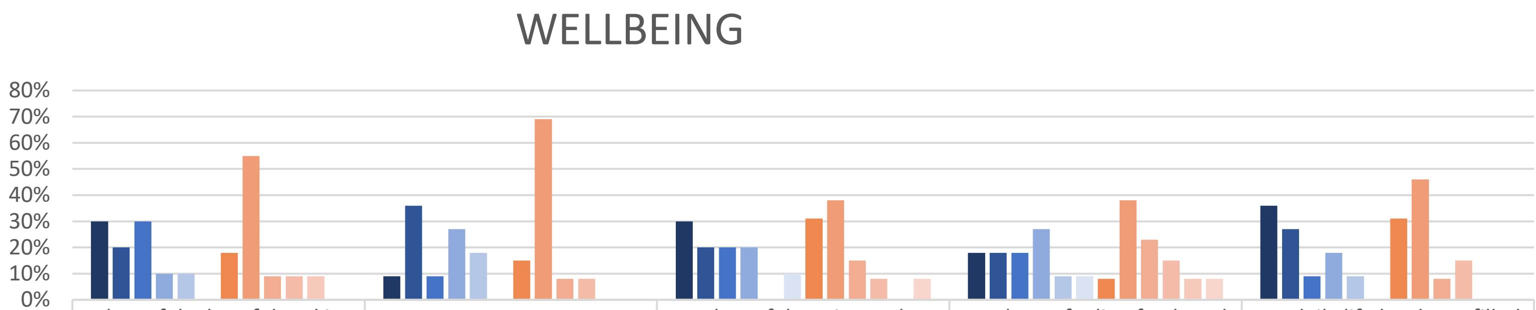
	How many friends do you see or hear from at least once a month?The level of effort or mental concentration that my job requires is	How many friends do you feel at ease with that you can talk about private matters?	How many friends do you feel close to such that you could call on them for help?
PRE NONE	0%	0%	9%
PRE ONE	20%	18%	27%
PRE TWO	10%	18%	9%
PRE THREE OR FOUR	20%	55%	45%
PRE FIVE TO EIGHT	40%	0%	0%
PRE NINE OR MORE	10%	9%	9%
POST NONE	0%	0%	8%
POST ONE	8%	15%	15%
POST TWO	31%	38%	23%
POST THREE OR FOUR	38%	38%	46%
POST FIVE TO EIGHT	8%	0%	0%
POST NINE OR MORE	15%	8%	8%

Figure 106: Social relationships. Friendships. Production use case.

2.3.3.2.7 General wellbeing & Quality of life

The volunteers were asked about their subjective wellbeing and quality of life, firstly with five questions that can be seen in detail in the next graph.

- "I have felt cheerful and in good spirits"
- "I have felt calm and relaxed"
- "I have felt active and vigorous"
- "I woke up feeling fresh and rested"
- "My daily life has been filled with things that interest me"



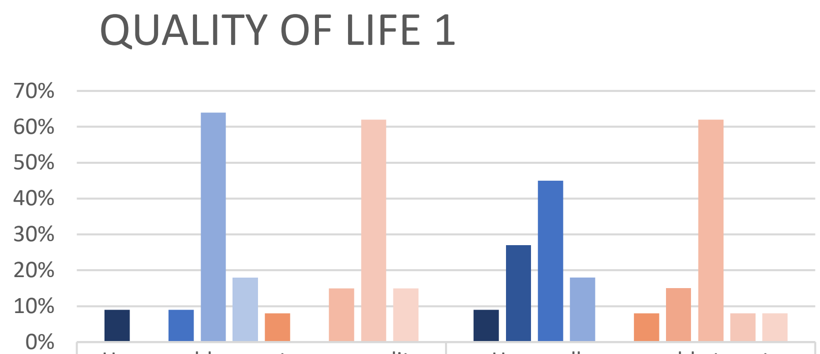
	I have felt cheerful and in good spirits	I have felt calm and relaxed	I have felt active and vigorous	I woke up feeling fresh and rested	My daily life has been filled with things that interest me
PRE All the time	30%	9%	30%	18%	36%
PRE Most of the time	20%	36%	20%	18%	27%
PRE More than half the time	30%	9%	20%	18%	9%
PRE Less than half the time	10%	27%	20%	27%	18%
PRE Some of the time	10%	18%	0%	9%	9%
PRE At no time	0%	0%	10%	9%	0%
POST All the time	18%	15%	31%	8%	31%
POST Most of the time	55%	69%	38%	38%	46%
POST More than half the time	9%	8%	15%	23%	8%
POST Less than half the time	9%	8%	8%	15%	15%
POST Some of the time	9%	0%	0%	8%	0%
POST At no time	0%	0%	8%	8%	0%

PRE All the time   PRE Most of the time   PRE More than half the time   PRE Less than half the time
PRE Some of the time   PRE At no time   POST All the time   POST Most of the time
POST More than half the time   POST Less than half the time   POST Some of the time   POST At no time

Figure 107: Wellbeing. Production use case

The next questions assess the quality of life of the participants. Can be seen in detail in the next six graphs.

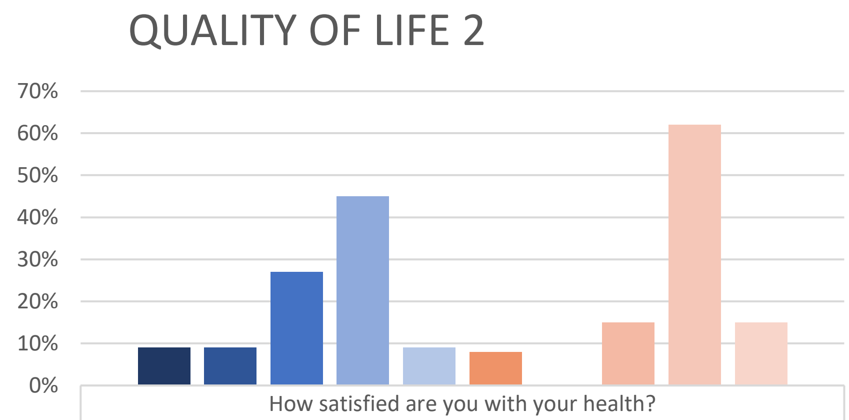
- "How would you rate your quality of life?"
- "How well are you able to get around?"



	How would you rate your quality of life?	How well are you able to get around?
PRE Very poor	9%	9%
PRE Poor	0%	27%
PRE Neither poor nor good	9%	45%
PRE Good	64%	18%
PRE Very good	18%	0%
POST Very poor	8%	8%
POST Poor	0%	15%
POST Neither poor nor good	15%	62%
POST Good	62%	8%
POST Very good	15%	8%

Figure 108: Quality of life 1. Production use case.

- "How satisfied are you with your health?"

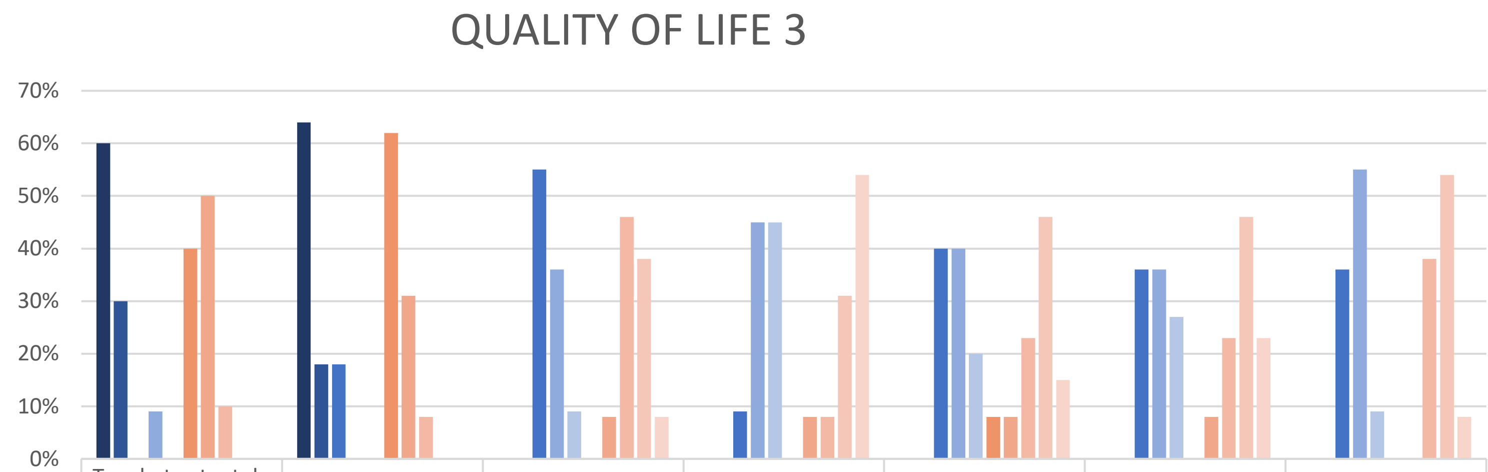


	How satisfied are you with your health?
PRE Very dissatisfied	9%
PRE Dissatisfied	9%
PRE Neither satisfied nor dissatisfied	27%
PRE Satisfied	45%
PRE Very satisfied	9%
POST Very dissatisfied	8%
POST Dissatisfied	0%
POST Neither satisfied nor dissatisfied	15%
POST Satisfied	62%
POST Very satisfied	15%

Figure 109: Quality of life 2. Production use case.

Next, you will see seven more questions that assess the quality of life of the WAOW tool users. The details are shown in the following image.

- "To what extent do you feel that physical pain prevents you from doing what you need to do?"
- "How much do you need any medical treatment to function in your daily life?"
- "How much do you enjoy life?"
- "To what extent do you feel your life to be meaningful?"
- "How well are you able to concentrate?"
- "How safe do you feel in your daily life?"
- "How healthy is your physical environment?"

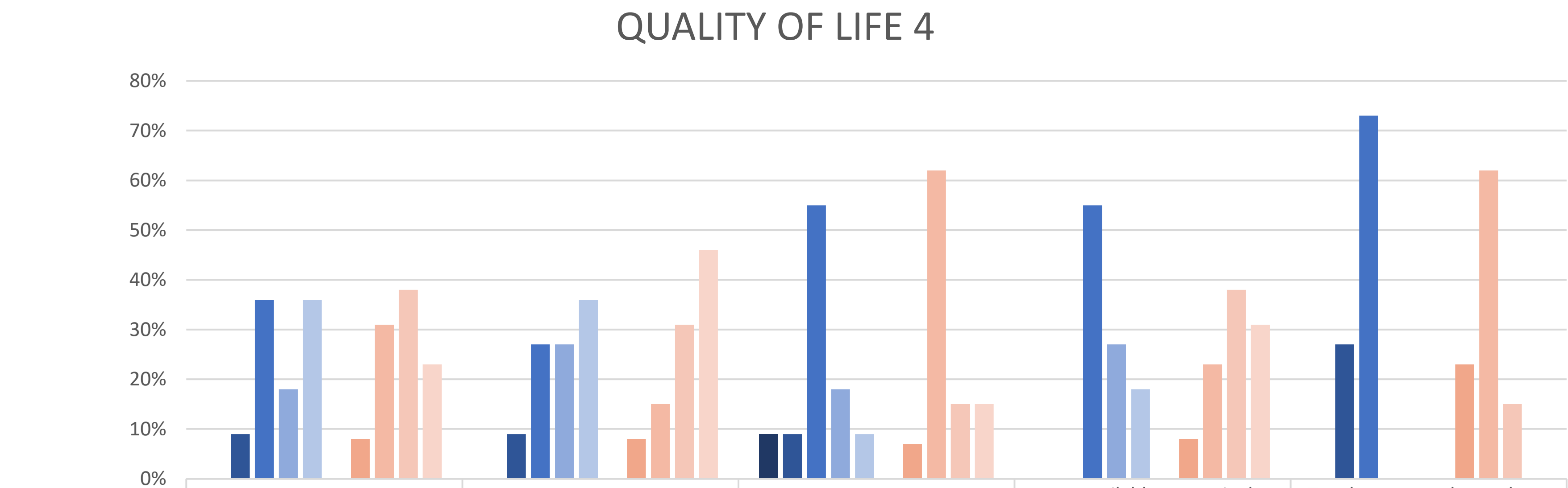


	To what extent do you feel that physical pain prevents you from doing what you need to do?	How much do you need any medical treatment to function in your daily life?	How much do you enjoy life?	To what extent do you feel your life to be meaningful?	How well are you able to concentrate?	How safe do you feel in your daily life?	How healthy is your physical environment?
PRE Not at all	60%	64%	0%	0%	0%	0%	0%
PRE A little	30%	18%	0%	0%	0%	0%	0%
PRE A moderate amount	0%	18%	55%	9%	40%	36%	36%
PRE Very much	9%	0%	36%	45%	40%	36%	55%
PRE An extreme amount	0%	0%	9%	45%	20%	27%	9%
POST Not at all	40%	62%	0%	0%	8%	0%	0%
POST A little	50%	31%	8%	8%	8%	8%	0%
POST A moderate amount	10%	8%	46%	8%	23%	23%	38%
POST Very much	0%	0%	38%	31%	46%	46%	54%
POST An extreme amount	0%	0%	8%	54%	15%	23%	8%

Figure 110: Quality of life 3. Production use case.

Another five questions that evaluate the quality of life are shown in the next graph: “Quality of life 4”.

- “Do you have enough energy for everyday life?”
- “Are you able to accept your bodily appearance?”
- “Have you enough money to meet your needs?”
- “How available to you is the information that you need in your day-to-day life?”
- “To what extent do you have the opportunity for leisure activities?”

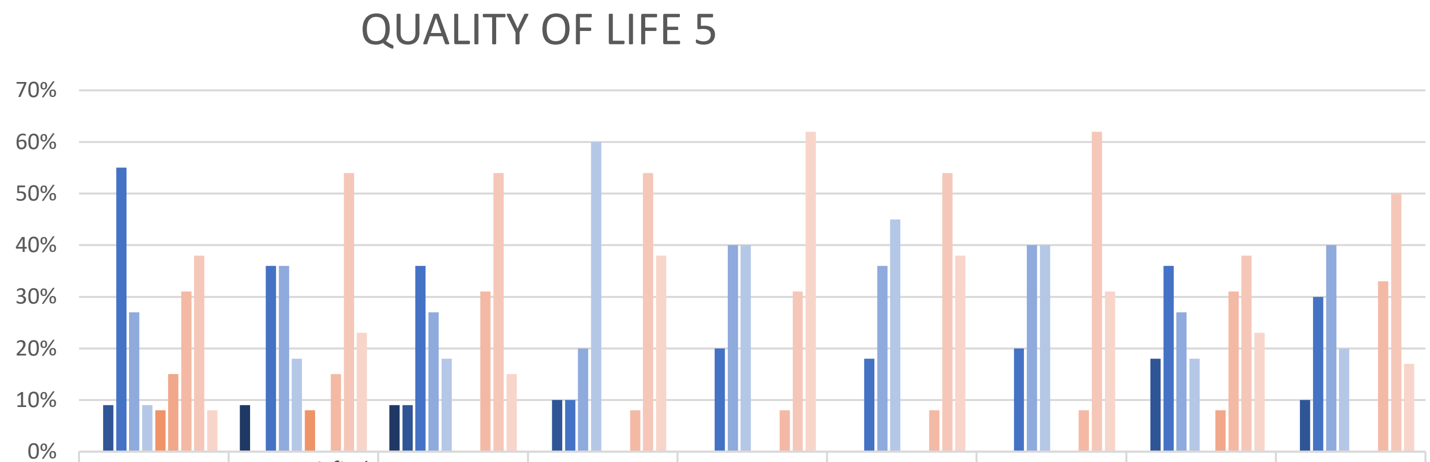


	Do you have enough energy for everyday life?	Are you able to accept your bodily appearance?	Have you enough money to meet your needs?	How available to you is the information that you need in your day-to-day life?	To what extent do you have the opportunity for leisure activities?
PRE Not at all	0%	0%	9%	0%	0%
PRE A little	9%	9%	9%	0%	27%
PRE Moderately	36%	27%	55%	55%	73%
PRE Mostly	18%	27%	18%	27%	0%
PRE Completely	36%	36%	9%	18%	0%
POST Not at all	0%	0%	0%	0%	0%
POST A little	8%	8%	7%	8%	23%
POST Moderately	31%	15%	62%	23%	62%
POST Mostly	38%	31%	15%	38%	15%
POST Completely	23%	46%	15%	31%	0%

Figure 111: Quality of life 4. Production use case.

Continuing assessing the quality of life, the participants answered the next nine questions. The details can be seen in the next graph: “Quality of life 5”.

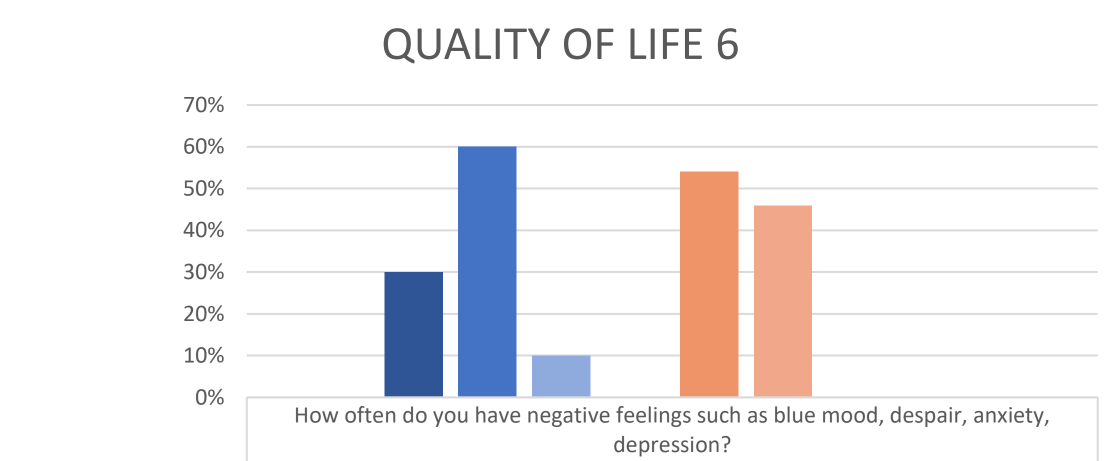
- “How satisfied are you with your sleep?”
- “How satisfied are you with your ability to perform your daily living activities?”
- “How satisfied are you with your capacity for work?”
- “How satisfied are you with yourself?”
- “How satisfied are you with your personal relationships?”
- “How satisfied are you with the support you get from your friends?”
- “How satisfied are you with the conditions of your living place?”
- “How satisfied are you with your access to health services?”
- “How satisfied are you with your transport?”



	How satisfied are you with your sleep?	How satisfied are you with your ability to perform your daily living activities?	How satisfied are you with your capacity for work?	How satisfied are you with yourself?	How satisfied are you with your personal relationships?	How satisfied are you with the support you get from your friends?	How satisfied are you with the conditions of your living place?	How satisfied are you with your access to health services?	How satisfied are you with your transport?
PRE Very dissatisfied	0%	9%	9%	0%	0%	0%	0%	0%	0%
PRE Dissatisfied	9%	0%	9%	10%	0%	0%	0%	18%	10%
PRE Neither satisfied nor dissatisfied	55%	36%	36%	10%	20%	18%	20%	36%	30%
PRE Satisfied	27%	36%	27%	20%	40%	36%	40%	27%	40%
PRE Very satisfied	9%	18%	18%	60%	40%	45%	40%	18%	20%
POST Very dissatisfied	8%	8%	0%	0%	0%	0%	0%	0%	0%
POST Dissatisfied	15%	0%	0%	0%	0%	0%	0%	8%	0%
POST Neither satisfied nor dissatisfied	31%	15%	31%	8%	8%	8%	8%	31%	33%
POST Satisfied	38%	54%	54%	54%	31%	54%	62%	38%	50%
POST Very satisfied	8%	23%	15%	38%	62%	38%	31%	23%	17%

Figure 112: Quality of life 5. Production use case.

The last question to assess the quality of life of the participants was "How often do you have negative feelings such as blue mood, despair, anxiety, depression?" The participants selected the following frequency:



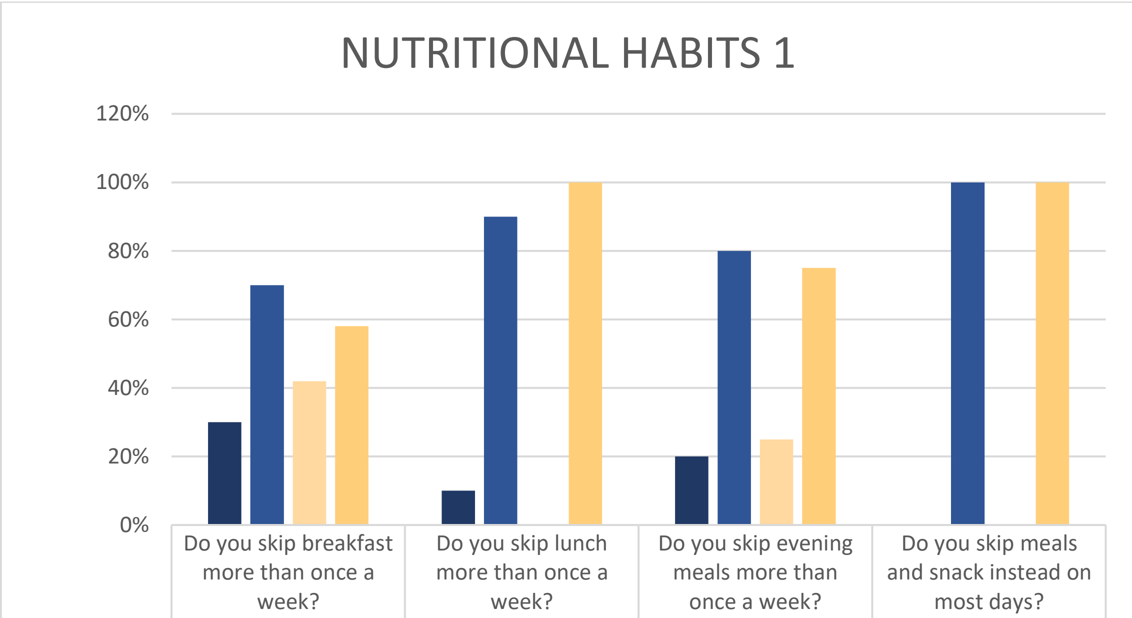
	How often do you have negative feelings such as blue mood, despair, anxiety, depression?
PRE Never	0%
PRE Seldom	30%
PRE Quite often	60%
PRE Very often	10%
PRE Always	0%
POST Never	54%
POST Seldom	46%
POST Quite often	0%
POST Very often	0%
POST Always	0%

Figure 113: Quality of life 6. Production use case.

For assessing the nutritional habits, very important in the general wellbeing of the participants, they were asked 30 questions that they had to answer with a simple yes or no.

The first four questions are related with the frequency they skip different meals during the week. A positive answer will mean a bad nutritional habit, while a negative will mean a good one. See the graph "Nutritional habits 1" below.

- "Do you skip breakfast more than once a week?"
- "Do you skip lunch more than once a week?"
- "Do you skip evening meals more than once a week?"
- "Do you skip meals and snack instead on most days?"

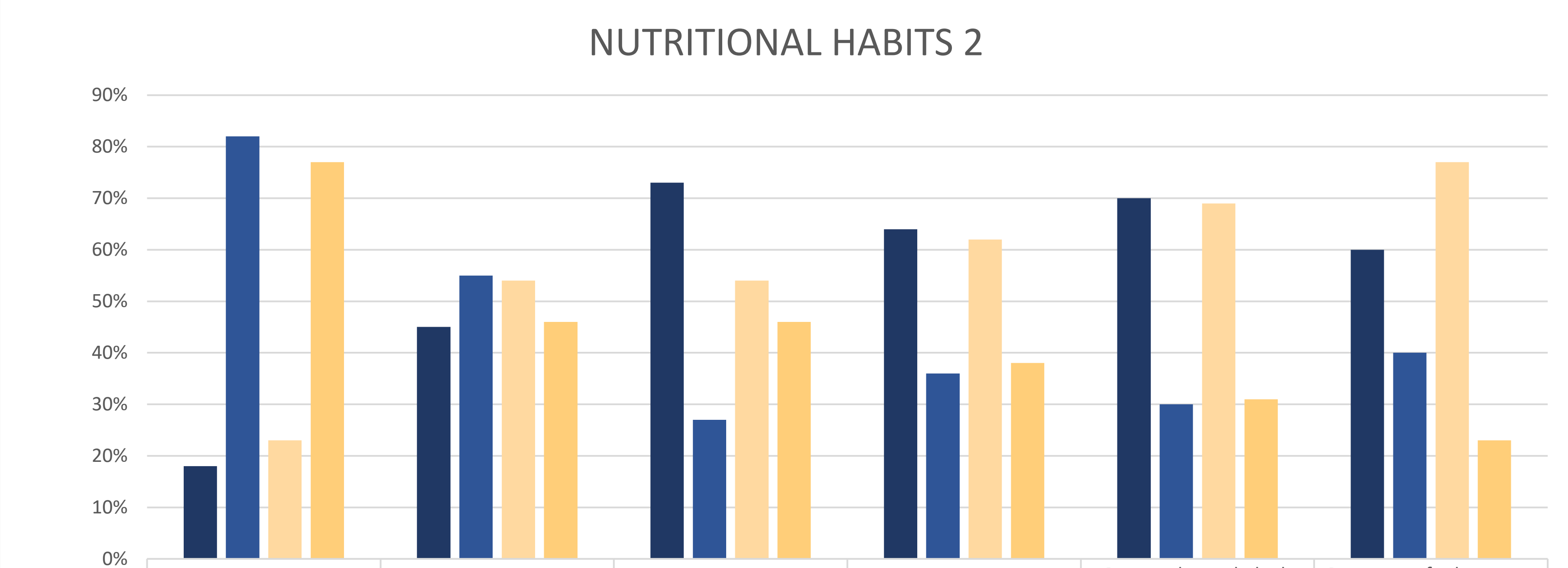


	Do you skip breakfast more than once a week?	Do you skip lunch more than once a week?	Do you skip evening meals more than once a week?	Do you skip meals and snack instead on most days?
PRE YES	30%	10%	20%	0%
PRE NO	70%	90%	80%	100%
POST YES	42%	0%	25%	0%
POST NO	58%	100%	75%	100%

Figure 114: Nutritional habits 1. Production use case

The next six questions are about the choice of the volunteers when they select their food or ingredients. A positive answer will mean a good nutritional habit, while a negative will mean a bad one. See the graph "Nutritional habits 2" below.

- "Do you eat more than 5 portions of fruit and/or vegetables every day?"
- "Do you eat more than 4 different varieties of fruit each week?"
- "Do you eat more than 4 different varieties of vegetables each week?"
- "Do you choose low-fat products when available?"
- "Do you choose baked, steamed, or grilled options when available, rather than fried foods (such as crisps and snacks, or fish and chips)?"
- "Do you opt for lean cuts of meat or remove visible fat – for example, removing the skin on chicken or the rind on bacon?"

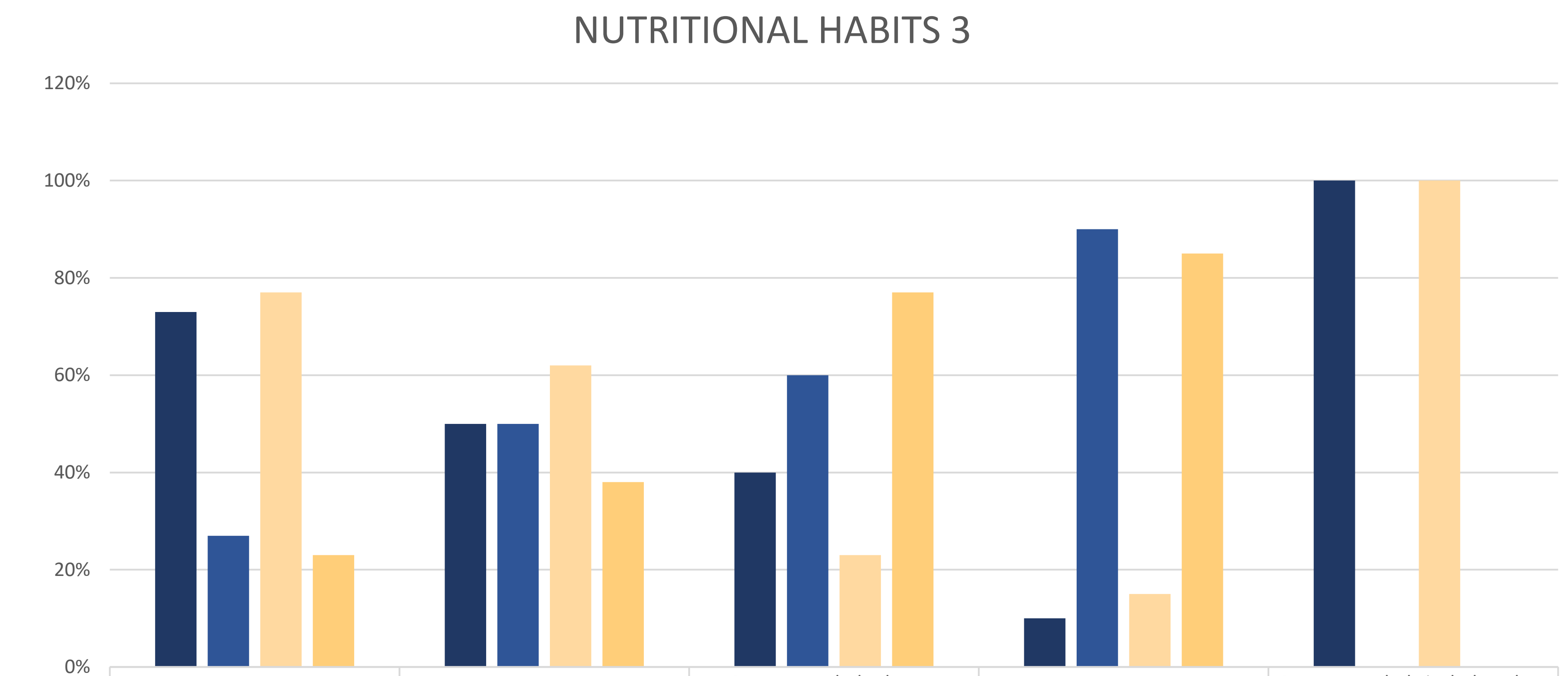


	Do you eat more than 5 portions of fruit and/or vegetables every day?	Do you eat more than 4 different varieties of fruit each week?	Do you eat more than 4 different varieties of vegetables each week?	Do you choose low-fat products when available?	Do you choose baked, steamed or grilled options when available, rather than fried foods (such as crisps and snacks, or fish and chips)?	Do you opt for lean cuts of meat or remove visible fat – for example, removing the skin on chicken or the rind on bacon?
PRE YES	18%	45%	73%	64%	70%	60%
PRE NO	82%	55%	27%	36%	30%	40%
POST YES	23%	54%	54%	62%	69%	77%
POST NO	77%	46%	46%	38%	31%	23%

Figure 115: Nutritional habits 2. Production use case

The next five questions are about the diet habits of the volunteers. A positive answer will mean a good nutritional habit, while a negative will mean a bad one. The details are shown in the graph “Nutritional habits 3” below.

- “Did you eat any oily fish last week?”
- “Do you include some unsalted nuts and seeds in your diet?”
- “Do you regularly choose wholemeal bread or rolls rather than white?”
- “Do you regularly eat wholegrain cereals, with no added sugar?”
- “Do you regularly include pulses in your diet? For example, beans and lentils”

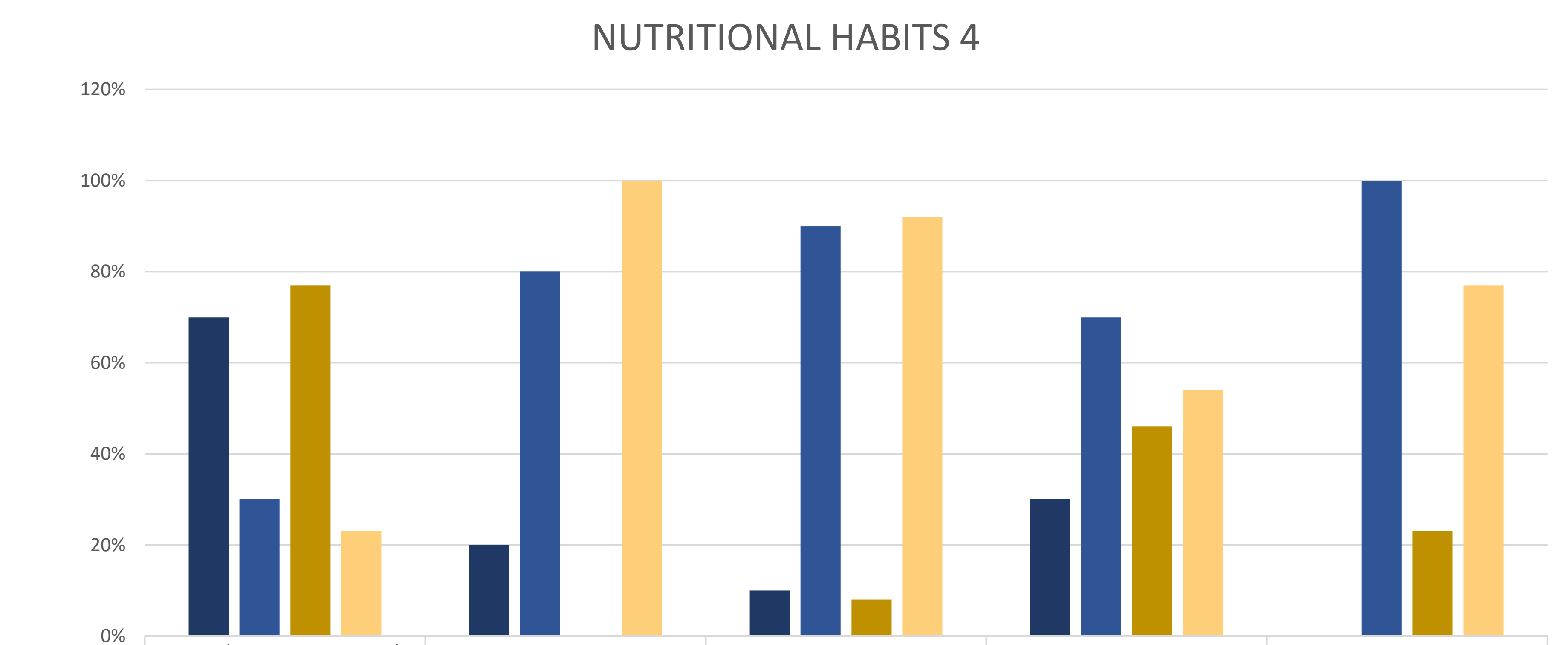


	Did you eat any oily fish last week?	Do you include some unsalted nuts and seeds in your diet?	Do you regularly choose wholemeal bread or rolls rather than white?	Do you regularly eat wholegrain cereals, with no added sugar?	Do you regularly include pulses in your diet? For example, beans and lentils.
PRE YES	73%	50%	40%	10%	100%
PRE NO	27%	50%	60%	90%	0%
POST YES	77%	62%	23%	15%	100%
POST NO	23%	38%	77%	85%	0%

Figure 116: Nutritional habits 3. Production use case

The next five questions are about starchy foods, sugar and sweets. A positive answer will mean a bad nutritional habit, while a negative will mean a good one. The details are shown in the graph "Nutritional habits 4" below.

- "Do you base your main meals around starchy foods? For example, potatoes, pasta, rice or bread"
- "Do you regularly eat sugar-coated breakfast cereals or add sugar to your breakfast cereals?"
- "Do you add sugar to your drinks?"
- "Do you regularly drink sweet fizzy drinks?"
- "Do you regularly eat cakes, sweets, chocolate, or biscuits at work?"

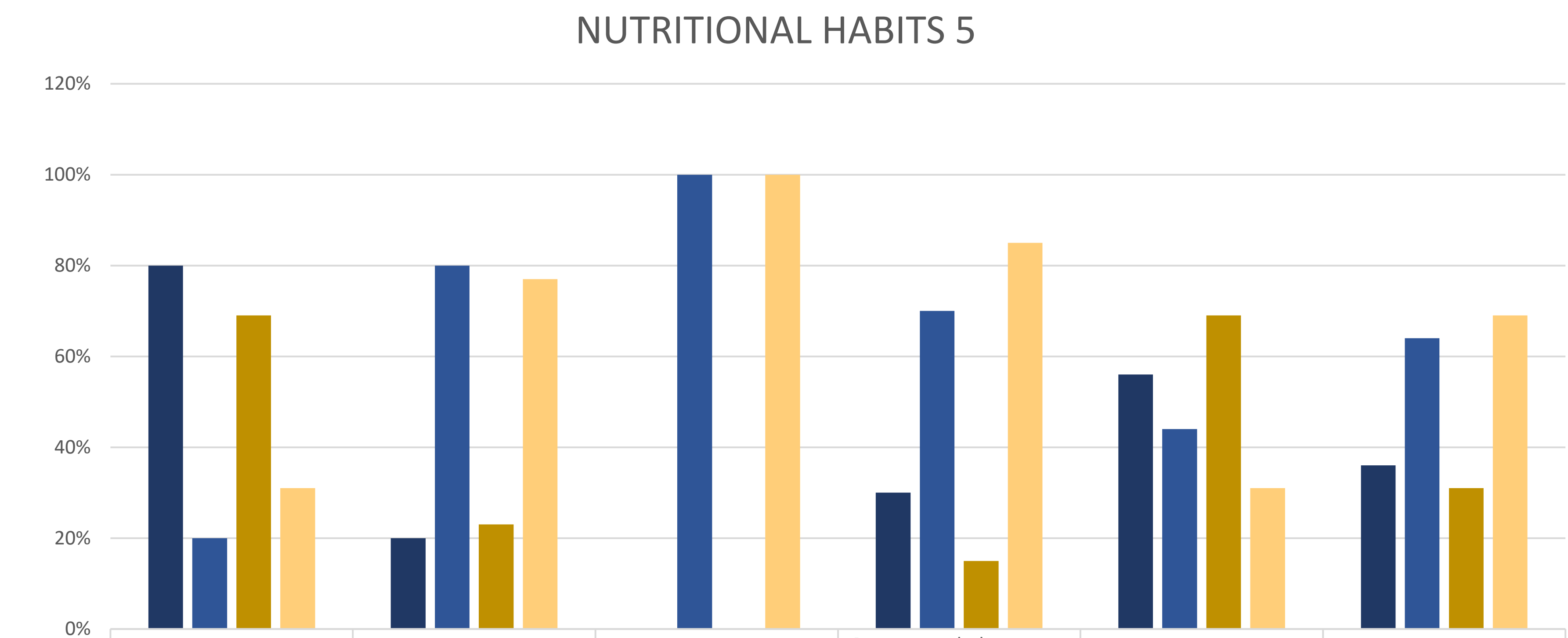


	Do you base your main meals around starchy foods? For example, potatoes, pasta, rice or bread.	Do you regularly eat sugar-coated breakfast cereals or add sugar to your breakfast cereals?	Do you add sugar to your drinks?	Do you regularly drink sweet fizzy drinks?	Do you regularly eat cakes, sweets, chocolate or biscuits at work?
PRE YES	70%	20%	10%	30%	0%
PRE NO	30%	80%	90%	70%	100%
POST YES	77%	0%	8%	46%	23%
POST NO	23%	100%	92%	54%	77%

Figure 117: Nutritional habits 4. Production use case

The next six questions are about bad nutritional habits and high blood pressure. A positive answer will mean a bad nutritional habit, while a negative will mean a good one. The details are shown in the graph “Nutritional habits 5” below.

- “Do you regularly add salt to food during cooking?”
- “Do you regularly add salt to meals at the table?”
- “Do you regularly eat savoury snacks at work? For example, crisps or salted nuts?”
- “Do you regularly eat pre-prepared meals? For example, pre-prepared sandwiches, ready meals, or canned soups?”
- “Do you regularly eat processed meats such as ham or bacon, or smoked fish?”
- “Has your GP advised you that you have high blood pressure?”

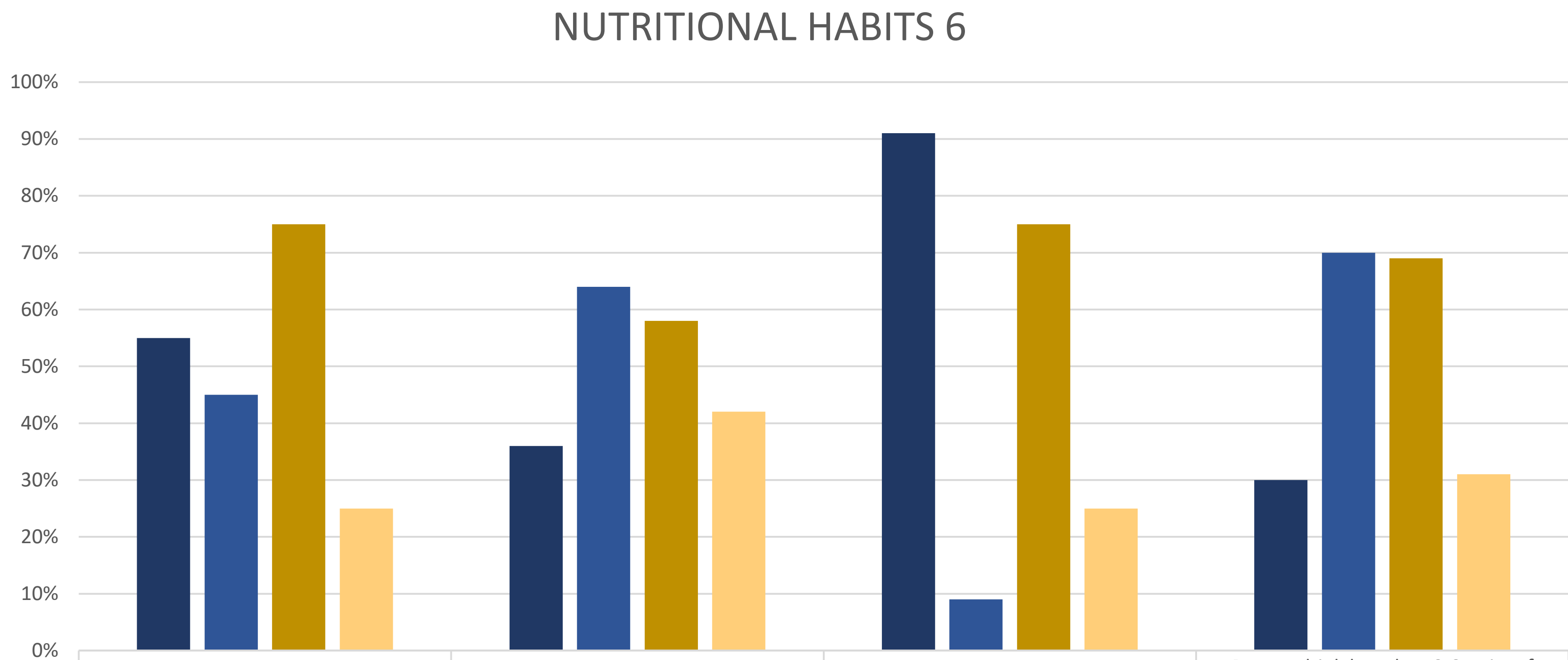


	Do you regularly add salt to food during cooking?	Do you regularly add salt to meals at the table?	Do you regularly eat savoury snacks at work? For example, crisps or salted nuts.	Do you regularly eat pre-prepared meals? For example, pre-prepared sandwiches, ready meals or canned soups.	Do you regularly eat processed meats such as ham or bacon, or smoked fish?	Has your GP advised you that you have high blood pressure?
PRE YES	80%	20%	0%	30%	56%	36%
PRE NO	20%	80%	100%	70%	44%	64%
POST YES	69%	23%	0%	15%	69%	31%
POST NO	31%	77%	100%	85%	31%	69%

Figure 118: Nutritional habits 5. Production use case

The last four questions are about drinks choice and drinking frequency. A positive answer will mean a good nutritional habit, while a negative will mean a bad one. The details are shown in the graph “Nutritional habits 6” below.

- “Do you drink plenty of fluids at regular intervals during the working day?”
- “Do you opt for a variety of different drinks, including water, at work?”
- “Do you avoid sugary fizzy drinks?”
- “Do you drink less than 2-3 units of alcohol a day if you’re a woman, or less than 3-4 units of alcohol a day if you’re a man? 1 unit of alcohol is equivalent to 100ml of 10% ABV (alcohol by volume)”



	Do you drink plenty of fluids at regular intervals during the working day?	Do you opt for a variety of different drinks, including water, at work?	Do you avoid sugary fizzy drinks?	Do you drink less than 2-3 units of alcohol a day if you're a woman, or less than 3-4 units of alcohol a day if you're a man? 1 unit of alcohol is equivalent to 100ml of 10% ABV (alcohol by volume)
PRE YES	55%	36%	91%	30%
PRE NO	45%	64%	9%	70%
POST YES	75%	58%	75%	69%
POST NO	25%	42%	25%	31%

Figure 119: Nutritional habits 6. Production use case

2.3.3.2.8 *Health-related physical activity*

To finish evaluating the health of the participants with respect to their physical activity, we have asked them 7 questions, about vigorous and moderate physical activities, walking and sitting time.

We consider them one by one in detail and with the corresponding graph.

- "During the last 7 days, on how many days did you do vigorous physical activities like heavy lifting, digging, aerobics, or fast bicycling?"

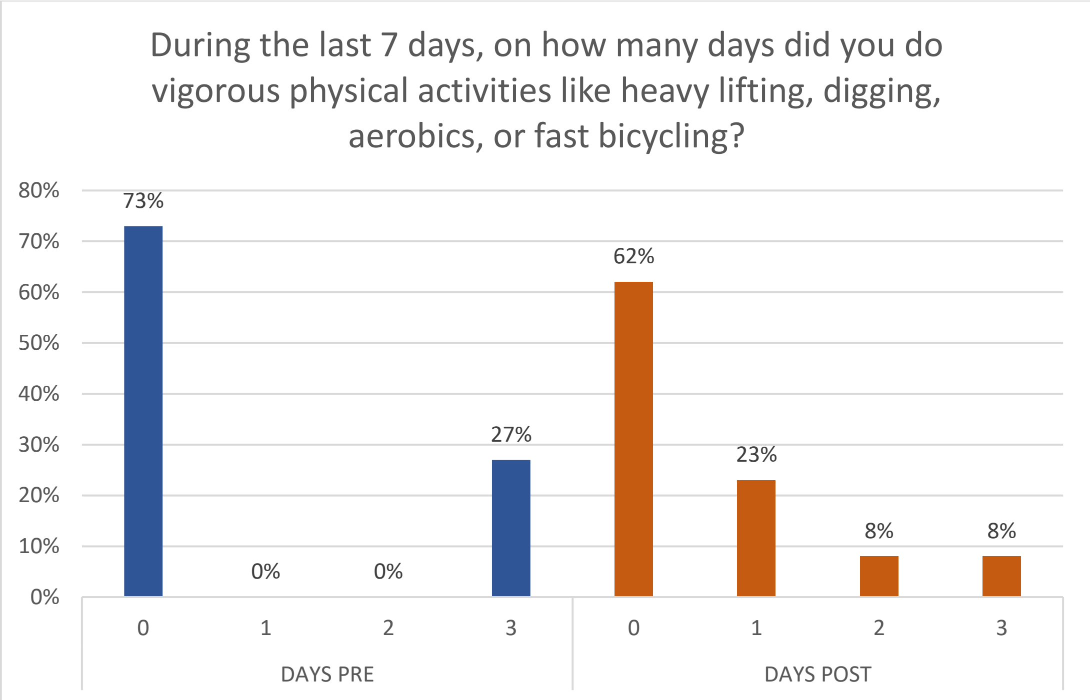


Figure 120: Health related vigorous physical activity 1. Production use case

- For those participants who answered yes to the previous question, we made an additional question: "How much time did you usually spend doing vigorous physical activities on one of those days?"

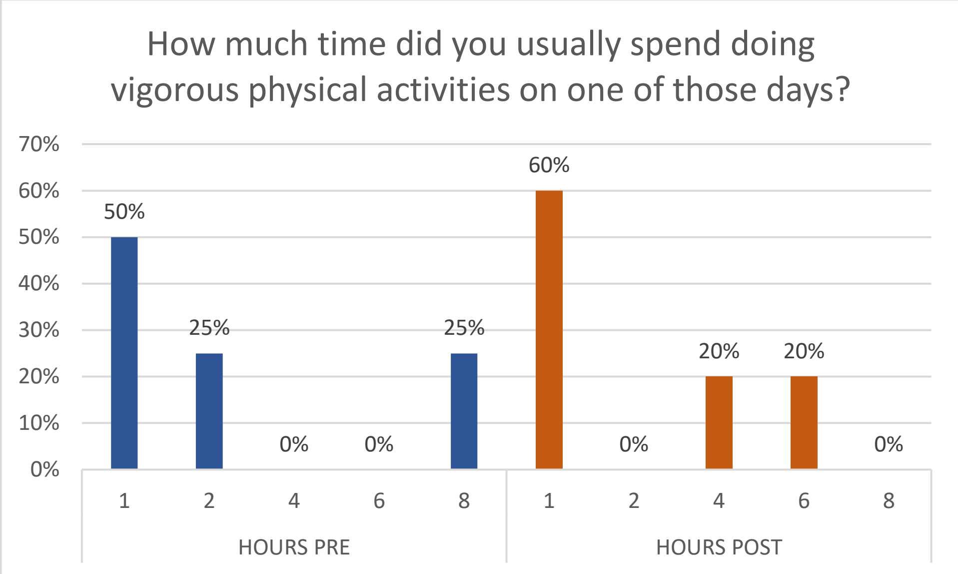


Figure 121: Health related vigorous physical activity 2. Production use case

- "During the last 7 days, on how many days did you do moderate physical activities like carrying light loads, bicycling at a regular pace, or doubles tennis?"

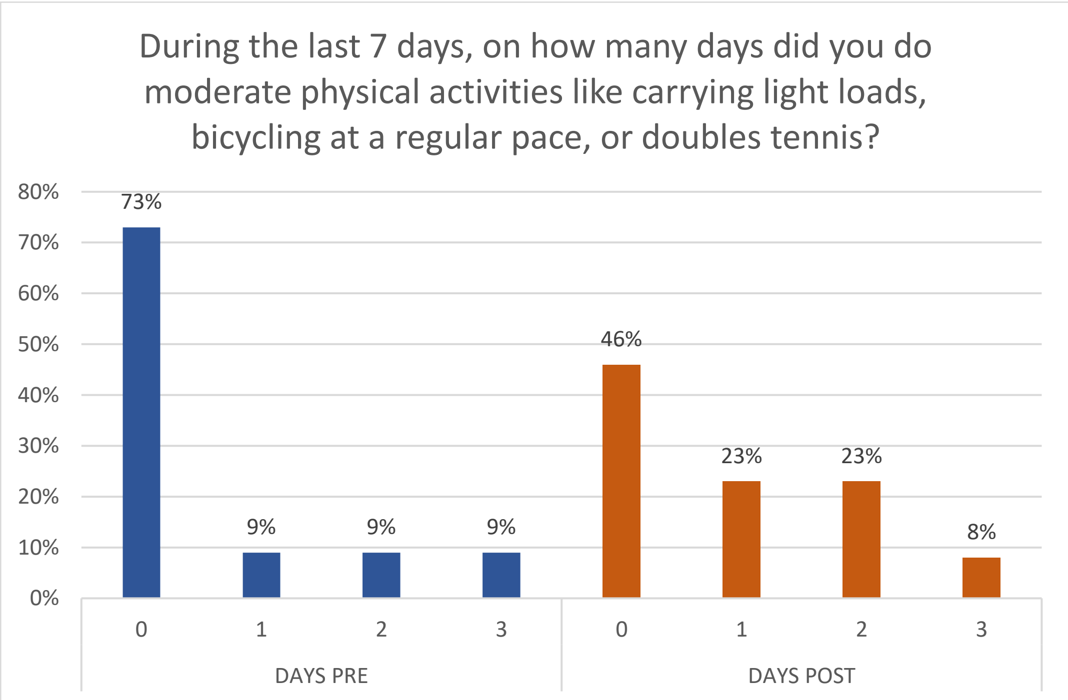


Figure 122: Health related moderate physical activity 1. Production use case

- For those volunteers who answered yes to the previous question, were asked "How much time did you usually spend doing moderate physical activities on one of those days?" and here are their answers in the pertinent graph:

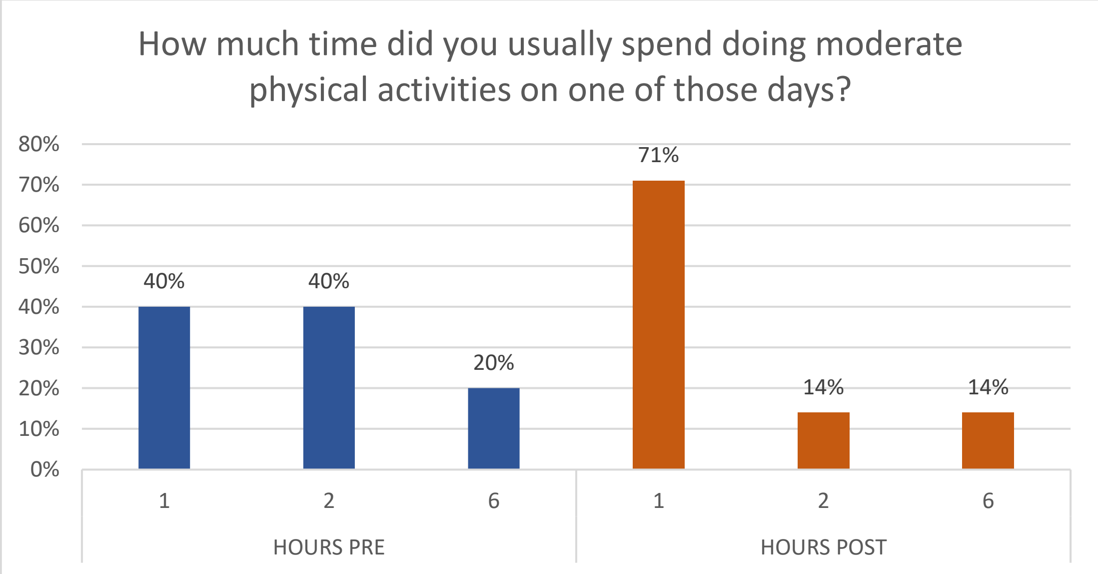


Figure 123: Health related moderate physical activity 2. Production use case.

- "During the last 7 days, on how many days did you walk for at least 10 minutes at a time?" the participants answered the next shown in the next graph:

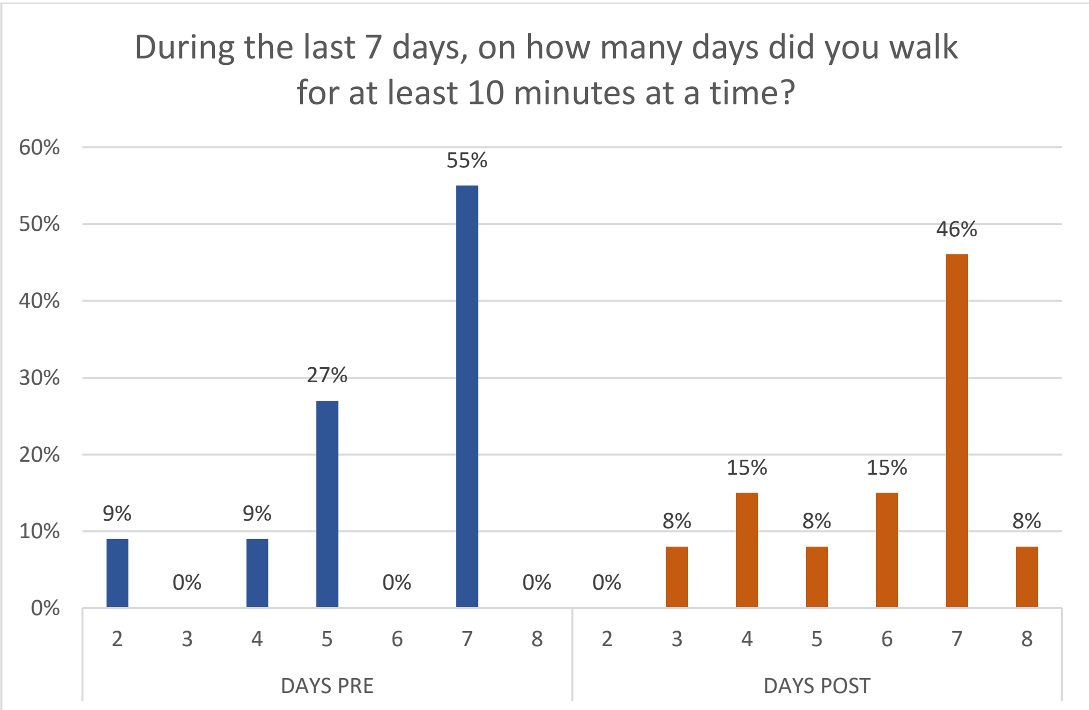


Figure 124: Health related walking physical activity 1. Production use case

- For those volunteers who answered yes to the previous question about walking, were asked about "How much time did you usually spend walking on one of those days?", and here are their answers in the related graph:

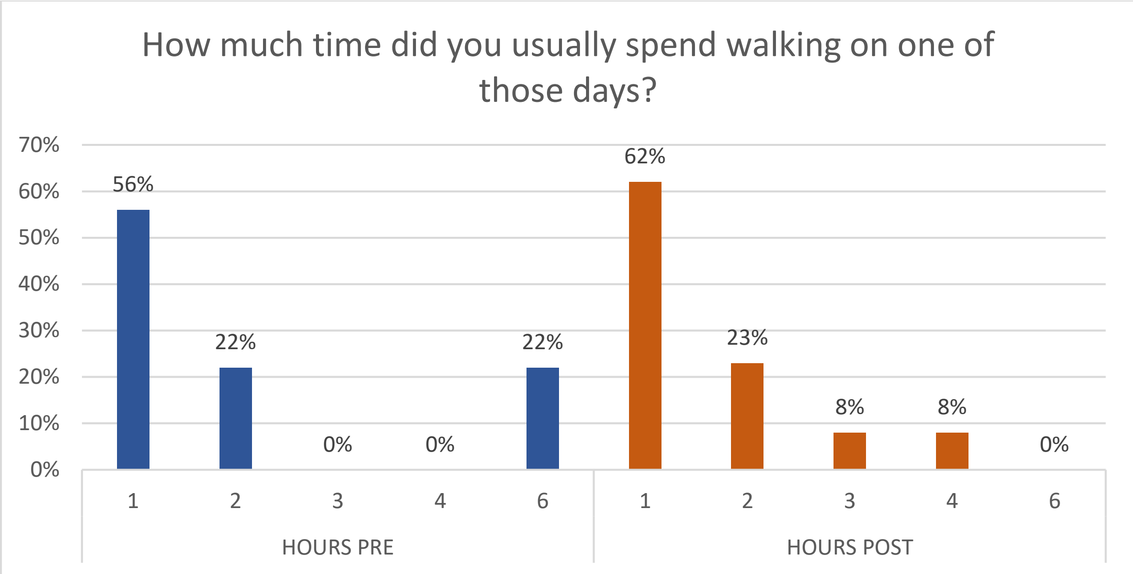


Figure 125: Health related walking physical activity 2. Production use case

- "During the last 7 days, how much time did you spend sitting on a weekday?"

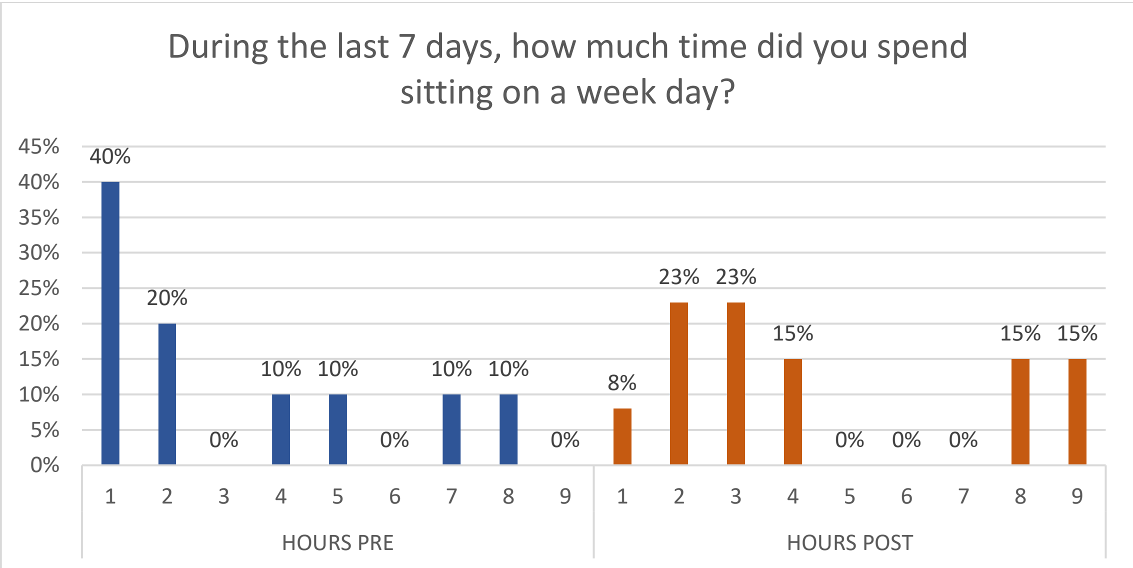


Figure 126: Health related sitting physical activity. Production use case

2.3.3.2.9 *Life-changing*

The last 2 additional questions were included in the post-questionnaire for the production workers too. The first one, asked the participants if they have had any significant situation that could have affected their lives. Only 3 of them answered this question in the affirmative (23%). These volunteers could have suffered changes in their life, but they may not be due to the WAOW tool usage. None of them wanted to explain what kind of situation was the one affecting their lives.

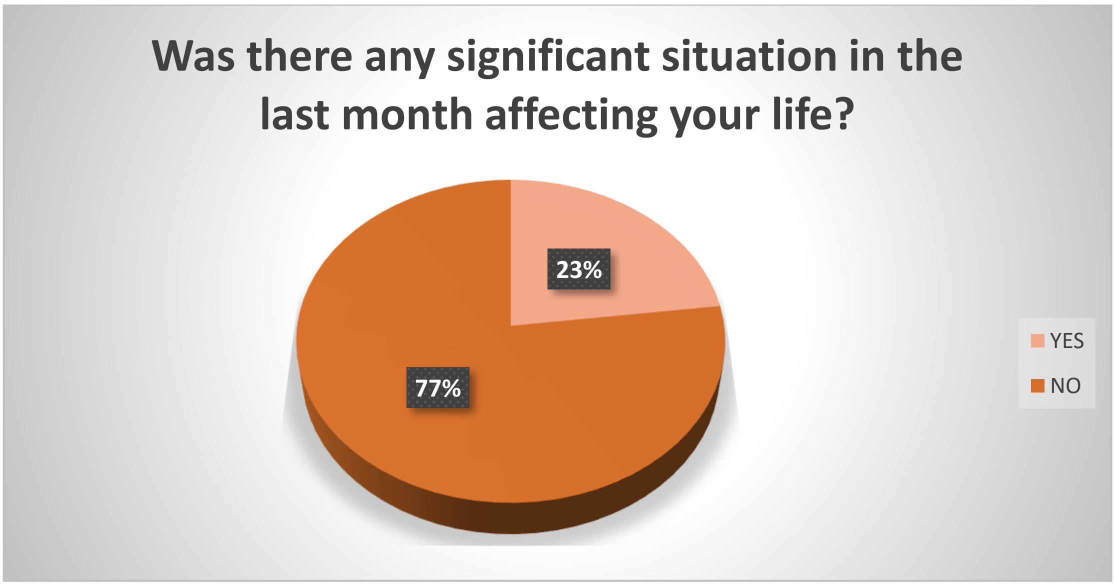


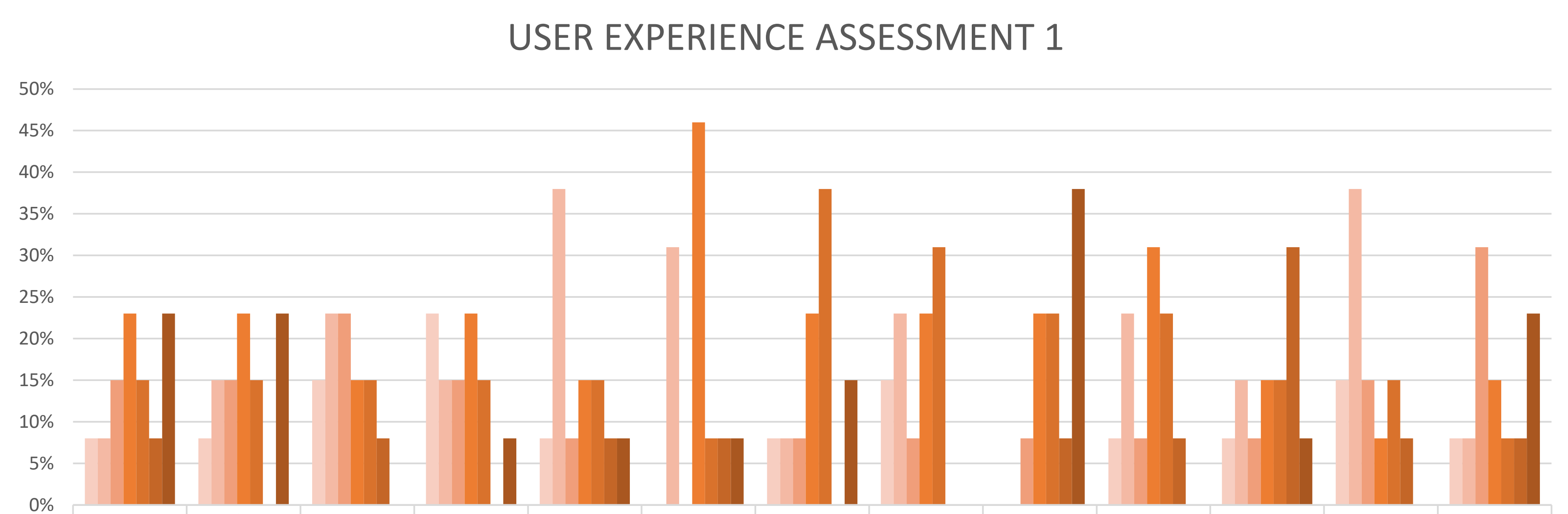
Figure 127: Life changing

2.3.3.2.10 User experience assessment

Next as in the office use case, the production volunteers were asked to evaluate their experience in understanding, learning, and using the WAOW tool. The details are described in the office use case user experience assessment.

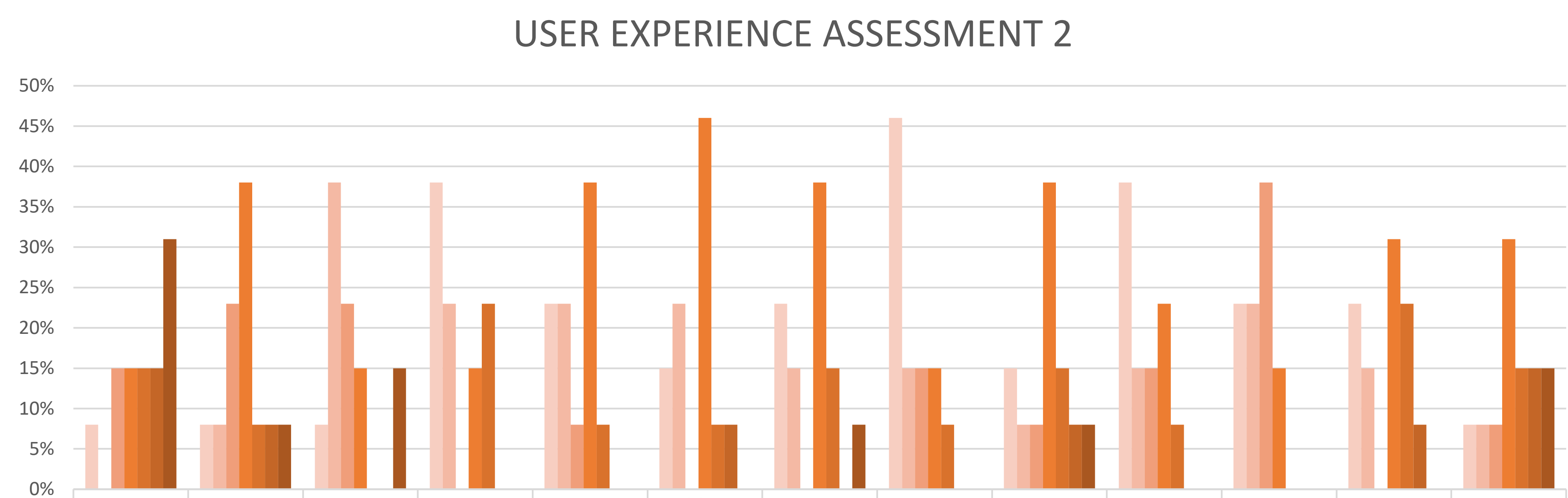
The numbers from 1 to 7 between the attributes represent gradations between the opposites. The volunteers can express their agreement with the attributes by selecting the number that most closely reflects their impression.

The graph is divided in two parts, to improve the viewing of the data.



	1 Annoying - 7 Enjoyable	1 Not understandable - 7 Understandable	1 Creative - 7 Dull	1 Easy to learn - 7 Difficult to learn	1 Valuable - 7 Inferior	1 Boring - 7 Exciting	1 Not interesting - 7 Interesting	1 Unpredictable - 7 Predictable	1 Fast - 7 Slow	1 Inventive - 7 Conventional	1 Obstructive - 7 Supportive	1 Good - 7 Bad	1 Complicated - 7 Easy
1	8%	8%	15%	23%	8%	0%	8%	15%	0%	8%	8%	15%	8%
2	8%	15%	23%	15%	38%	31%	8%	23%	0%	23%	15%	38%	8%
3	15%	15%	23%	15%	8%	0%	8%	8%	8%	8%	8%	15%	31%
4	23%	23%	15%	23%	15%	46%	23%	23%	23%	31%	15%	8%	15%
5	15%	15%	15%	15%	15%	8%	38%	31%	23%	23%	15%	15%	8%
6	8%	0%	8%	0%	8%	8%	0%	0%	8%	8%	31%	8%	8%
7	23%	23%	0%	8%	8%	8%	15%	0%	38%	0%	8%	0%	23%

Figure 128: User Experience Assessment 1. Production use case



	1 Unlikeable - 7 Pleasing	1 Usual - 7 Leading edge	1 Unpleasant - 7 Pleasant	1 Secure - 7 Not secure	1 Motivating - 7 Demotivating	1 Meets expectations - 7 Does not meet expectations	1 Inefficient - 7 Efficient	1 Clear - 7 Confusing	1 Impractical - 7 Practical	1 Organized - 7 Cluttered	1 Attractive - 7 Unattractive	1 Friendly - 7 Unfriendly	1 Conservative - 7 Innovative
1	8%	8%	8%	38%	23%	15%	23%	46%	15%	38%	23%	23%	8%
2	0%	8%	38%	23%	23%	23%	15%	15%	8%	15%	23%	15%	8%
3	15%	23%	23%	0%	8%	0%	0%	15%	8%	15%	38%	0%	8%
4	15%	38%	15%	15%	38%	46%	38%	15%	38%	23%	15%	31%	31%
5	15%	8%	0%	23%	8%	8%	15%	8%	15%	8%	0%	23%	15%
6	15%	8%	0%	0%	0%	8%	0%	0%	8%	0%	0%	8%	15%
7	31%	8%	15%	0%	0%	0%	8%	0%	8%	0%	0%	0%	15%

Figure 129: User Experience Assessment 2. Production use case

2.3.3.3 *Use Case: Teleworking (EXUS)*

2.3.3.3.1 Physical Working Conditions

To the question "Does your job ever require that you wear personal protective equipment?", all the volunteers answered that they don't need to wear this kind of equipment, as it is showed in the next figure.

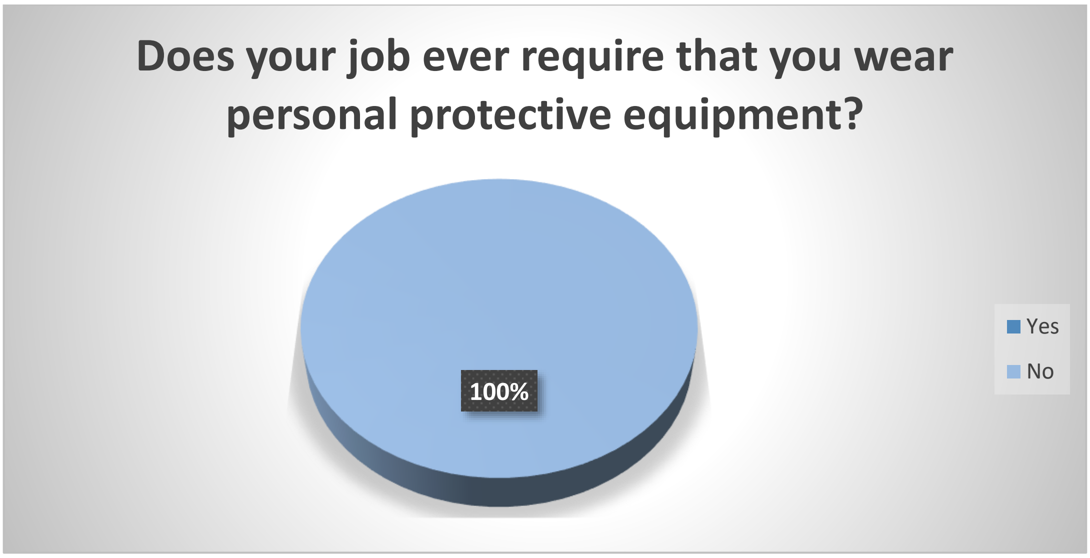


Figure 130: Protective equipment. Teleworking use case.

Next are the answers of the Teleworking use case workers comparing the pre and post.

When the teleworking participants were asked if they health or safety were at risk because of their work, they gave the following answers:

Figure 131: Risk because of work. Teleworking use case.

The volunteers were asked if their work affected their health. This is shown in the next figure.

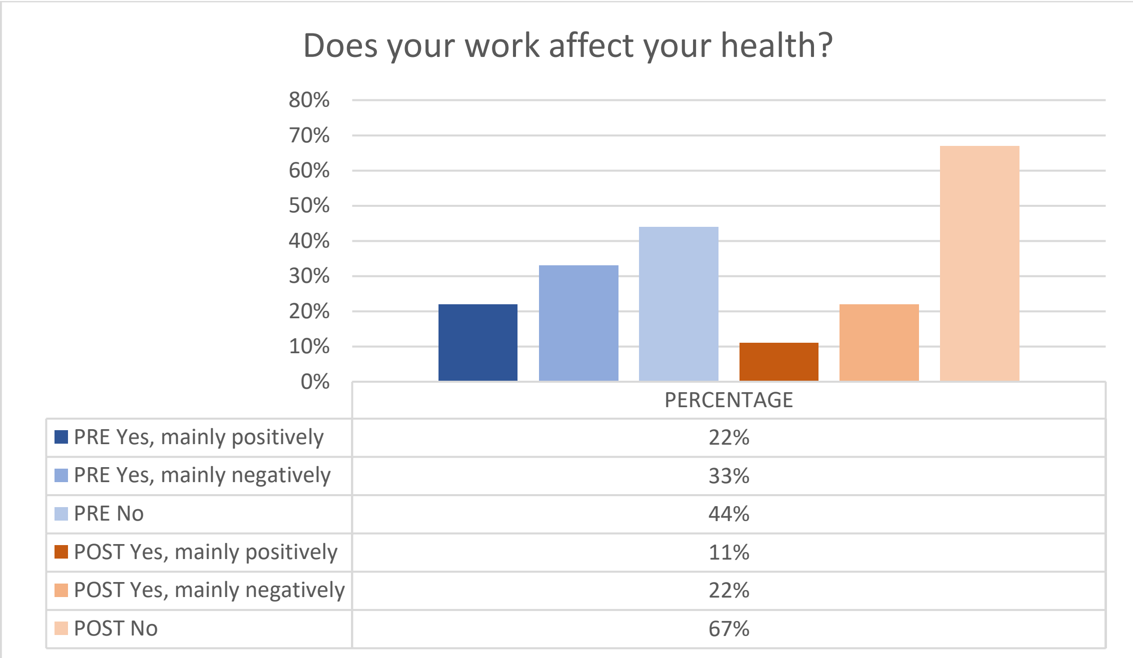
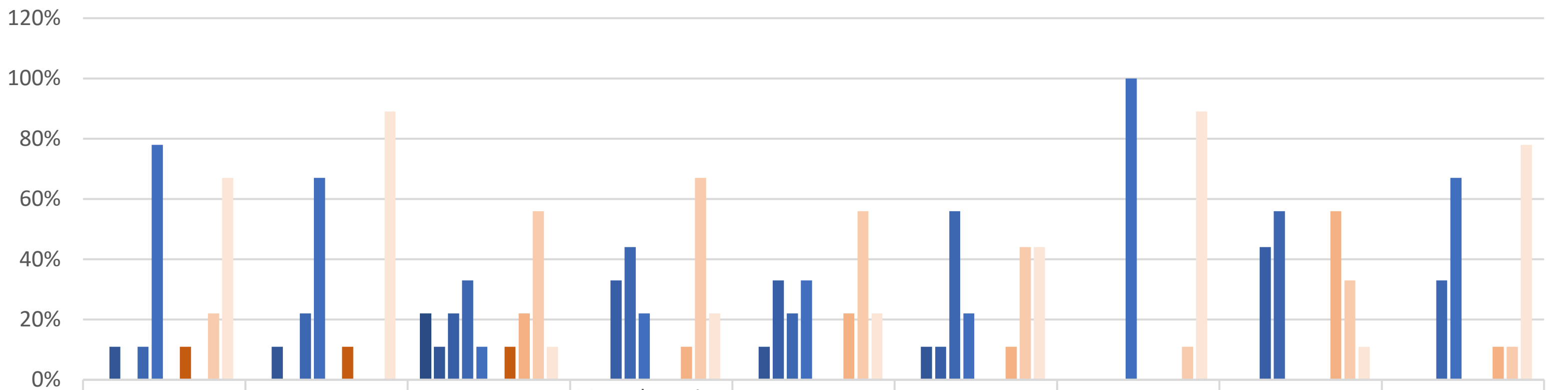


Figure 132: Health affected by the work. Teleworking use case.

The next graph compares different health problems that the participants suffered at the beginning and at the end of the pilots.

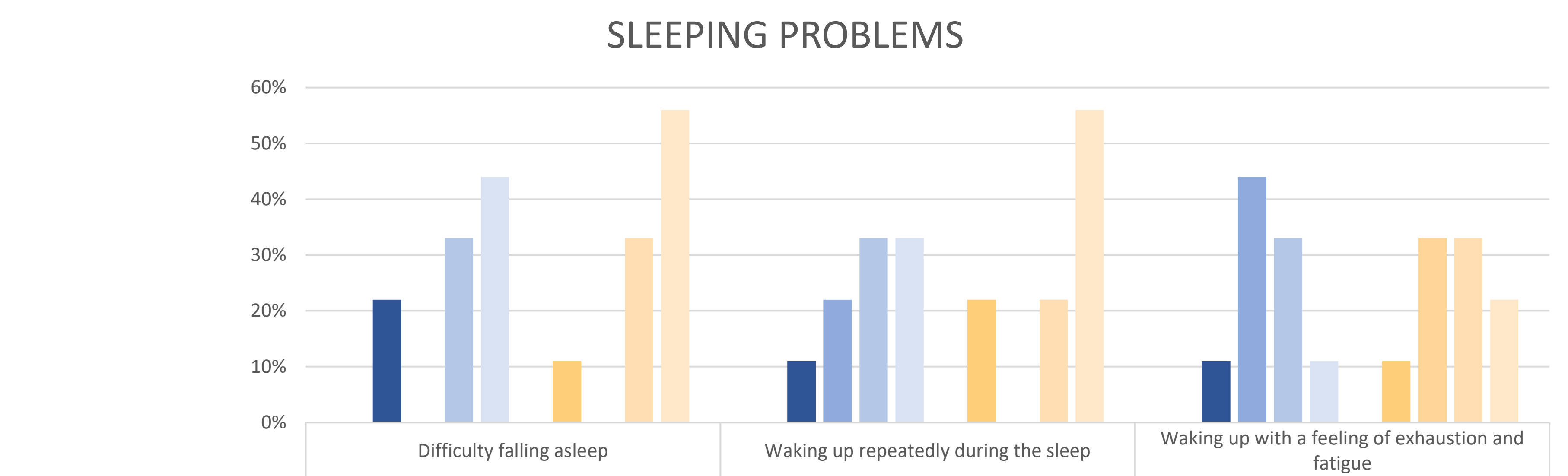
Over the last 6 months, did you have any of the following health problems?



	Hearing problems	Skin problems	Backache	Muscular pains in shoulders, neck and/or upper limbs	Muscular pains in lower limbs	Headaches, eyestrain	Injury	Overall fatigue	Other (Spontaneus)
PRE All the time	0%	0%	22%	0%	0%	0%	0%	0%	0%
PRE A large part of the time	11%	11%	11%	0%	11%	11%	0%	0%	0%
PRE Part of the time	0%	0%	22%	33%	33%	11%	0%	44%	0%
PRE A small part of the time	11%	22%	33%	44%	22%	56%	0%	56%	33%
PRE Not at all	78%	67%	11%	22%	33%	22%	100%	0%	67%
POST All the time	0%	0%	0%	0%	0%	0%	0%	0%	0%
POST A large part of the time	11%	11%	11%	0%	0%	0%	0%	0%	0%
POST Part of the time	0%	0%	22%	11%	22%	11%	0%	56%	11%
POST A small part of the time	22%	0%	56%	67%	56%	44%	11%	33%	11%
POST Not at all	67%	89%	11%	22%	22%	44%	89%	11%	78%

Figure 133: Health problems. Teleworking use case.

The next question is "Over the last 6 months, how often did you have any of the following sleep related problems?". The volunteers were asked about 3 different aspects about sleeping problems:



	Difficulty falling asleep	Waking up repeatedly during the sleep	Waking up with a feeling of exhaustion and fatigue
PRE Daily	0%	0%	0%
PRE Several times a week	22%	11%	11%
PRE Several times a month	0%	22%	44%
PRE Less often	33%	33%	33%
PRE Never	44%	33%	11%
POST Daily	0%	0%	0%
POST Several times a week	11%	22%	11%
POST Several times a month	0%	0%	33%
POST Less often	33%	22%	33%
POST Never	56%	56%	22%

Figure 134: Sleeping problems. Teleworking use case.

The next questions are about times when the participants might have been absent from work.

- Over the past 6 months how many days in total were you absent from work due to sick leave or health-related leave? (Figure 135).

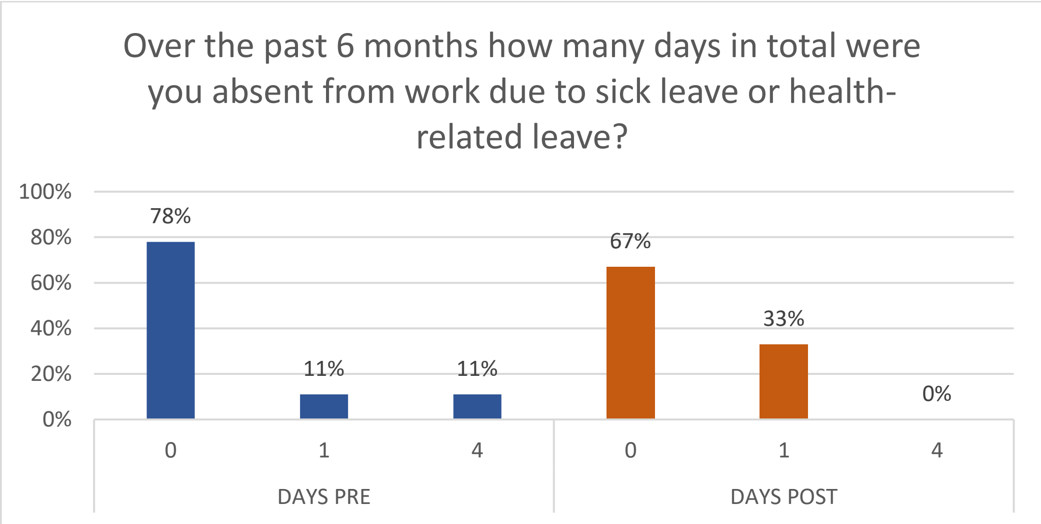


Figure 135: Absences from work. Teleworking use case.

- Asked if these days were caused by accidents at work, the participants, answered the following:

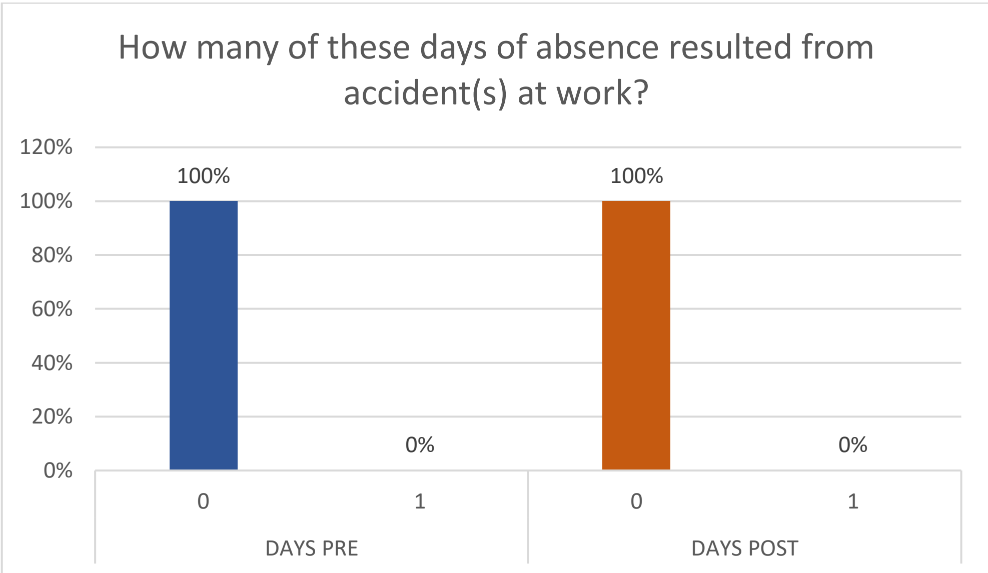


Figure 136: Accidents at work. Teleworking use case.

- To the question if these days of absence resulted from health problems caused or made worse by their work (excluding accidents), the participants answered this way:

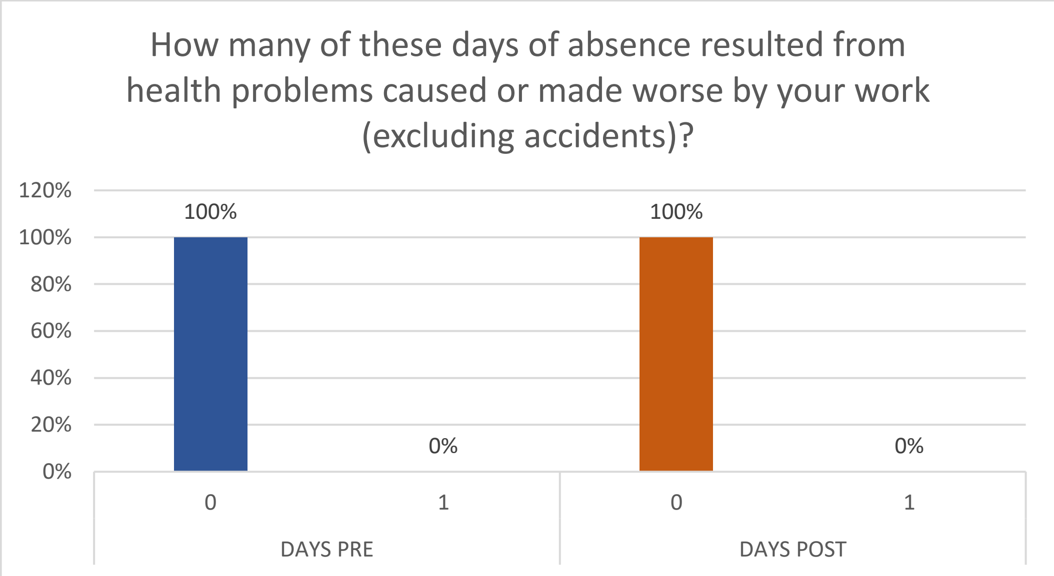


Figure 137: Health problems caused or made worse by work. Teleworking use case.

- The volunteers were asked if they worked when they were sick, and the number of days. To have a more graphic view of the data, you can see the graphs below.

Figure 138: Working when sick. Teleworking use case.

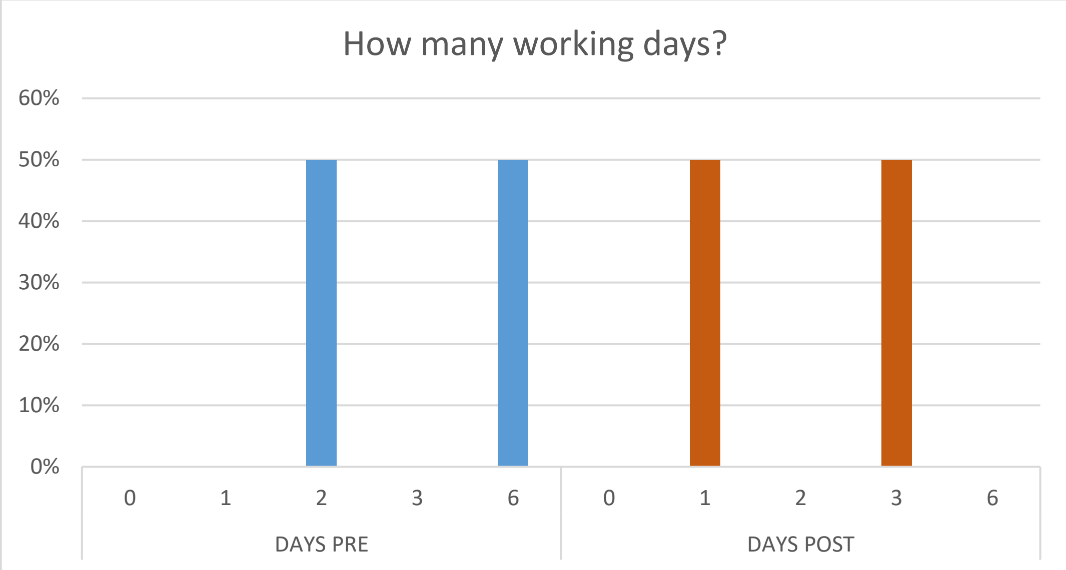
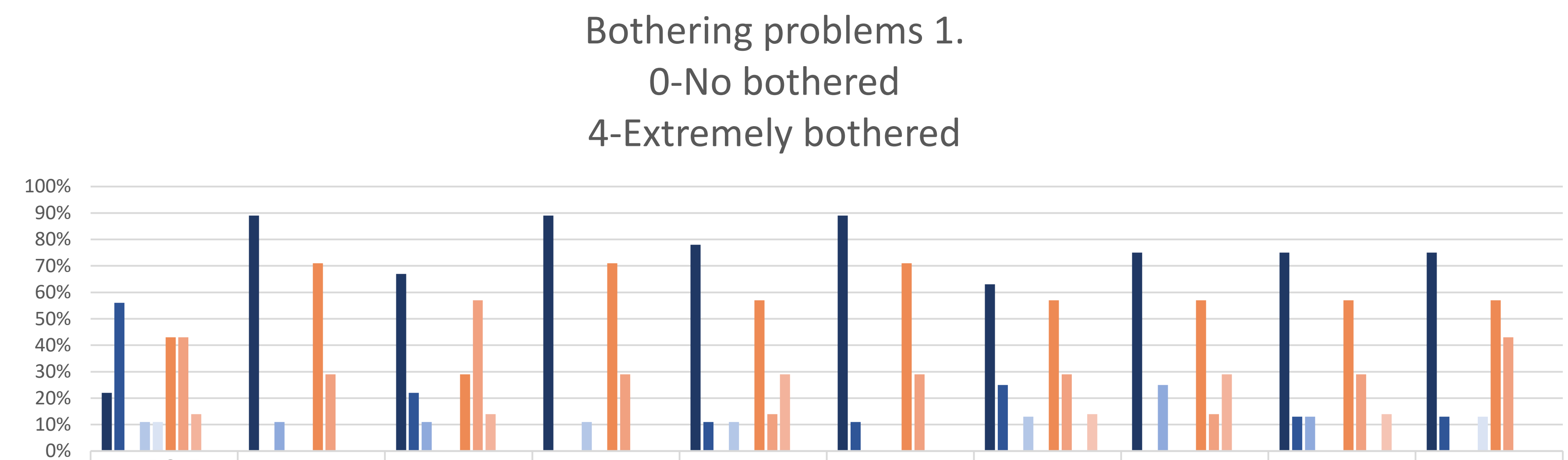


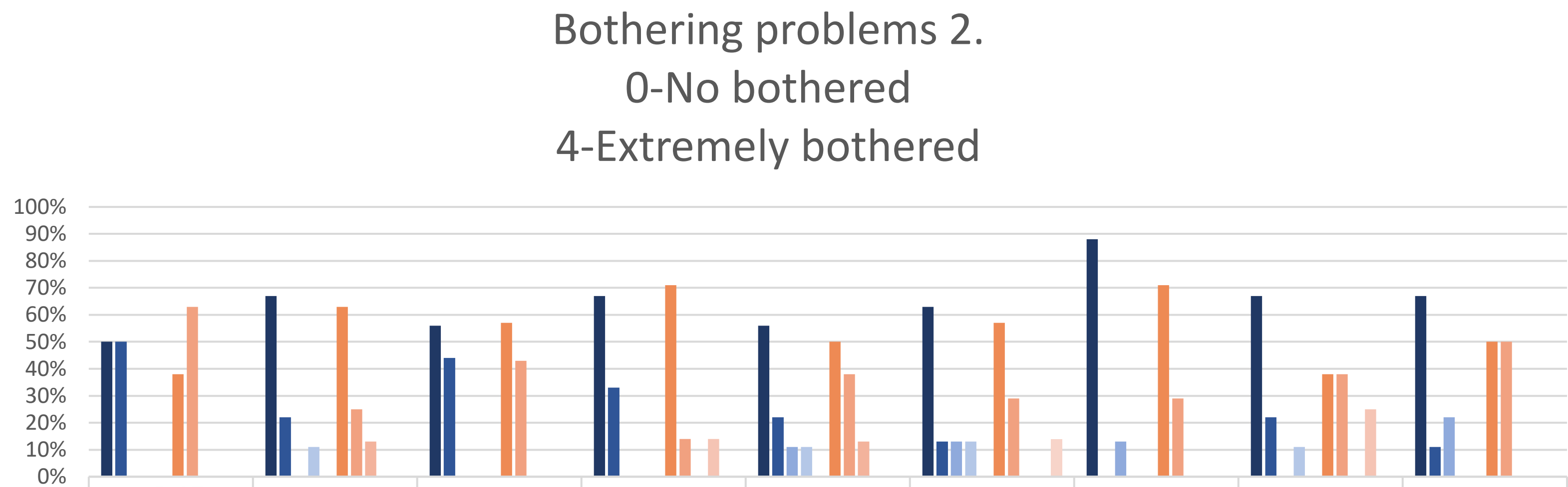
Figure 139: Working days when sick. Teleworking use case.

To end with the questions about physical working conditions, the volunteers were asked about the physical symptoms they had suffered during the last few weeks and how much discomfort these problems caused them. The answering options are divided from 0 to 4, where 0 means that the participant has not been bothered by the problem, and 4 means that the problem has been an extreme bother. The graph is divided in two, due to the large number of symptoms related.



	WEIGHT CHANGE (gain or loss of 5 libs. or more)	CONSTIPATION	DIZZINES	DIARRHEA	FAINTNESS	NAUSEA AND/OR VOMITING	ACID STOMACH OR INDIGESTION	STOMACH PAIN (E.G., CRAMPS)	HOT OR COLD SPELLS	HANDS TREMBLING
PRE 0	22%	89%	67%	89%	78%	89%	63%	75%	75%	75%
PRE 1	56%	0%	22%	0%	11%	11%	25%	0%	13%	13%
PRE 2	0%	11%	11%	0%	0%	0%	0%	25%	13%	0%
PRE 3	11%	0%	0%	11%	11%	0%	13%	0%	0%	0%
PRE 4	11%	0%	0%	0%	0%	0%	0%	0%	0%	13%
POST 0	43%	71%	29%	71%	57%	71%	57%	57%	57%	57%
POST 1	43%	29%	57%	29%	14%	29%	29%	14%	29%	43%
POST 2	14%	0%	14%	0%	29%	0%	0%	29%	0%	0%
POST 3	0%	0%	0%	0%	0%	0%	14%	0%	14%	0%
POST 4	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Figure 140: Bothering problems in the last 2 weeks 1. Teleworking use case.



	HEART PALPITATIONS (Heart pounding/racing)	SHORTNESS OF BREATH when not exercising or working hard	NUMBNESS OR TINGLING in parts of your body	PAINS IN HEART OR CHEST	STUFFY HEAD OR NOSE	BLURRED VISION	NOSEBLEED	JOINT PAINS	COLD OR COUGH
PRE 0	50%	67%	56%	67%	56%	63%	88%	67%	67%
PRE 1	50%	22%	44%	33%	22%	13%	0%	22%	11%
PRE 2	0%	0%	0%	0%	11%	13%	13%	0%	22%
PRE 3	0%	11%	0%	0%	11%	13%	0%	11%	0%
PRE 4	0%	0%	0%	0%	0%	0%	0%	0%	0%
POST 0	38%	63%	57%	71%	50%	57%	71%	38%	50%
POST 1	63%	25%	43%	14%	38%	29%	29%	38%	50%
POST 2	0%	13%	0%	0%	13%	0%	0%	0%	0%
POST 3	0%	0%	0%	14%	0%	0%	0%	25%	0%
POST 4	0%	0%	0%	0%	0%	14%	0%	0%	0%

Figure 141: Bothering problems in the last 2 weeks 1. Teleworking use case.

2.3.3.3.2 *Psychosocial working conditions*

To evaluate the psychological working conditions, the volunteers were asked the following questions in the pre and post questionnaires:

- "Is your work meaningful?"

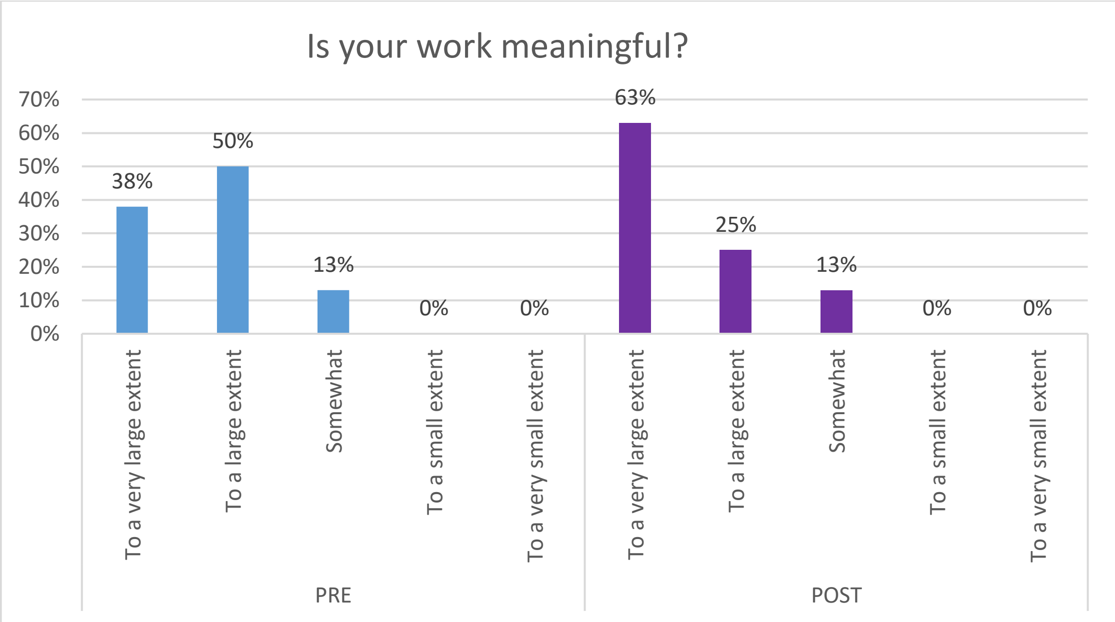


Figure 142: Meaning of work. Teleworking use case.

- "How pleased are you with your job as a whole?"

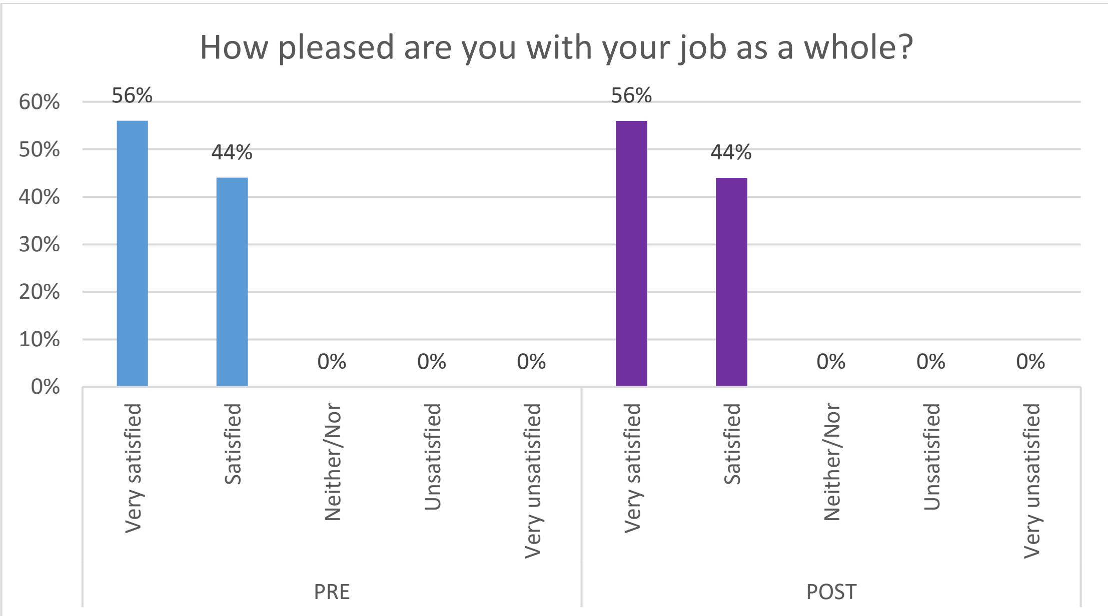


Figure 143: Job satisfaction. Teleworking use case.

- "How often have you felt worn out?"

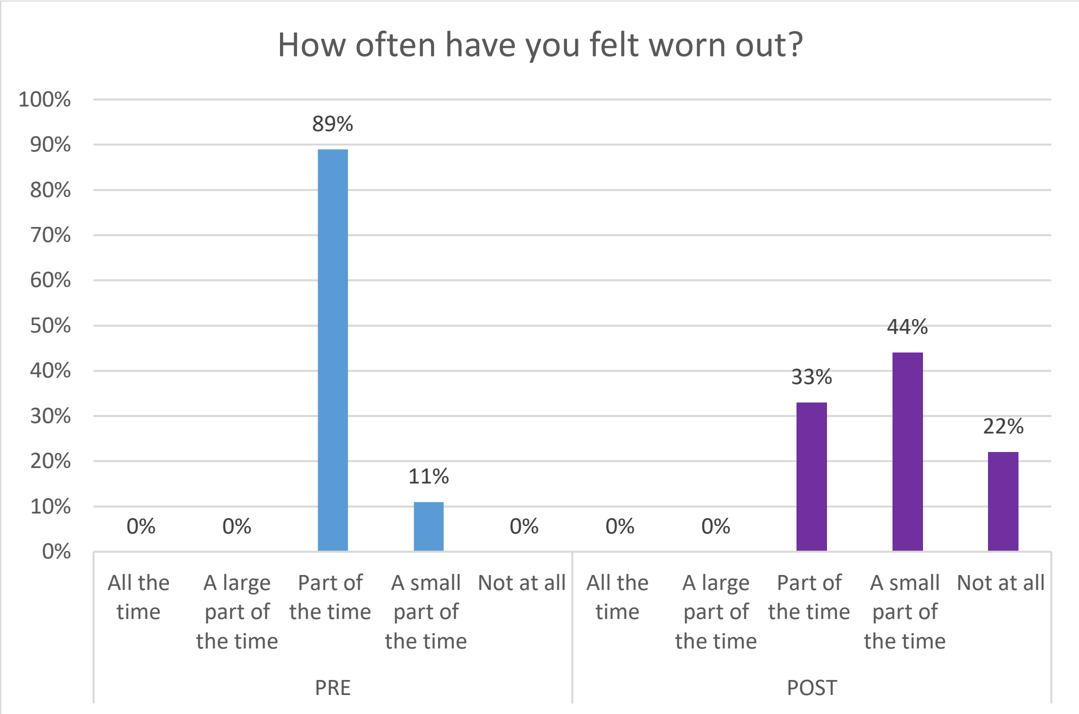


Figure 144: Burn out and stress 1. Teleworking use case.

- "How often have you been physically exhausted?"

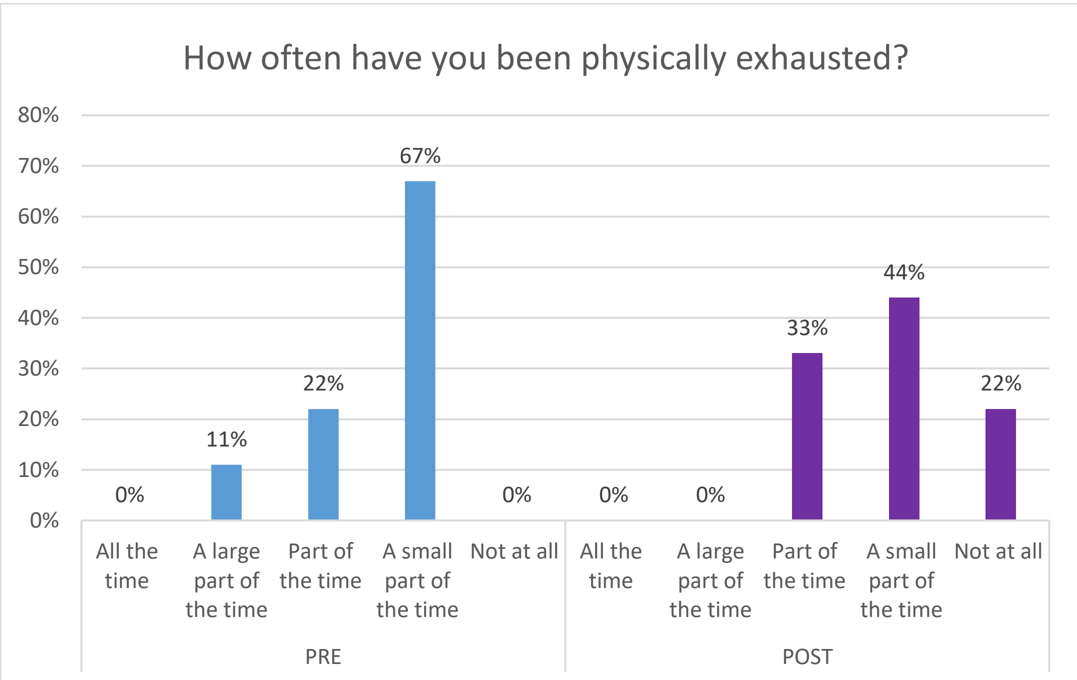


Figure 145: Burn out and stress 2. Teleworking use case.

- "How often have you been emotionally exhausted?"

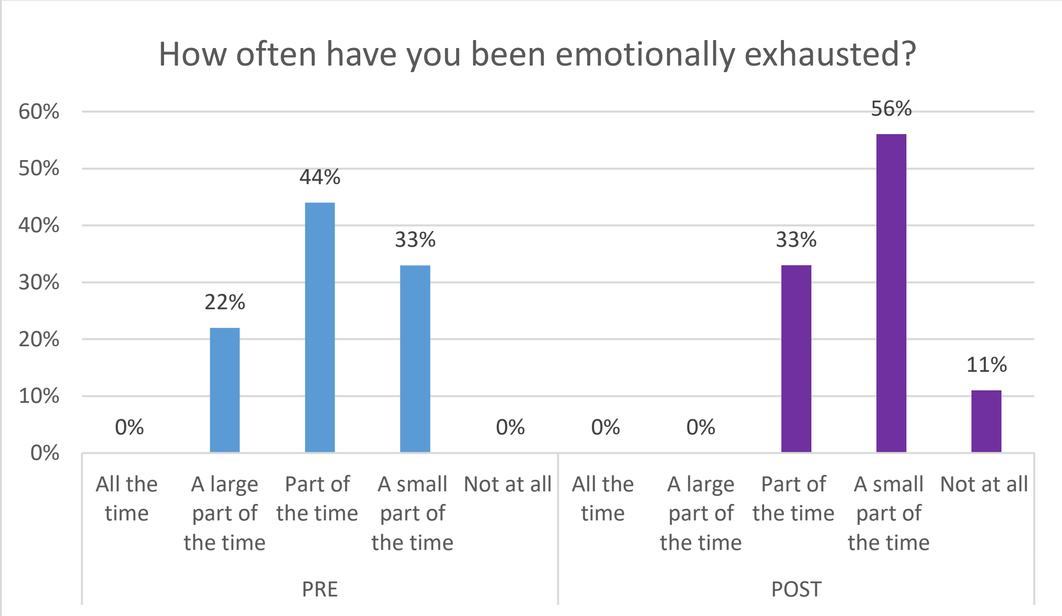


Figure 146: Burn out and stress 3. Teleworking use case.

- "How often have you been tense?"

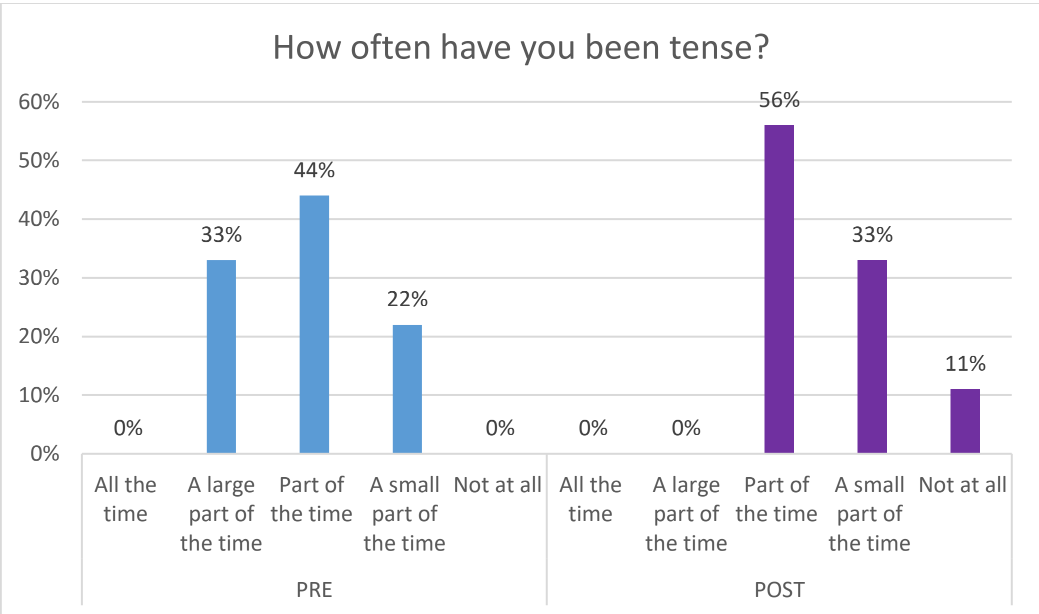


Figure 147: Burn out and stress 4. Teleworking use case.

- “Is your work recognized and appreciated by the management?”

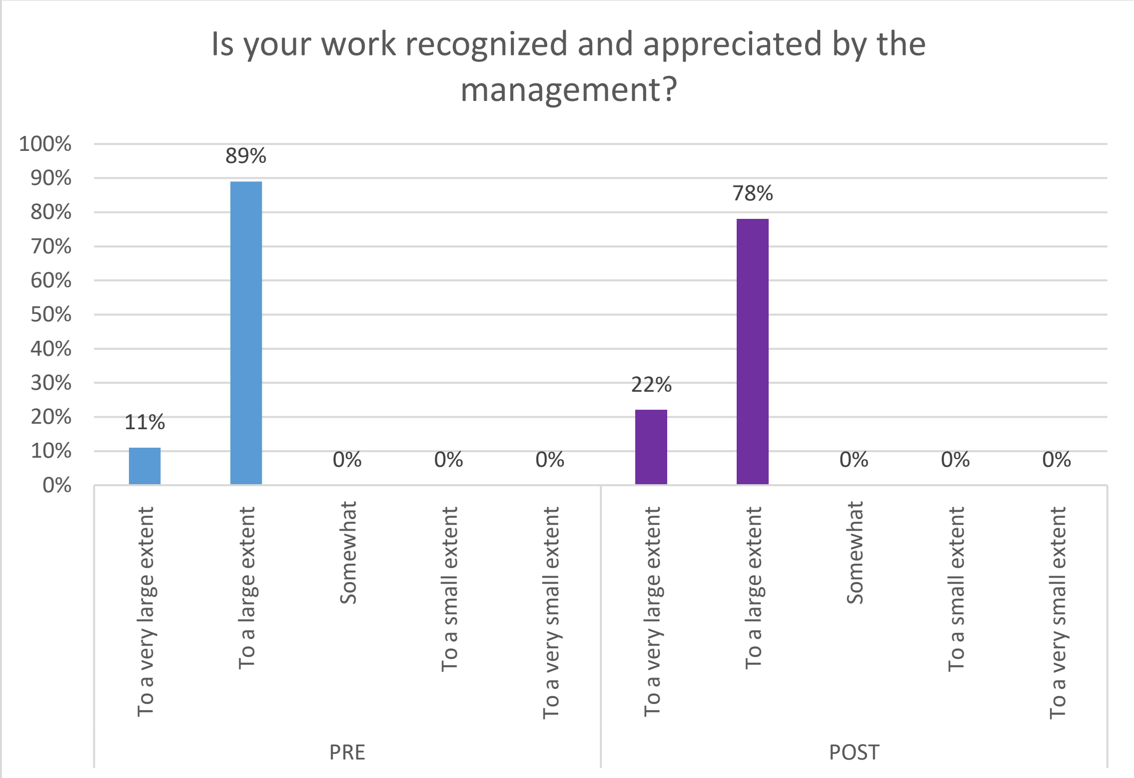
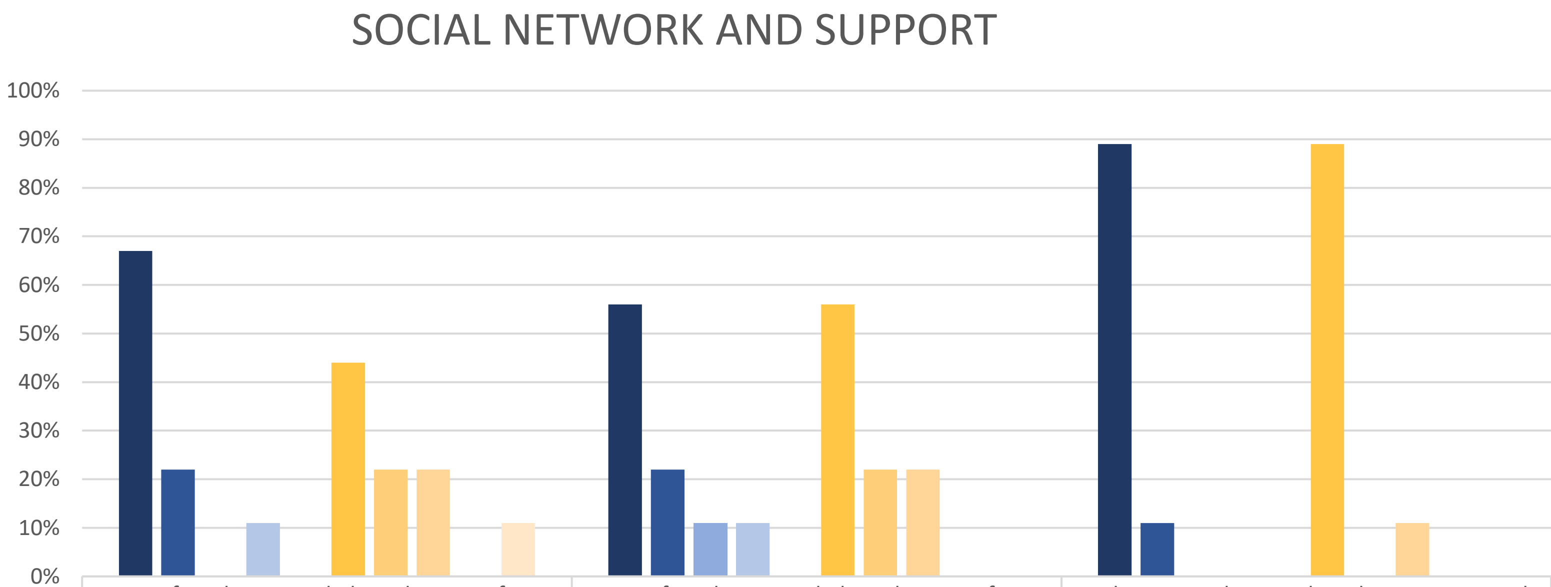


Figure 148: Social network and Support 1. Teleworking use case

The next figure shows the social network and support felt by the volunteers in the pre and post questionnaire. To evaluate this aspect, three questions were asked to the participants.

- “How often do you get help and support from your immediate superior, if needed?”
- “How often do you get help and support from your colleagues, if needed?”
- “Is there a good atmosphere between you and your colleagues?”

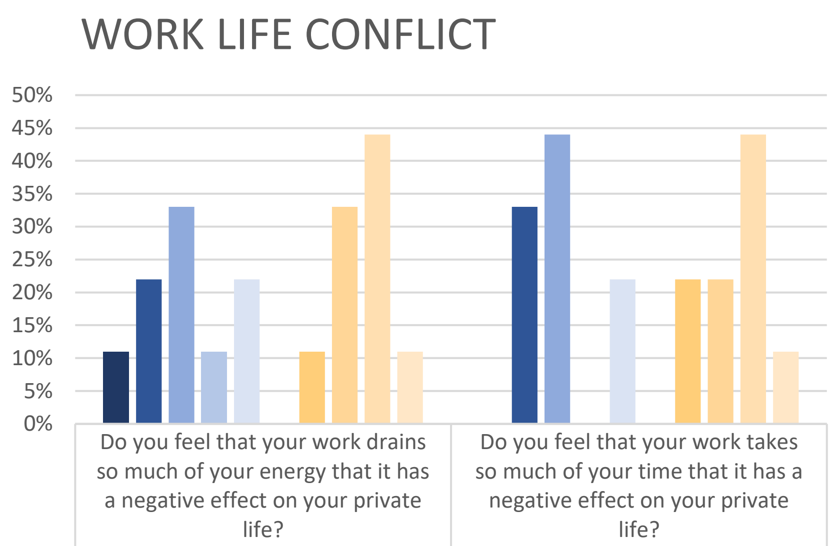


	How often do you get help and support from your immediate superior, if needed?	How often do you get help and support from your colleagues, if needed?	Is there a good atmosphere between you and your colleagues?
PRE Always	67%	56%	89%
PRE Often	22%	22%	11%
PRE Sometimes	0%	11%	0%
PRE Seldom	11%	11%	0%
PRE Never/hardly ever	0%	0%	0%
POST Always	44%	56%	89%
POST Often	22%	22%	0%
POST Sometimes	22%	22%	11%
POST Seldom	0%	0%	0%
POST Never/hardly ever	11%	0%	0%

Figure 149: Social network and support 2. Teleworking use case.

We continue with two questions related to work life conflict. They can be seen in detail in the next graphic.

- "Do you feel that your work drains so much of your energy that it has a negative effect on your private life?"
- "Do you feel that your work takes so much of your time that it has a negative effect on your private life?"



	Do you feel that your work drains so much of your energy that it has a negative effect on your private life?	Do you feel that your work takes so much of your time that it has a negative effect on your private life?
PRE To a very large extent	11%	0%
PRE To a large extent	22%	33%
PRE Somewhat	33%	44%
PRE To a small extent	11%	0%
PRE To a very small extent	22%	22%
POST To a very large extent	0%	0%
POST To a large extent	11%	22%
POST Somewhat	33%	22%
POST To a small extent	44%	44%
POST To a very small extent	11%	11%

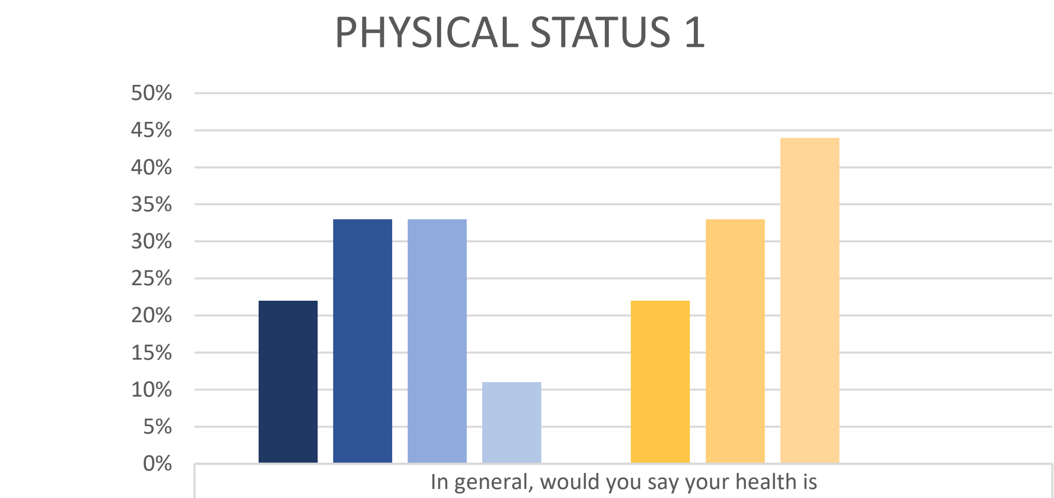
Figure 150: Work life conflict. Teleworking use case

2.3.3.3.3 *Physical status*

To evaluate the physical status of the participants at the beginning and at the end of the pilots, they were asked seven questions related with this subject.

Five graphs will show the detailed data of their answers.

- "In general, would you say your health is…"

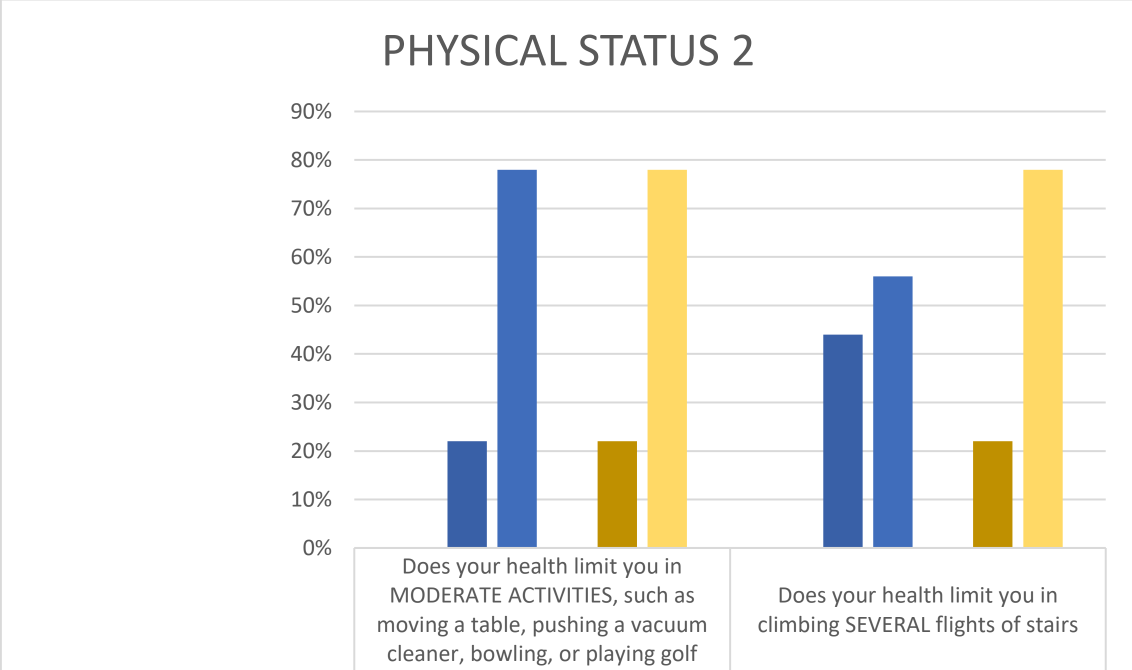


	In general, would you say your health is
PRE Excellent	22%
PRE Very good	33%
PRE Good	33%
PRE Fair	11%
PRE Poor	0%
POST Excellent	22%
POST Very good	33%
POST Good	44%
POST Fair	0%
POST Poor	0%

Figure 151: Physical status1. Teleworking use case.

The next figure shows two questions related with the physical status of the participants at the beginning and at the end of the pilots.

- "Does your health limit you in MODERATE ACTIVITIES, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf?"
- "Does your health limit you in climbing SEVERAL flights of stairs?"



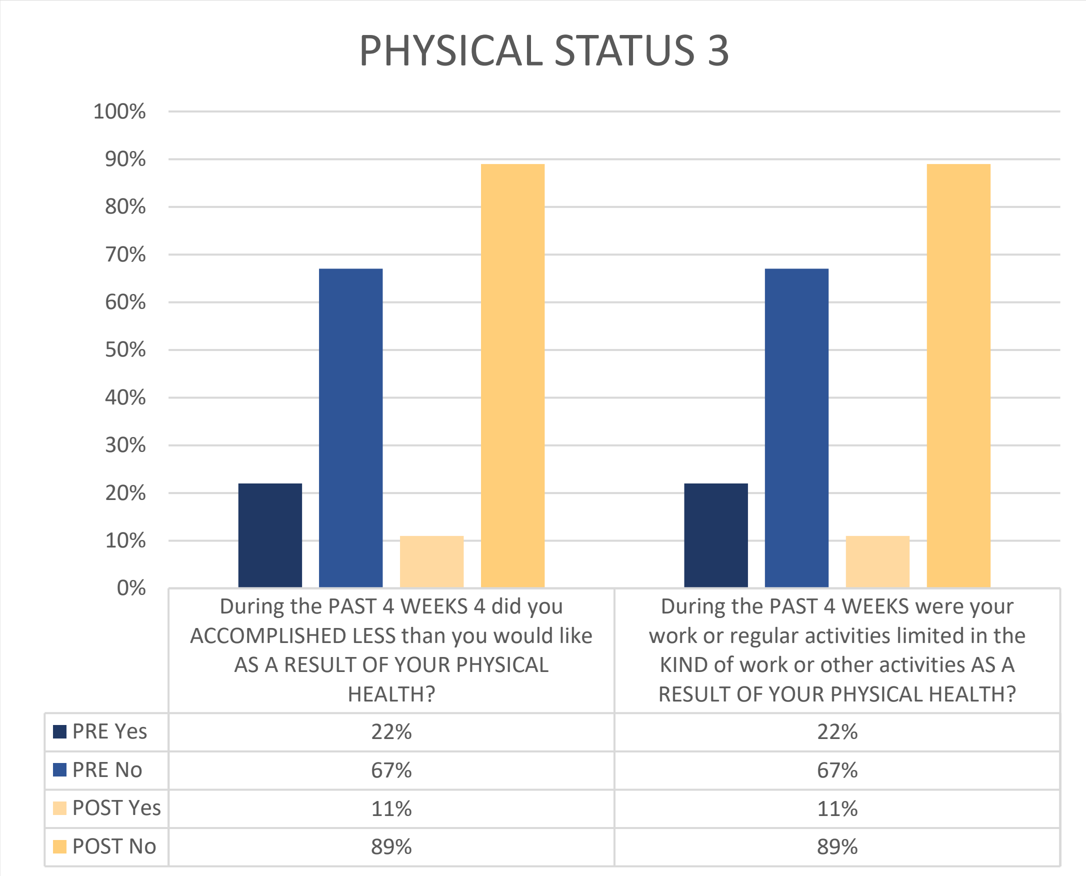
	Does your health limit you in MODERATE ACTIVITIES, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf	Does your health limit you in climbing SEVERAL flights of stairs
PRE Yes, Limited A Lot	0%	0%
PRE Yes, Limited A Little	22%	44%
PRE No, Not Limited At All	78%	56%
POST Yes, Limited A Lot	0%	0%
POST Yes, Limited A Little	22%	22%
POST No, Not Limited At All	78%	78%

Figure 152: Physical status 2. Teleworking use case.

Again, the next figure shows two questions related with the physical status of the participants in a period of four weeks in advance.

- “During the PAST 4 WEEKS 4 did you ACCOMPLISHED LESS than you would like AS A RESULT OF YOUR PHYSICAL HEALTH?”

- “During the PAST 4 WEEKS were your work or regular activities limited in the KIND of work or other activities AS A RESULT OF YOUR PHYSICAL HEALTH?”

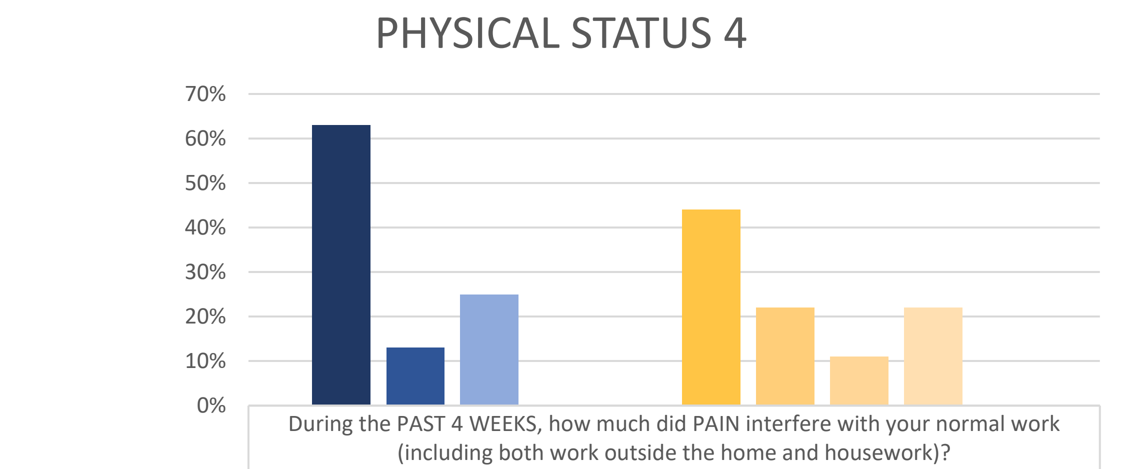


	During the PAST 4 WEEKS 4 did you ACCOMPLISHED LESS than you would like AS A RESULT OF YOUR PHYSICAL HEALTH?	During the PAST 4 WEEKS were your work or regular activities limited in the KIND of work or other activities AS A RESULT OF YOUR PHYSICAL HEALTH?
PRE Yes	22%	22%
PRE No	67%	67%
POST Yes	11%	11%
POST No	89%	89%

Figure 153: Physical status 3. Teleworking use case.

The next physical status question is reflected in the graphic located below.

- "During the PAST 4 WEEKS, how much did PAIN interfere with your normal work (including both work outside the home and housework)?"

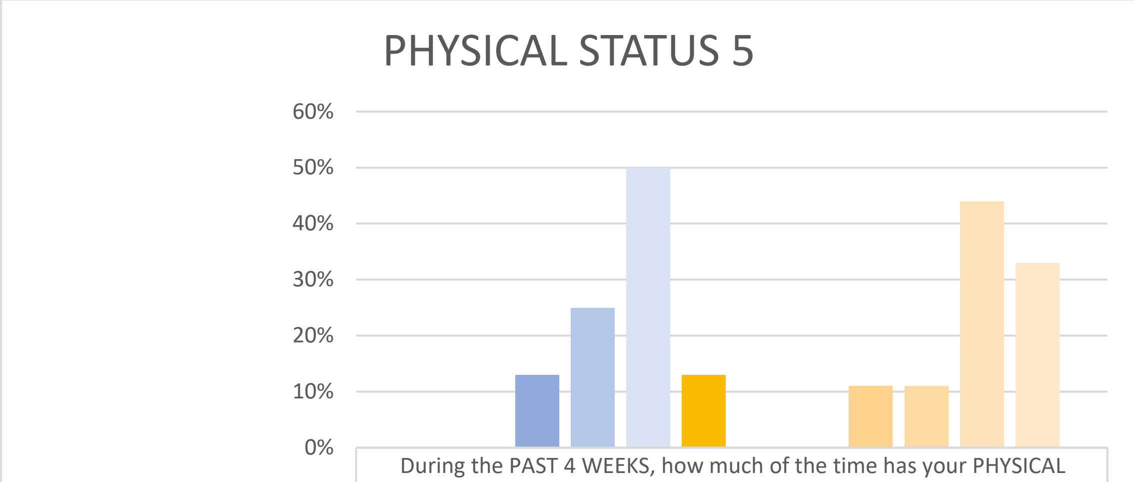


	During the PAST 4 WEEKS, how much did PAIN interfere with your normal work (including both work outside the home and housework)?
PRE Not at all	63%
PRE A little bit	13%
PRE Moderately	25%
PRE Quite a bit	0%
PRE Extremely	0%
POST Not at all	44%
POST A little bit	22%
POST Moderately	11%
POST Quite a bit	22%
POST Extremely	0%

Figure 154: Physical status 4. Teleworking use case.

The last graphic about physical status of the participants is shown below.

- "How much of the time has your PHYSICAL HEALTH OR EMOTIONAL PROBLEMS interfered with your social activities (like visiting with friends, relatives, etc.)?"



	During the PAST 4 WEEKS, how much of the time has your PHYSICAL HEALTH OR EMOTIONAL PROBLEMS interfered with your social activities (like visiting with friends, relatives, etc.)?
PRE All the time	0%
PRE Most of the time	0%
PRE A good bit of the time	13%
PRE Some of the time	25%
PRE A little of the time	50%
POST None of the time	13%
POST All the time	0%
POST Most of the time	0%
POST A good bit of the time	11%
POST Some of the time	11%
POST A little of the time	44%
POST None of the time	33%

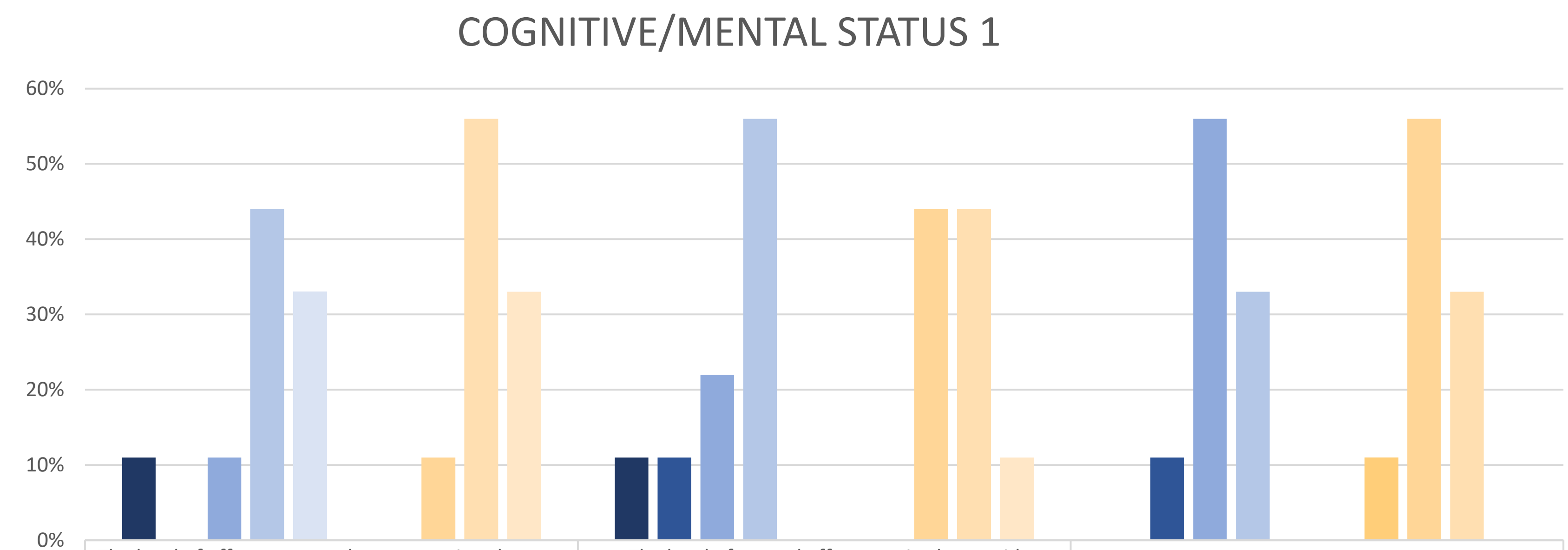
Figure 155: Physical status 5. Teleworking use case.

2.3.3.3.4 *Cognitive/Mental status*

For the assessment of the cognitive/mental status of the Teleworking use case workers, they were asked six questions that address this topic.

The next graphic shows the detailed answers of the participants to the three first questions.

- "The level of effort or mental concentration that my job requires is…"
- "The level of mental effort required to avoid mistakes in my work is…"
- "The fatigue that my work produces is…"

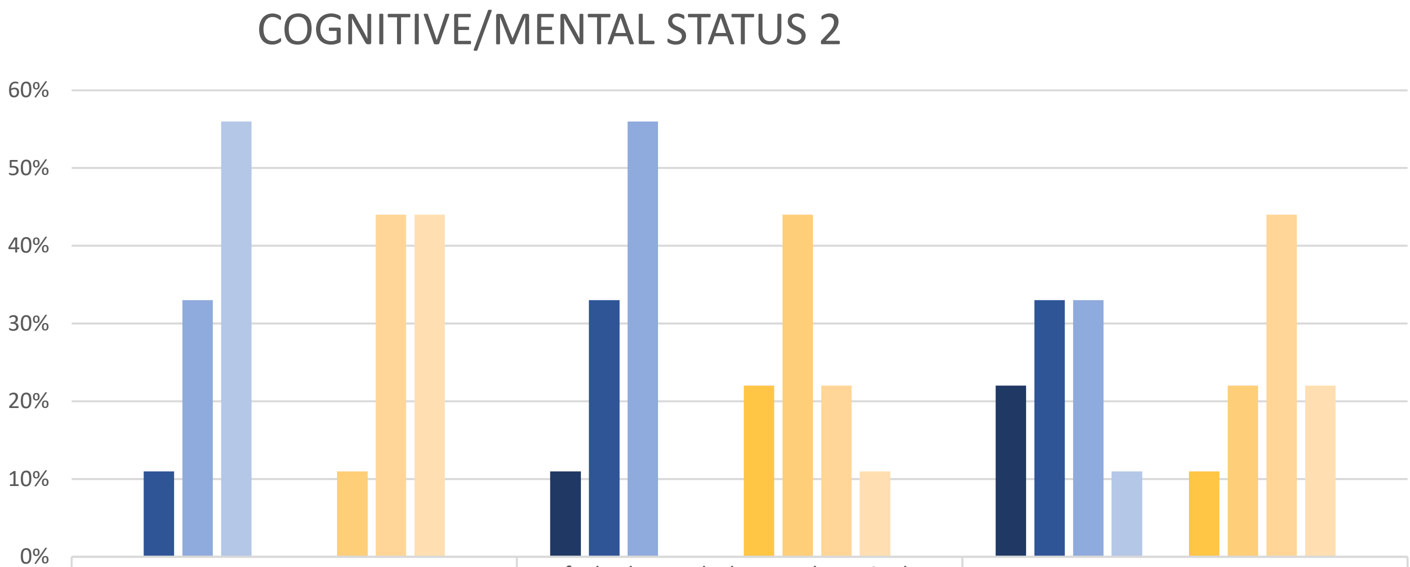


	The level of effort or mental concentration that my job requires is	The level of mental effort required to avoid mistakes in my work is	The fatigue that my work produces is
PRE Very low	11%	11%	0%
PRE Low	0%	11%	11%
PRE Medium	11%	22%	56%
PRE High	44%	56%	33%
PRE Very high	33%	0%	0%
POST Very low	0%	0%	0%
POST Low	0%	0%	11%
POST Medium	11%	44%	56%
POST High	56%	44%	33%
POST Very high	33%	11%	0%

Figure 156: Cognitive/Mental status 1. Teleworking use case.

The next three questions about cognitive/mental status of the volunteers are reflected in the next graph.

- “At the end of the workday I feel exhausted”
- “I feel exhausted when I wake up in the morning and have to face another day of work”
- “I have a hard time relaxing after work”



	At the end of the work day I feel exhausted	I feel exhausted when I wake up in the morning and have to face another day of work	I have a hard time relaxing after work
PRE Strongly disagree	0%	11%	22%
PRE Somewhat disagree	11%	33%	33%
PRE Neither disagree, nor agree	33%	56%	33%
PRE Somewhat agree	56%	0%	11%
PRE Strongly agree	0%	0%	0%
POST Strongly disagree	0%	22%	11%
POST Somewhat disagree	11%	44%	22%
POST Neither disagree, nor agree	44%	22%	44%
POST Somewhat agree	44%	11%	22%
POST Strongly agree	0%	0%	0%

Figure 157: Cognitive/Mental status 2. Teleworking use case

2.3.3.3.5 Social relationships

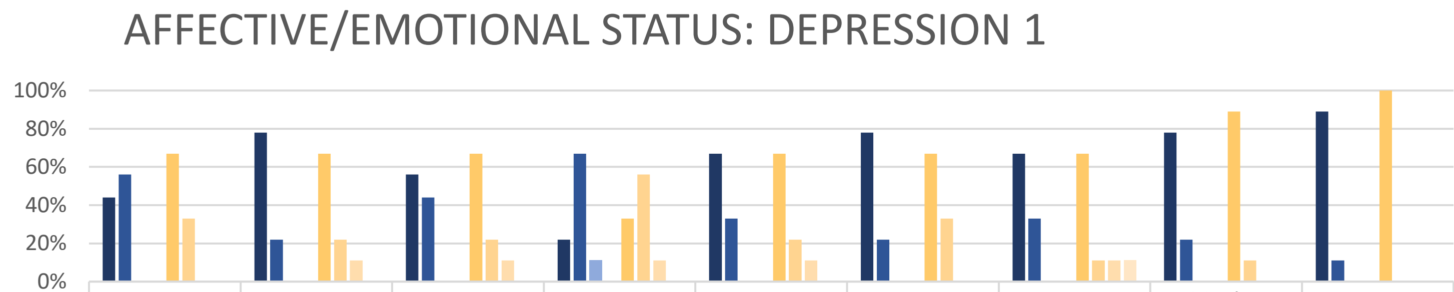
To evaluate the affective and emotional status of the Teleworking volunteers, they were asked first with ten questions, that collect their general state with respect to a possible depression. They were asked if over the last 2 weeks, how often have you been bothered by any of the following problems:

- “Little interest or pleasure in doing things”
- “Feeling down, depressed, or hopeless”
- “Trouble falling or staying asleep, or sleeping too much”
- “Feeling tired or having little energy”
- “Poor appetite or overeating”
- “Feeling bad about yourself — or that you are a failure or have let yourself or your family down”
- “Trouble concentrating on things, such as reading the newspaper or watching television”
- “Moving or speaking so slowly that other people could have noticed. Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual”
- “Thoughts that you would be better off dead or of hurting yourself in some way”

The volunteers’ answers to these questions can be seen in the next graph: Depression 1.

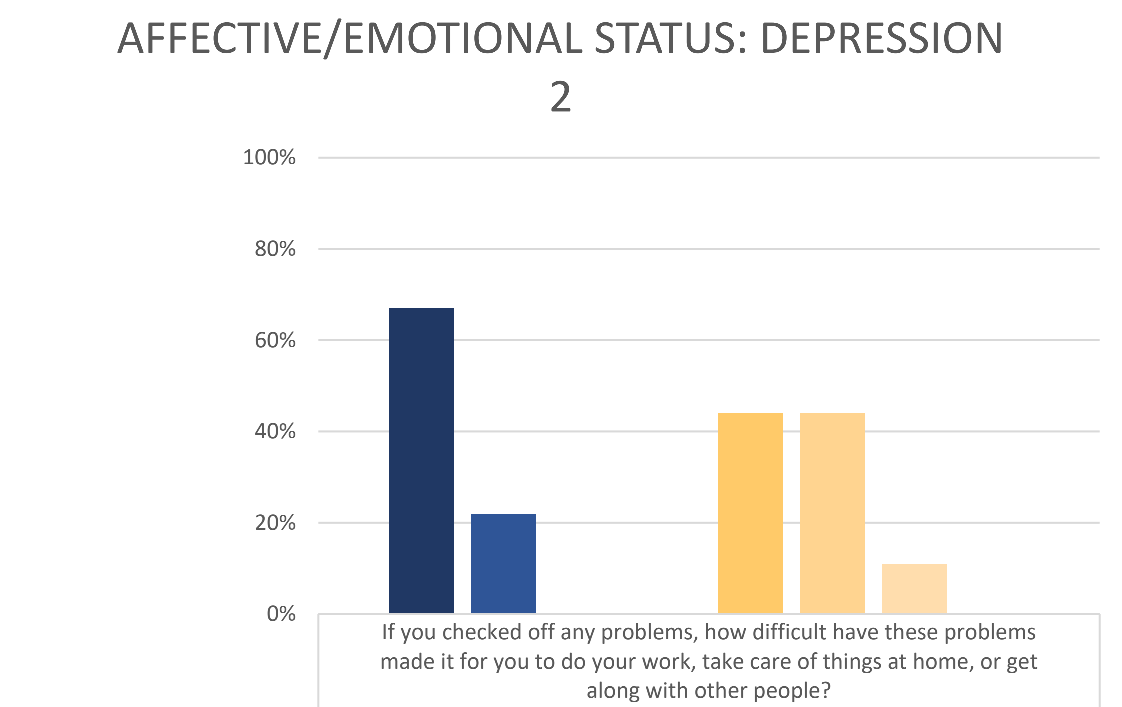
There is one final question reflected in the Affective/Emotional status: Depression 2.

- “If you checked off any problems, how difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?”



	Little interest or pleasure in doing things	Feeling down, depressed, or hopeless	Trouble falling or staying asleep, or sleeping too much	Feeling tired or having little energy	Poor appetite or overeating	Feeling bad about yourself — or that you are a failure or have let yourself or your family down	Trouble concentrating on things, such as reading the newspaper or watching television	Moving or speaking so slowly that other people could have noticed. Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual	Thoughts that you would be better off dead or of hurting yourself in some way
PRE Not at all	44%	78%	56%	22%	67%	78%	67%	78%	89%
PRE Several days	56%	22%	44%	67%	33%	22%	33%	22%	11%
PRE More than half of the days	0%	0%	0%	11%	0%	0%	0%	0%	0%
PRE Nearly every day	0%	0%	0%	0%	0%	0%	0%	0%	0%
POST Not at all	67%	67%	67%	33%	67%	67%	67%	89%	100%
POST Several days	33%	22%	22%	56%	22%	33%	11%	11%	0%
POST More than half of the days	0%	11%	11%	11%	11%	0%	11%	0%	0%
POST Nearly every day	0%	0%	0%	0%	0%	0%	11%	0%	0%

Figure 158: Affective/Emotional status: Depression. Teleworking use case.



	If you checked off any problems, how difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?
PRE Not difficult at all	67%
PRE Somewhat difficult	22%
PRE Very difficult	0%
PRE Extremely difficult	0%
POST Not difficult at all	44%
POST Somewhat difficult	44%
POST Very difficult	11%
POST Extremely difficult	0%

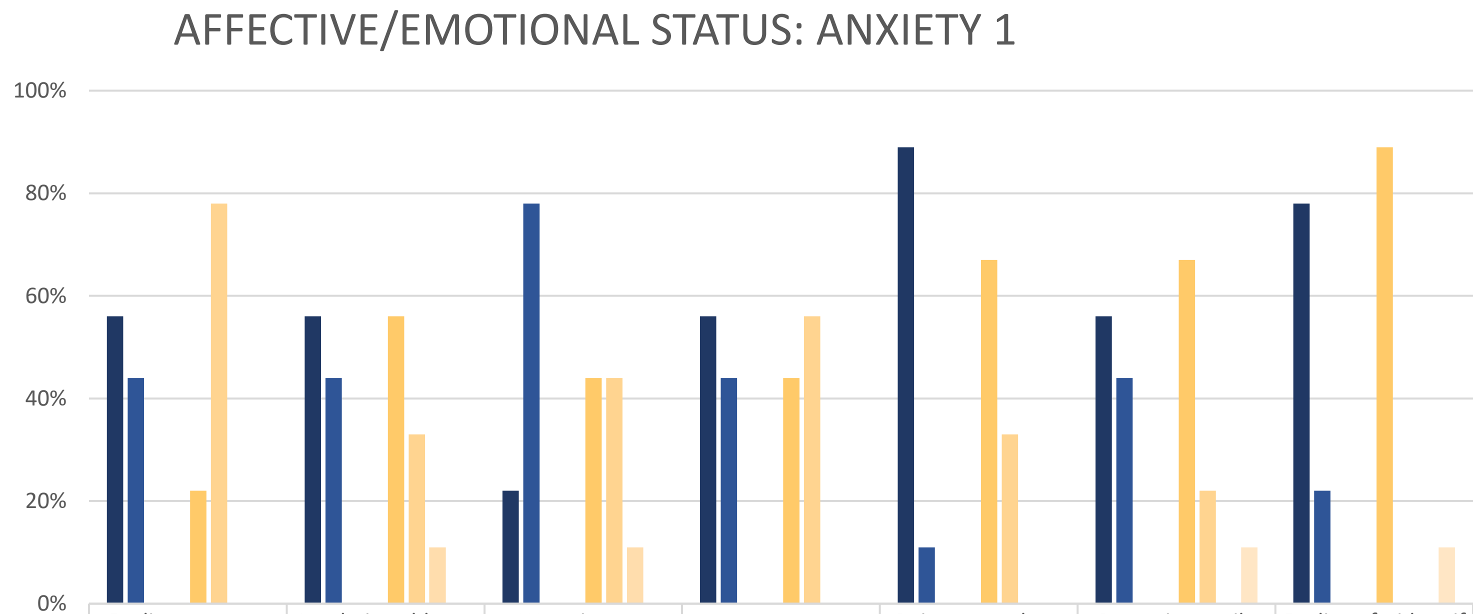
Figure 159: Affective/Emotional status: Depression 2. Teleworking use case.

Continuing with the affective/emotional status assessment, the participants were asked eight questions, that collect their general state with respect to a possible anxiety problem. As in the previous section, they were asked if over the last 2 weeks, how often have you been bothered by any of the following problems and this is reflected in the next graph:

- “Feeling nervous, anxious, or on edge”
- “Not being able to stop or control worrying”
- “Worrying too much about different things”
- “Trouble relaxing”
- “Being so restless that it is hard to sit still”
- “Becoming easily annoyed or irritable”
- “Feeling afraid, as if something awful might happen”

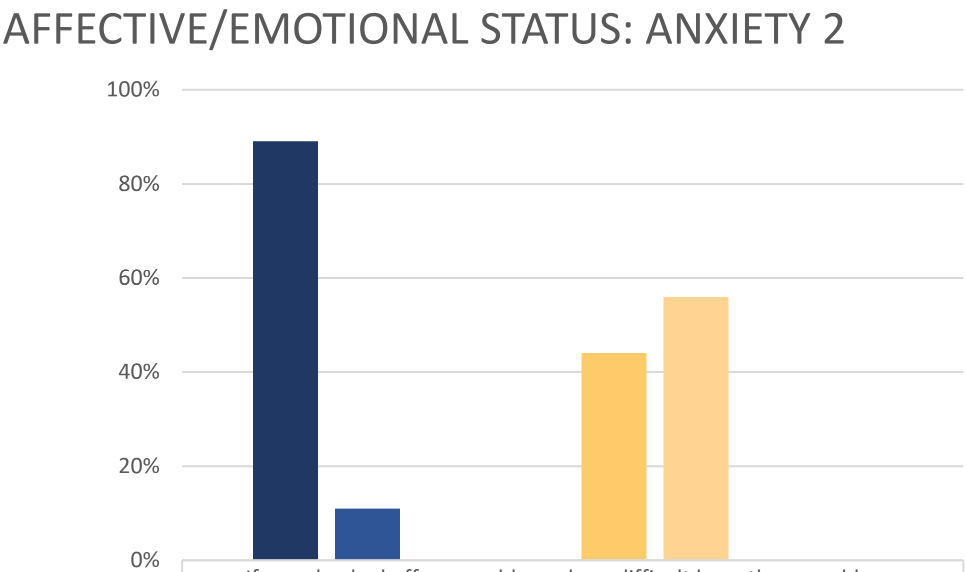
One final question completes this evaluation and is reflected in the Affective/Emotional status: Anxiety 2.

- “If you checked off any problems, how difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?”



	Feeling nervous, anxious, or on edge	Not being able to stop or control worrying	Worrying too much about different things	Trouble relaxing	Being so restless that it is hard to sit still	Becoming easily annoyed or irritable	Feeling afraid, as if something awful might happen
PRE Not at all	56%	56%	22%	56%	89%	56%	78%
PRE Several days	44%	44%	78%	44%	11%	44%	22%
PRE More than half of the days	0%	0%	0%	0%	0%	0%	0%
PRE Nearly every day	0%	0%	0%	0%	0%	0%	0%
POST Not at all	22%	56%	44%	44%	67%	67%	89%
POST Several days	78%	33%	44%	56%	33%	22%	0%
POST More than half of the days	0%	11%	11%	0%	0%	0%	0%
POST Nearly every day	0%	0%	0%	0%	0%	11%	11%

Figure 160: Affective/Emotional status: Anxiety 1. Teleworking use case.



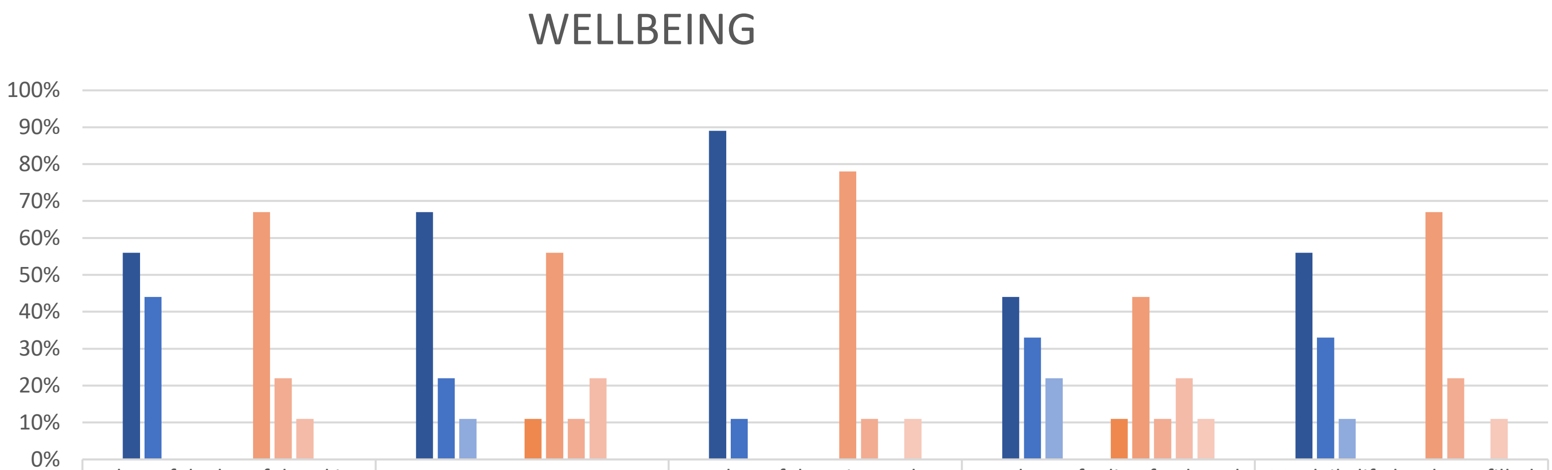
	If you checked off any problems, how difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?
PRE Not difficult at all	89%
PRE Somewhat difficult	11%
PRE Very difficult	0%
PRE Extremely difficult	0%
POST Not difficult at all	44%
POST Somewhat difficult	56%
POST Very difficult	0%
POST Extremely difficult	0%

Figure 161: Affective/Emotional status: Anxiety 2. Teleworking use case.

2.3.3.3.6 *General wellbeing & Quality of life*

The volunteers were asked about their subjective wellbeing and quality of life, firstly with five questions that can be seen in detail in the next graph.

- “I have felt cheerful and in good spirits”
- “I have felt calm and relaxed”
- “I have felt active and vigorous”
- “I woke up feeling fresh and rested”
- “My daily life has been filled with things that interest me”

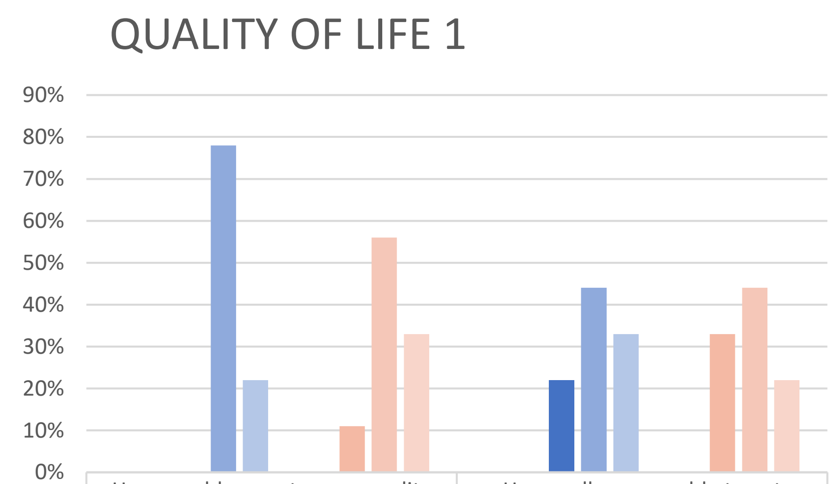


	I have felt cheerful and in good spirits	I have felt calm and relaxed	I have felt active and vigorous	I woke up feeling fresh and rested	My daily life has been filled with things that interest me
PRE All the time	0%	0%	0%	0%	0%
PRE Most of the time	56%	67%	89%	44%	56%
PRE More than half the time	44%	22%	11%	33%	33%
PRE Less than half the time	0%	11%	0%	22%	11%
PRE Some of the time	0%	0%	0%	0%	0%
PRE At no time	0%	0%	0%	0%	0%
POST All the time	0%	11%	0%	11%	0%
POST Most of the time	67%	56%	78%	44%	67%
POST More than half the time	22%	11%	11%	11%	22%
POST Less than half the time	11%	22%	0%	22%	0%
POST Some of the time	0%	0%	11%	11%	11%
POST At no time	0%	0%	0%	0%	0%

Figure 162: Wellbeing. Teleworking use case

The next questions assess the quality of life of the participants. Can be seen in detail in the next six graphs.

- "How would you rate your quality of life?"
- "How well are you able to get around?"



	How would you rate your quality of life?	How well are you able to get around?
PRE Very poor	0%	0%
PRE Poor	0%	0%
PRE Neither poor nor good	0%	22%
PRE Good	78%	44%
PRE Very good	22%	33%
POST Very poor	0%	0%
POST Poor	0%	0%
POST Neither poor nor good	11%	33%
POST Good	56%	44%
POST Very good	33%	22%

Figure 163: Quality of life 1. Teleworking use case.

- "How satisfied are you with your health?"

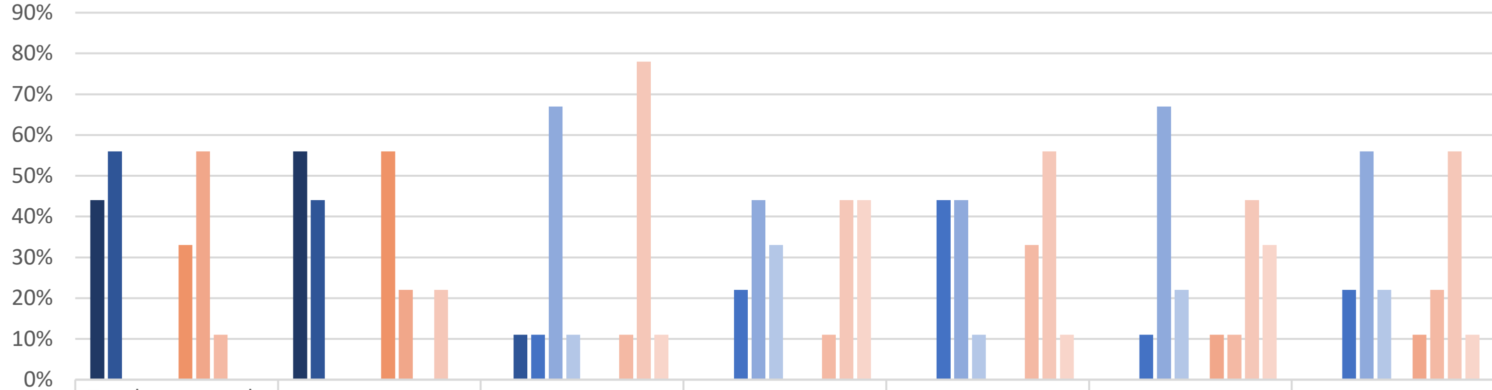
	How satisfied are you with your health?
PRE Very dissatisfied	0%
PRE Dissatisfied	11%
PRE Neither satisfied nor dissatisfied	11%
PRE Satisfied	56%
PRE Very satisfied	22%
POST Very dissatisfied	0%
POST Dissatisfied	11%
POST Neither satisfied nor dissatisfied	22%
POST Satisfied	44%
POST Very satisfied	22%

Figure 164: Quality of life 2. Teleworking use case.

Next, there are seven more questions that assess the quality of life of the WAOW tool users. The details are shown in the following image.

- "To what extent do you feel that physical pain prevents you from doing what you need to do?"
- "How much do you need any medical treatment to function in your daily life?"
- "How much do you enjoy life?"
- "To what extent do you feel your life to be meaningful?"
- "How well are you able to concentrate?"
- "How safe do you feel in your daily life?"
- "How healthy is your physical environment?"

QUALITY OF LIFE 3

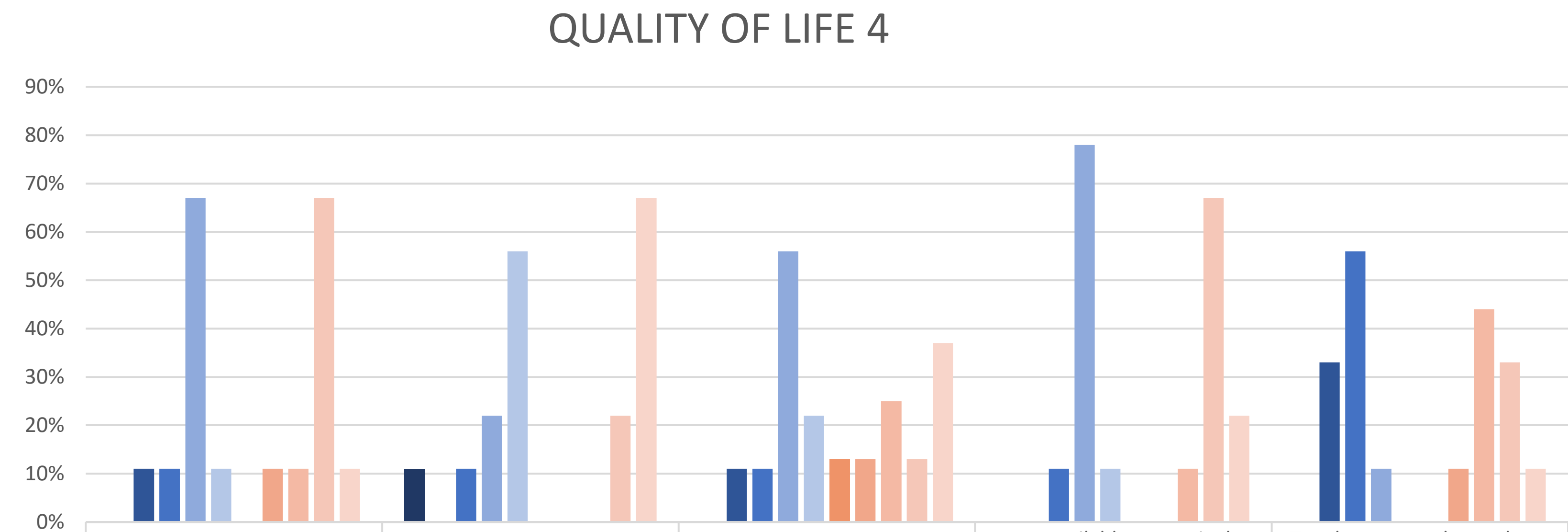


	To what extent do you feel that physical pain prevents you from doing what you need to do?	How much do you need any medical treatment to function in your daily life?	How much do you enjoy life?	To what extent do you feel your life to be meaningful?	How well are you able to concentrate?	How safe do you feel in your daily life?	How healthy is your physical environment?
PRE Not at all	44%	56%	0%	0%	0%	0%	0%
PRE A little	56%	44%	11%	0%	0%	0%	0%
PRE A moderate amount	0%	0%	11%	22%	44%	11%	22%
PRE Very much	0%	0%	67%	44%	44%	67%	56%
PRE An extreme amount	0%	0%	11%	33%	11%	22%	22%
POST Not at all	33%	56%	0%	0%	0%	0%	0%
POST A little	56%	22%	0%	0%	0%	11%	11%
POST A moderate amount	11%	0%	11%	11%	33%	11%	22%
POST Very much	0%	22%	78%	44%	56%	44%	56%
POST An extreme amount	0%	0%	11%	44%	11%	33%	11%

Figure 165: Quality of life 3. Teleworking use case.

Another five questions that evaluate the quality of life are shown in the next graph: “Quality of life 4”.

- “Do you have enough energy for everyday life?”
- “Are you able to accept your bodily appearance?”
- “Have you enough money to meet your needs?”
- “How available to you is the information that you need in your day-to-day life?”
- “To what extent do you have the opportunity for leisure activities?”

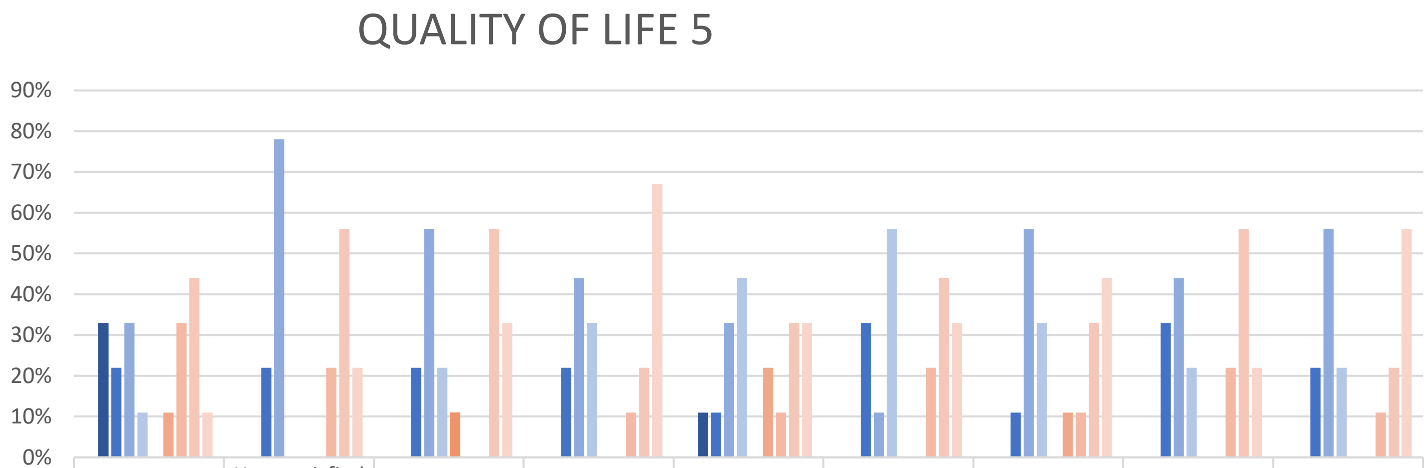


	Do you have enough energy for everyday life?	Are you able to accept your bodily appearance?	Have you enough money to meet your needs?	How available to you is the information that you need in your day-to-day life?	To what extent do you have the opportunity for leisure activities?
PRE Not at all	0%	11%	0%	0%	0%
PRE A little	11%	0%	11%	0%	33%
PRE Moderately	11%	11%	11%	11%	56%
PRE Mostly	67%	22%	56%	78%	11%
PRE Completely	11%	56%	22%	11%	0%
POST Not at all	0%	0%	13%	0%	0%
POST A little	11%	0%	13%	0%	11%
POST Moderately	11%	0%	25%	11%	44%
POST Mostly	67%	22%	13%	67%	33%
POST Completely	11%	67%	37%	22%	11%

Figure 166: Quality of life 4. Teleworking use case.

Continuing assessing the quality of life, the volunteers answered the next nine questions. The details can be seen in the next graph: "Quality of life 5".

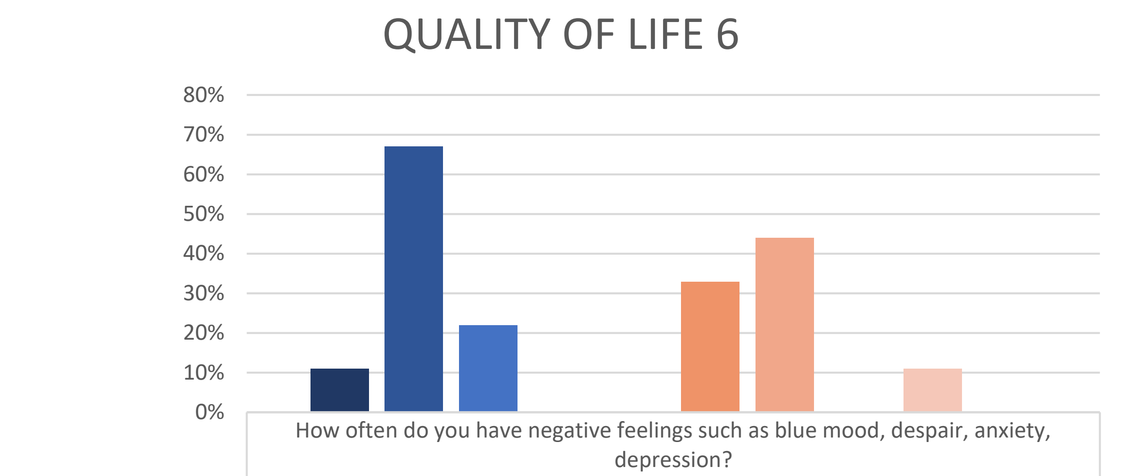
- "How satisfied are you with your sleep?"
- "How satisfied are you with your ability to perform your daily living activities?"
- "How satisfied are you with your capacity for work?"
- "How satisfied are you with yourself?"
- "How satisfied are you with your personal relationships?"
- "How satisfied are you with the support you get from your friends?"
- "How satisfied are you with the conditions of your living place?"
- "How satisfied are you with your access to health services?"
- "How satisfied are you with your transport?"



	How satisfied are you with your sleep?	How satisfied are you with your ability to perform your daily living activities?	How satisfied are you with your capacity for work?	How satisfied are you with yourself?	How satisfied are you with your personal relationships?	How satisfied are you with the support you get from your friends?	How satisfied are you with the conditions of your living place?	How satisfied are you with your access to health services?	How satisfied are you with your transport?
PRE Very dissatisfied	0%	0%	0%	0%	0%	0%	0%	0%	0%
PRE Dissatisfied	33%	0%	0%	0%	11%	0%	0%	0%	0%
PRE Neither satisfied nor dissatisfied	22%	22%	22%	22%	11%	33%	11%	33%	22%
PRE Satisfied	33%	78%	56%	44%	33%	11%	56%	44%	56%
PRE Very satisfied	11%	0%	22%	33%	44%	56%	33%	22%	22%
POST Very dissatisfied	0%	0%	11%	0%	0%	0%	0%	0%	0%
POST Dissatisfied	11%	0%	0%	0%	22%	0%	11%	0%	0%
POST Neither satisfied nor dissatisfied	33%	22%	0%	11%	11%	22%	11%	22%	11%
POST Satisfied	44%	56%	56%	22%	33%	44%	33%	56%	22%
POST Very satisfied	11%	22%	33%	67%	33%	33%	44%	22%	56%

Figure 167: Quality of life 5. Teleworking use case.

The last question to assess the quality of life of the participants was "How often do you have negative feelings such as blue mood, despair, anxiety, depression?" The participants selected the following frequency:



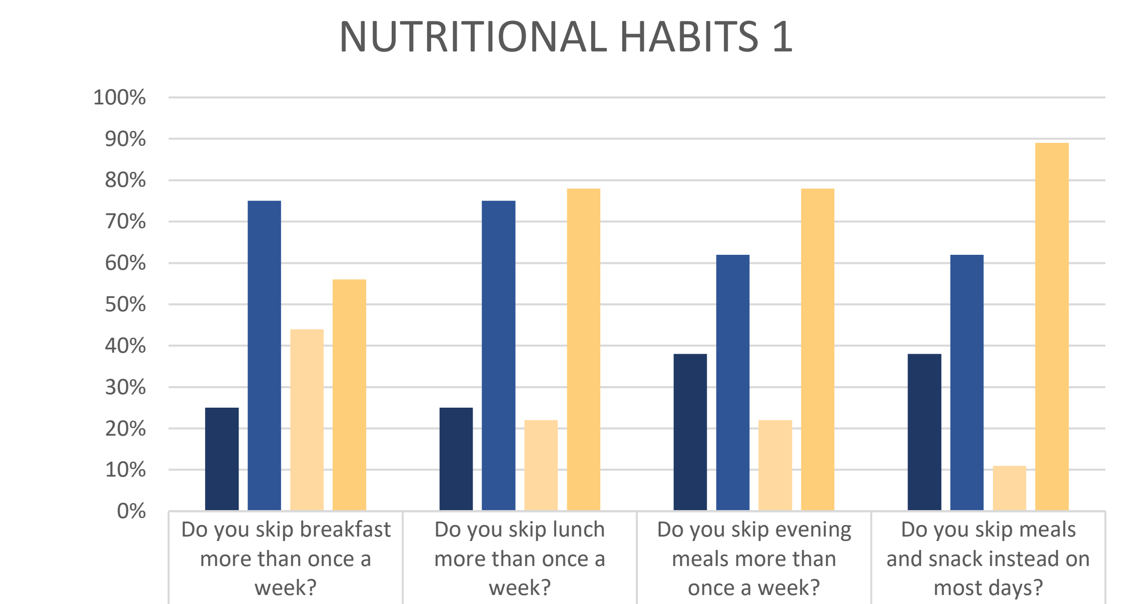
	How often do you have negative feelings such as blue mood, despair, anxiety, depression?
PRE Never	11%
PRE Seldom	67%
PRE Quite often	22%
PRE Very often	0%
PRE Always	0%
POST Never	33%
POST Seldom	44%
POST Quite often	0%
POST Very often	11%
POST Always	0%

Figure 168: Quality of life 6. Teleworking use case.

To assess the nutritional habits, essential in the general wellbeing of the participants, they were asked 30 questions that they had to answer with a simple yes or no.

The first four questions are related with the frequency they skip different meals during the week. A positive answer will mean a bad nutritional habit, while a negative will mean a good one. See the graph "Nutritional habits 1" below.

- "Do you skip breakfast more than once a week?"
- "Do you skip lunch more than once a week?"
- "Do you skip evening meals more than once a week?"
- "Do you skip meals and snack instead on most days?"



	Do you skip breakfast more than once a week?	Do you skip lunch more than once a week?	Do you skip evening meals more than once a week?	Do you skip meals and snack instead on most days?
PRE YES	25%	25%	38%	38%
PRE NO	75%	75%	62%	62%
POST YES	44%	22%	22%	11%
POST NO	56%	78%	78%	89%

Figure 169: Nutritional habits 1. Teleworking use case

The next six questions are about the choice of the teleworkers when they select their food or ingredients. A positive answer will mean a good nutritional habit, while a negative will mean a bad one. See the graph "Nutritional habits 2" below.

- "Do you eat more than 5 portions of fruit and/or vegetables every day?"
- "Do you eat more than 4 different varieties of fruit each week?"
- "Do you eat more than 4 different varieties of vegetables each week?"
- "Do you choose low-fat products when available?"
- "Do you choose baked, steamed, or grilled options when available, rather than fried foods (such as crisps and snacks, or fish and chips)?"
- "Do you opt for lean cuts of meat or remove visible fat – for example, removing the skin on chicken or the rind on bacon?"

	Do you eat more than 5 portions of fruit and/or vegetables every day?	Do you eat more than 4 different varieties of fruit each week?	Do you eat more than 4 different varieties of vegetables each week?	Do you choose low-fat products when available?	Do you choose baked, steamed or grilled options when available, rather than fried foods (such as crisps and snacks, or fish and chips)?	Do you opt for lean cuts of meat or remove visible fat – for example, removing the skin on chicken or the rind on bacon?
PRE YES	0%	33%	44%	67%	67%	62%
PRE NO	100%	67%	56%	33%	33%	38%
POST YES	11%	44%	78%	44%	78%	67%
POST NO	89%	56%	22%	56%	22%	33%

Figure 170: Nutritional habits 2. Teleworking use case

The following five questions are about the diet habits of the volunteers. A positive answer will mean a good nutritional habit, while a negative will mean a bad one. The details are shown in the graph "Nutritional habits 3" below.

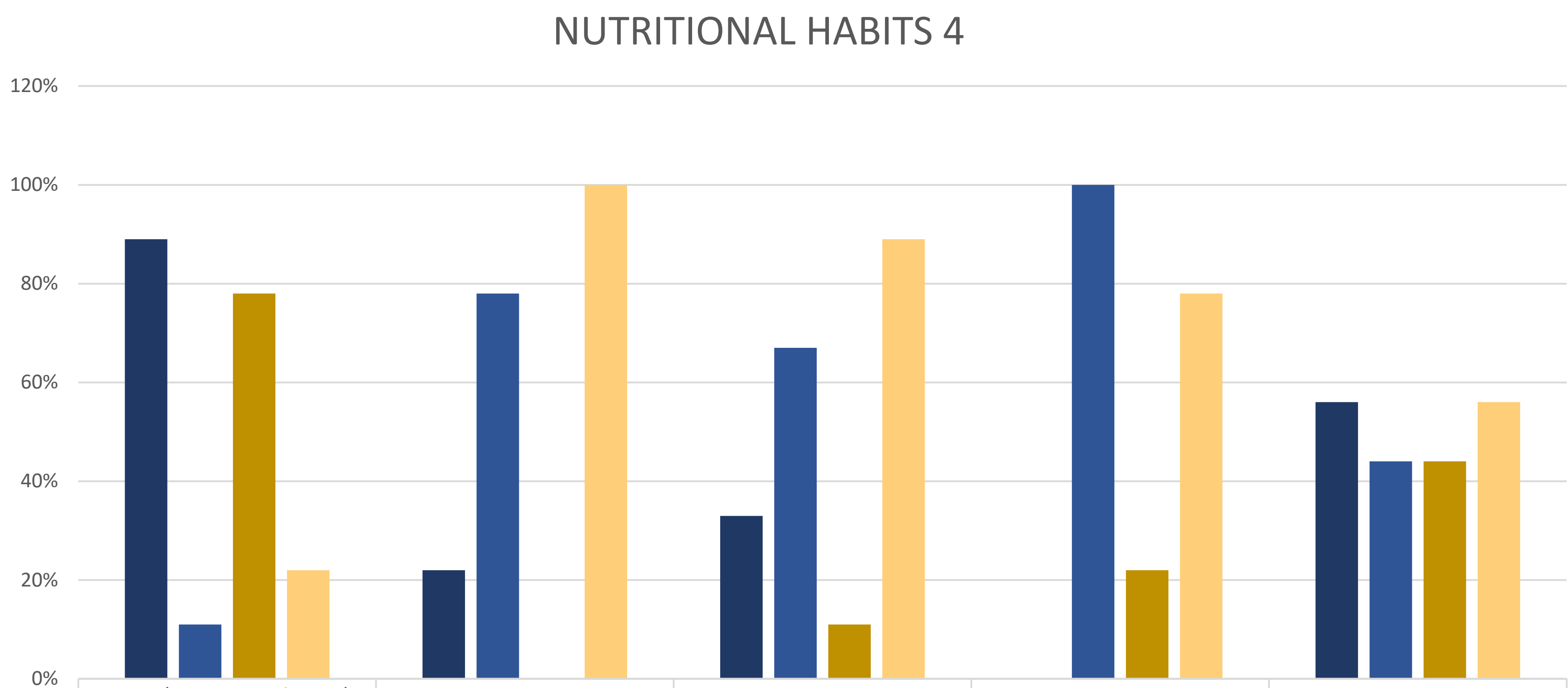
- "Did you eat any oily fish last week?"
- "Do you include some unsalted nuts and seeds in your diet?"
- "Do you regularly choose wholemeal bread or rolls rather than white?"
- "Do you regularly eat wholegrain cereals, with no added sugar?"
- "Do you regularly include pulses in your diet? For example, beans and lentils"

	Did you eat any oily fish last week?	Do you include some unsalted nuts and seeds in your diet?	Do you regularly choose wholemeal bread or rolls rather than white?	Do you regularly eat wholegrain cereals, with no added sugar?	Do you regularly include pulses in your diet? For example, beans and lentils.
PRE YES	33%	67%	44%	22%	67%
PRE NO	67%	33%	56%	78%	33%
POST YES	44%	56%	44%	22%	78%
POST NO	56%	44%	56%	67%	22%

Figure 171: Nutritional habits 3. Teleworking use case

The next five questions are about starchy foods, sugar and sweets. A positive answer will mean a bad nutritional habit, while a negative will mean a good one. The details are shown in the graph "Nutritional habits 4" below.

- "Do you base your main meals around starchy foods? For example, potatoes, pasta, rice, or bread"
- "Do you regularly eat sugar-coated breakfast cereals or add sugar to your breakfast cereals?"
- "Do you add sugar to your drinks?"
- "Do you regularly drink sweet fizzy drinks?"
- "Do you regularly eat cakes, sweets, chocolate, or biscuits at work?"

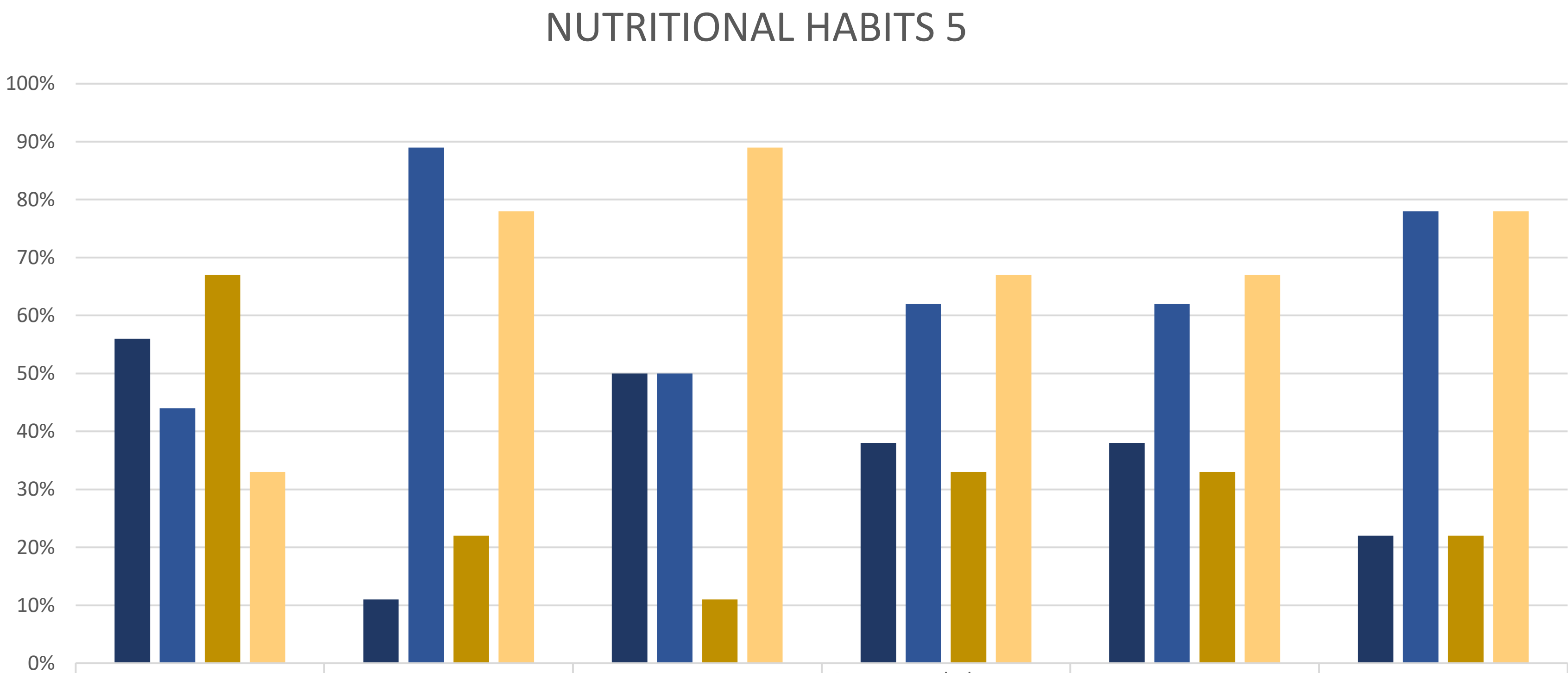


	Do you base your main meals around starchy foods? For example, potatoes, pasta, rice or bread.	Do you regularly eat sugar-coated breakfast cereals or add sugar to your breakfast cereals?	Do you add sugar to your drinks?	Do you regularly drink sweet fizzy drinks?	Do you regularly eat cakes, sweets, chocolate or biscuits at work?
PRE YES	89%	22%	33%	0%	56%
PRE NO	11%	78%	67%	100%	44%
POST YES	78%	0%	11%	22%	44%
POST NO	22%	100%	89%	78%	56%

Figure 172: Nutritional habits 4. Teleworking use case

The next six questions are about bad nutritional habits and high blood pressure. A positive answer will mean a bad nutritional habit, while a negative will mean a good one. The details are shown in the graph "Nutritional habits 5" below.

- "Do you regularly add salt to food during cooking?"
- "Do you regularly add salt to meals at the table?"
- "Do you regularly eat savoury snacks at work? For example, crisps or salted nuts?"
- "Do you regularly eat pre-prepared meals? For example, pre-prepared sandwiches, ready meals, or canned soups?"
- "Do you regularly eat processed meats such as ham or bacon, or smoked fish?"
- "Has your GP advised you that you have high blood pressure?"

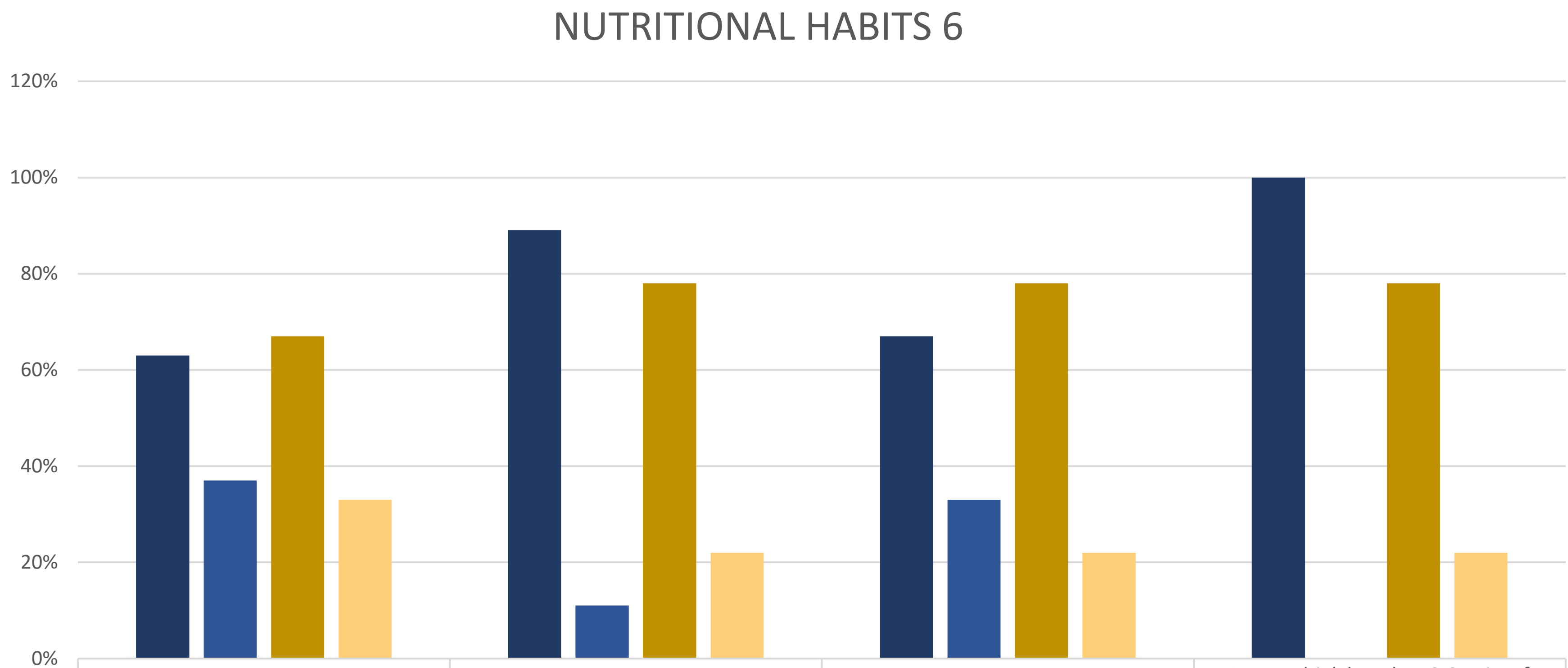


	Do you regularly add salt to food during cooking?	Do you regularly add salt to meals at the table?	Do you regularly eat savoury snacks at work? For example, crisps or salted nuts.	Do you regularly eat pre-prepared meals? For example, pre-prepared sandwiches, ready meals or canned soups.	Do you regularly eat processed meats such as ham or bacon, or smoked fish?	Has your GP advised you that you have high blood pressure?
PRE YES	56%	11%	50%	38%	38%	22%
PRE NO	44%	89%	50%	62%	62%	78%
POST YES	67%	22%	11%	33%	33%	22%
POST NO	33%	78%	89%	67%	67%	78%

Figure 173: Nutritional habits 5. Teleworking use case

The last four questions are about drinks choice and drinking frequency. A positive answer will mean a good nutritional habit, while a negative will mean a bad one. The details are shown in the graph “Nutritional habits 6” below.

- “Do you drink plenty of fluids at regular intervals during the working day?”
- “Do you opt for a variety of different drinks, including water, at work?”
- “Do you avoid sugary fizzy drinks?”
- “Do you drink less than 2-3 units of alcohol a day if you’re a woman, or less than 3-4 units of alcohol a day if you’re a man? 1 unit of alcohol is equivalent to 100ml of 10% ABV (alcohol by volume)”



	Do you drink plenty of fluids at regular intervals during the working day?	Do you opt for a variety of different drinks, including water, at work?	Do you avoid sugary fizzy drinks?	Do you drink less than 2-3 units of alcohol a day if you're a woman, or less than 3-4 units of alcohol a day if you're a man? 1 unit of alcohol is equivalent to 100ml of 10% ABV (alcohol by volume)
PRE YES	63%	89%	67%	100%
PRE NO	37%	11%	33%	0%
POST YES	67%	78%	78%	78%
POST NO	33%	22%	22%	22%

Figure 174: Nutritional habits 6. Teleworking use case

2.3.3.3.7 Life-changing

The last 2 additional questions were included in the post-questionnaire for the teleworkers. The first one, asked the participants if they have had any significant situation that could have affected to their lives. 4 of them answered this question in the affirmative, the 44%. These volunteers could have suffered some changes in their life, but they may not be due to the WAOW tool usage. None of them wanted to explain what kind of situation was the one affecting their lives.

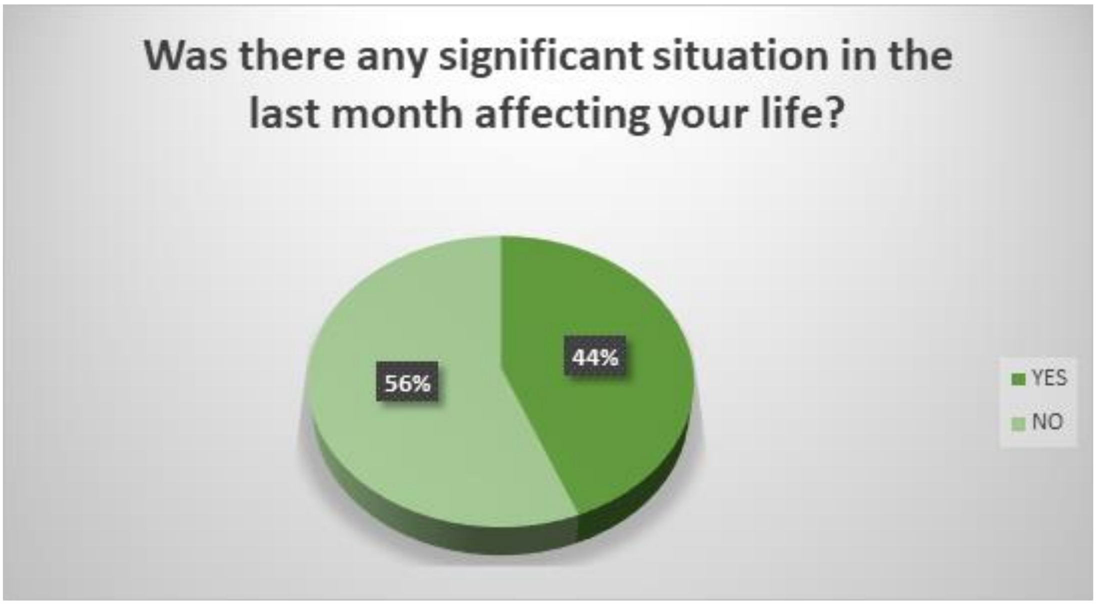


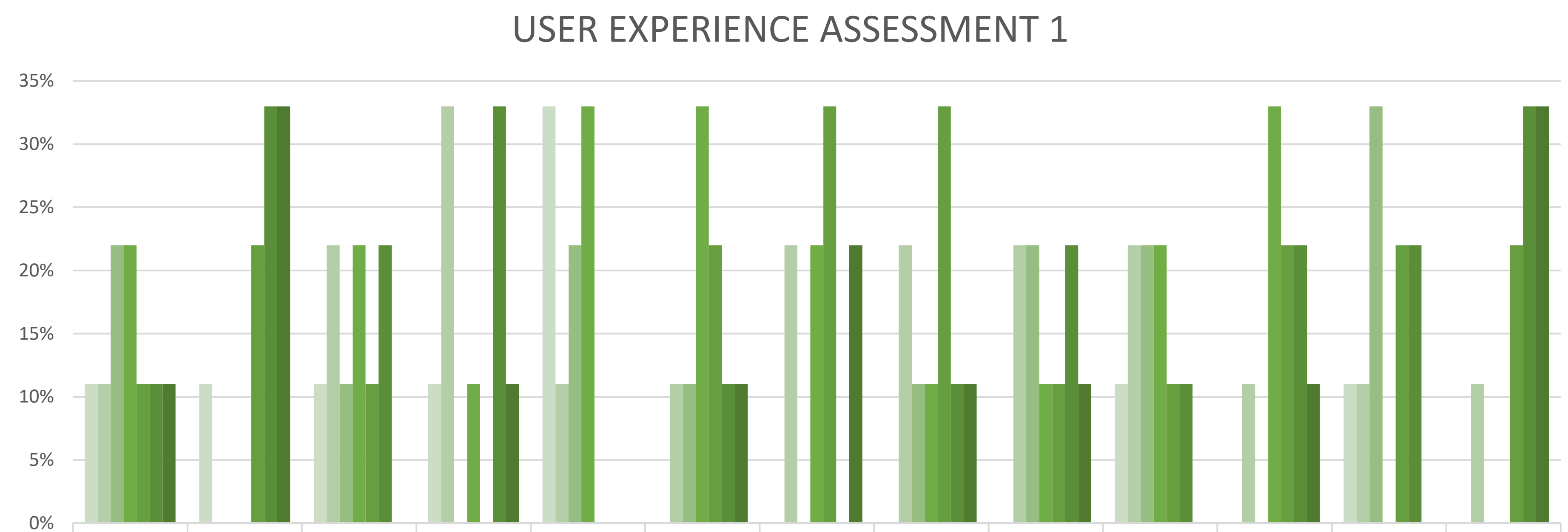
Figure 175: Life changing

2.3.3.3.8 User experience assessment

Next as in the other two use cases, the Teleworking volunteers were asked to evaluate their experience in understanding, learning, and using the WAOW tool. The details are described in the office use case user experience assessment.

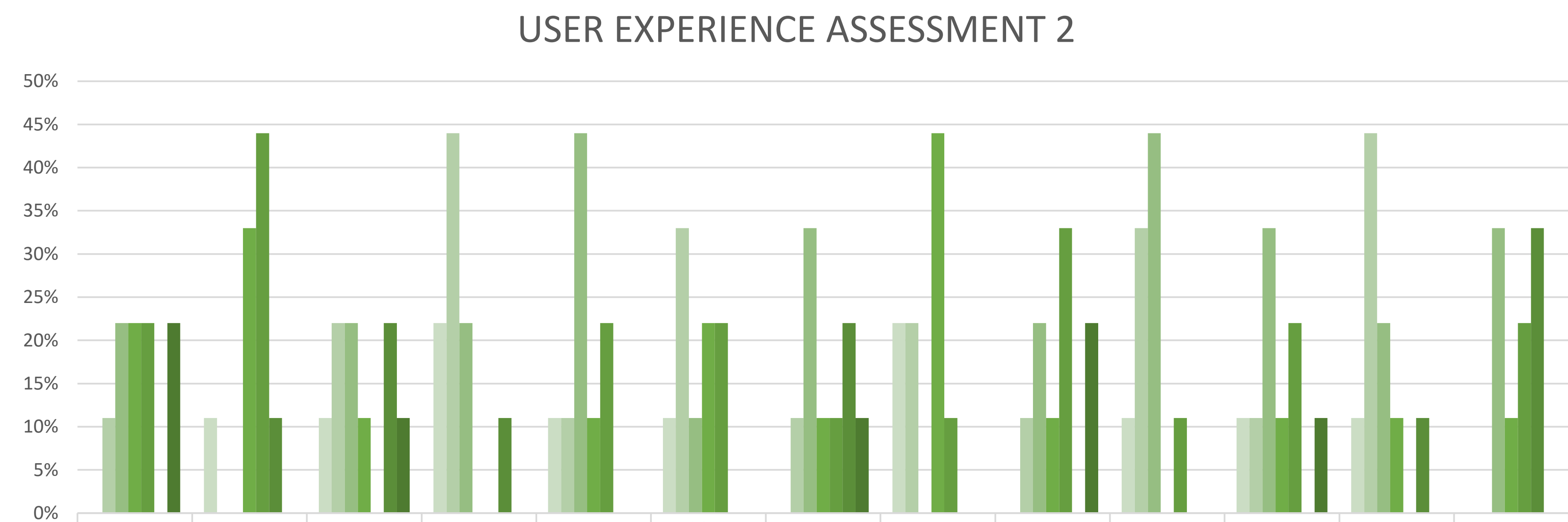
The numbers from 1 to 7 between the attributes represent gradations between the opposites. The volunteers can express their agreement with the attributes by selecting the number that most closely reflects their impression.

As there are many attributes to evaluate, the graph is divided in two parts.



	1 Annoying - 7 Enjoyable	1 Not understandable - 7 Understandable	1 Creative - 7 Dull	1 Easy to learn - 7 Difficult to learn	1 Valuable - 7 Inferior	1 Boring - 7 Exciting	1 Not interesting - 7 Interesting	1 Unpredictable - 7 Predictable	1 Fast - 7 Slow	1 Inventive - 7 Conventional	1 Obstructive - 7 Supportive	1 Good - 7 Bad	1 Complicated - 7 Easy
1	11%	11%	11%	11%	33%	0%	0%	0%	0%	11%	0%	11%	0%
2	11%	0%	22%	33%	11%	11%	22%	22%	22%	22%	11%	11%	11%
3	22%	0%	11%	0%	22%	11%	0%	11%	22%	22%	0%	33%	0%
4	22%	0%	22%	11%	33%	33%	22%	11%	11%	22%	33%	0%	0%
5	11%	22%	11%	0%	0%	22%	33%	33%	11%	11%	22%	22%	22%
6	11%	33%	22%	33%	0%	11%	0%	11%	22%	11%	22%	22%	33%
7	11%	33%	0%	11%	0%	11%	22%	11%	11%	0%	11%	0%	33%

Figure 176: User Experience Assessment 1. Teleworking use case



	1 Unlikeable - 7 Pleasing	1 Usual - 7 Leading edge	1 Unpleasant - 7 Pleasant	1 Secure - 7 Not secure	1 Motivating - 7 Demotivating	1 Meets expectations - 7 Does not meet expectations	1 Inefficient - 7 Efficient	1 Clear - 7 Confusing	1 Impractical - 7 Practical	1 Organized - 7 Cluttered	1 Attractive - 7 Unattractive	1 Friendly - 7 Unfriendly	1 Conservative - 7 Innovative
1	0%	11%	11%	22%	11%	11%	0%	22%	0%	11%	11%	11%	0%
2	11%	0%	22%	44%	11%	33%	11%	22%	11%	33%	11%	44%	0%
3	22%	0%	22%	22%	44%	11%	33%	0%	22%	44%	33%	22%	33%
4	22%	33%	11%	0%	11%	22%	11%	44%	11%	0%	11%	11%	11%
5	22%	44%	0%	0%	22%	22%	11%	11%	33%	11%	22%	0%	22%
6	0%	11%	22%	11%	0%	0%	22%	0%	0%	0%	0%	11%	33%
7	22%	0%	11%	0%	0%	0%	11%	0%	22%	0%	11%	0%	0%

Figure 177: User Experience Assessment 2. Teleworking use case

2.3.4 Discussion

There have been some changes in the participants' answers related with different aspects assessed.

We must consider that we asked to the participants in the post-questionnaire, if they had suffered any significant situations during the pilots that affected to their lives, and 4 participants in the office use case, 3 in the production, and 4 in the teleworking use case had, so the positive changes expected may be due to external causes to the WAOW tool.

Here are reflected the most positive changes in each use case. We are selected those aspects that have changed at least a 5% from the pre to the post in a positive way.

OFFICE

The volunteers reflected positive changes in their *physical working conditions* in:

- Hearing problems: These problems were selected a small part of the time or not at all with a frequency from 93% in the pre and 100% in the post.
- Muscular pains in shoulders, neck and/or upper limbs: The same happen with this health problem. Went from 64% in the pre to the 71% in the post.
- Headaches, eyestrain: These problems were reduced from the 43% in the pre to the 58% in the post.

The participants of the office use case pointed out that they felt that their safety wasn't at risk because of their work with less frequency. They answered positively with a 21% in the pre, but this was reduced to the 14% in the post.

In fact, these workers changed their mind when they were asked if their work affected to their health, because they answered yes, positively 14% in the pre, and 21% in the post.

The number of days of absence resulted from accidents, were reduce from 17% to 0% in the post, as those days resulted from health problems caused or made worse by work, that those who chose 0 days, increased the percentage from 50% to 63%.

About bothering problems, the participants selected the options 0 and 1 that mean that they were not bothered at all by the problem or very little, and we have selected these answers to reflect the positive changes. We have found an increase of the percentage from the pre to the post in the following problems:

- WEIGHT CHANGE (gain or loss of 5 libs. or more)
- DIARRHEA
- FAINTNESS
- NAUSEA AND/OR VOMITING
- STOMACH PAIN (E.G., CRAMPS)
- HEART PALPITATIONS (Heart pounding/racing)
- SHORTNESS OF BREATH when not exercising or working hard
- NUMBNESS OR TINGLING in parts of your body

- JOINT PAINS

The positive changes in the *psychosocial working conditions* in the office use case were:

When the participants were asked "How often have you been physically exhausted?" they selected not at all or a small part of the time in the pre by the 79%, while in the post, the 85%.

To the question "Is your work recognized and appreciated by the management?", the volunteers chose to a very large extent or to a large extent by 36% of them in the pre, and 43% in the post.

To the question How often do you get help and support from your immediate superior, if needed? the office workers selected always and often by the 35% in the pre and by the 43% in the post.

In the *physical status aspect*, the positive changes were the following:

To the question, "Does your health limit you in climbing SEVERAL flights of stairs?", the participants answered with No, not limited at all with a 79% in the pre and an 86% in the post.

When they were asked if "During the PAST 4 WEEKS 4 did you ACCOMPLISHED LESS than you would like AS A RESULT OF YOUR PHYSICAL HEALTH?", they answered no with a 79% in the pre and a 93% in the post.

The next question "During the PAST 4 WEEKS were your work or regular activities limited in the KIND of work or other activities AS A RESULT OF YOUR PHYSICAL HEALTH?", the participants changed from a 71% in the pre, to a 93% in the post.

To the question "During the PAST 4 WEEKS, how much did PAIN interfere with your normal work (including both work outside the home and housework)?" the participants selected the options not at all and a little bit with the 64% in the pre, and 79% in the post.

There was a positive change to the question "During the PAST 4 WEEKS, how much of the time has your PHYSICAL HEALTH OR EMOTIONAL PROBLEMS interfered with your social activities (like visiting with friends, relatives, etc.)?". The 64% of the volunteers chose in the pre a little of the time or none of the time, while in the post was selected by the 72%.

Now, we will describe the main positive changes in the *affective/mental status* of the office use case workers.

When they were asked about possible depression problems, their answers have a positive change in the following items:

- Feeling down, depressed, or hopeless: The option not at all, was selected by the 57% of the participants in the pre and by the 64% in the post.
- Feeling bad about yourself — or that you are a failure or have let yourself or your family down: The same frequency was chosen by 43% of the participants in the pre, and by the 71% in the post.

- Moving or speaking so slowly that other people could have noticed. Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual: The volunteers selected the option not at all 57% at the beginning of the pilots, and 79% at the end.

Asked about possible anxiety problems, we found positive changes in:

- Feeling afraid, as if something awful might happen: This sentence was marked as not at all by the 79% in the pre, and by the 93% in the post.
- Being asked about the anxiety problems, the participants found that these problems were not difficult at all for them for working, take care of things or get along with other people, by the 43% in the pre, and by the 71% in the post.

The *social relationships* section had positive changes in the following:

FAMILY

- The participants were asked about the number of relatives that they had seen or heard from at least once a month, they chose the amount of five to more than nine, by the 57% in the pre and 71% in the post.
- The same amount was chosen by the 29% of the participants in the pre, and the 50% in the post, when they were asked about with how many relatives can talk about private matters.
- The percentage increased too from 21% in the pre to 50% in the post, when the volunteers were asked about the number of relatives that could call for help.

FRIENDSHIPS

- The volunteers were asked about the number of friends that they had seen or heard from at least once a month, they chose the amount of five to more than nine, by the 57% in the pre and 64% in the post.
- The same amount was chosen by the 14% of the office workers in the pre, and the 36% in the post, when they were asked about with how many friends can talk about private matters.
- The percentage increased too from 14% in the pre to 29% in the post, when the volunteers were asked about the number of friends that could call for help.

When assessed the general *wellbeing and their quality of life,* we found the following positive changes:

- The participants felt cheerful and in good spirits with a frequency of all the time and most of the time by the 57% in the pre and 64% in the post.
- The same frequency was chosen by the 36% in the pre and the 43% in the post, when asked if their daily life was filled with interesting things.
- The volunteers were very satisfied or satisfied with their health with the 71% of them in the pre, while this figure increased to the 78% in the post.

- When asked "To what extent do you feel that physical pain prevents you from doing what you need to do?" the participants chose not at all or a little, from the 86% at the beginning of the pilots to the 92% at the end.
- To the question "How healthy is your physical environment?", the participants selected the same options by the 7% in the post, while none of them was selected in the pre.
- When the volunteers were asked if were able to accept their bodily appearance, in the pre, the 72% chose mostly or completely, while in the post the percentage increased to the 79%.
- Asked about their satisfaction with their sleep, the volunteers chose satisfied or very satisfied by the 21% in the pre, and by the 28% in the post.
- They felt that their satisfaction with their friends have increased from the 64% in the pre to the 72% in the post, selecting the same options.
- About their satisfaction with the conditions of their living place, the 79% from the pre changed to the 86% in the post.
- About the satisfaction with the health services, the 72% in the pre, was increased to 85% in the post.
- About the transport, the participants, chose very satisfied or satisfied with a 57% in the pre and a 64% in the post.
- To the question" How often do you have negative feelings such as blue mood, despair, anxiety, depression?" the participants selected never or seldom in the pre by the 71% while in the post increased to the 78%.

About the nutritional habits, the participants' positive changes were the following:

- The percentage of participants increased from the pre to the post in good nutritional habits such:
    - Eat more than 5 portions of fruit and/or vegetables every day.
    - Eat more than 4 different varieties of fruit each week.
    - Choose low-fat products when available.
    - Choose baked, steamed, or grilled options when available, rather than fried foods.
    - Eat oily fish last week.
    - Drink plenty of fluids at regular intervals during the working day
- The percentage of participants that avoid the following bad nutritional habits increased from the pre to the post:
    - Skip meals and snack instead most of the days
    - Base your main meals around starchy foods.
    - Drink sweet fizzy drinks.

We continue with the *health-related physical activity* positive changes.

When the participants were asked about the vigorous physical activities, the 57% said that didn't do any at all in the pre, while in the post, this figure lowered to the 29%.

To the number of days that they walked at least 10 minutes, those who chose the 7 days of the week, increased the percentage from 50% to the 72%.

PRODUCTION

The volunteers reflected improvement in their *physical working conditions* in:

- The participants of the production use case pointed out that they felt that their safety wasn't at risk because of their work with less frequency. They answered positively with a 27% in the pre, but this was reduced to the 15% in the post.
- A positive change was observed in the sleeping problems section, when to the problem, waking up repeatedly during the sleep, the 45% chose less often and never, in the pre, while in the post, the figure increased to the 54%.

About bothering problems, the participants selected the options 0 and 1 that mean that they were not bothered at all by the problem or very little, and we have selected these problems to reflect the positive changes where the percentage of these options has increased:

- DIZZINES
- FAINTNESS
- HANDS TREMBLING
- HEART PALPITATIONS (Heart pounding/racing)
- JOINT PAINS
- COLD OR COUGH

The positive changes in the *psychosocial working conditions* in the production use case were:

When the participants were asked "How often have you felt worn out?" they selected not at all or a small part of the time in the pre by the 45%, while in the post, the 69%. The same happens with the question "How often have you been tense?", the same options were selected by the 36% in the pre, while in the post this figure increased to the 54%.

To the question "Is your work recognized and appreciated by the management?", the volunteers chose to a very large extent or to a large extent by 18% of them in the pre, and 38% in the post.

To the question How often do you get help and support from your immediate superior, if needed? the production workers selected always and often by the 45% in the pre and by the 69% in the post.

The participants were asked two questions about work life conflict. To the question "Do you feel that your work drains so much of your energy that it has a negative effect on your private life?", the participants selected to a very small extent or to a small extent by the 72% of them in the pre, while this figure increased to the 77% in the post. The other question, "Do you feel that your work takes so much of your time that it has a negative effect on your private life?", the participants, chose the same options by the 72% in the pre, and by the 84% in the post.

In the *physical status aspect*, the positive changes were the following:

The production workers value their health as excellent or very good by the 9% in the pre, while in the post this percentage has an increase to the 23%.

To the question "During the PAST 4 WEEKS, how much did PAIN interfere with your normal work (including both work outside the home and housework)?" the participants selected the options not at all and a little bit with the 54% in the pre, and 85% in the post.

There was a positive change to the question "During the PAST 4 WEEKS, how much of the time has your PHYSICAL HEALTH OR EMOTIONAL PROBLEMS interfered with your social activities (like visiting with friends, relatives, etc.)?". The 54% of the volunteers chose in the pre a little of the time or none of the time, while in the post was selected by the 61%.

In the section of *cognitive/mental status*, the participants' answers showed a positive change in the following:

- The level of effort or mental concentration that their job requires was valued as very low or low by the 9% of the participants in the pre, while this percentage was increased to the 23% in the post.
- The participants selected the options strongly disagree or disagree, when they were asked if they had a hard time relaxing after work, by the 45% in the pre, and the 77% in the post.

Now, we will describe the main positive changes in the *affective/mental status* of the production use case workers.

When they were asked about possible depression problems, their answers have a positive change in the following items:

- Little interest or pleasure in doing things: The 27% of the volunteers chose the option not at all in the pre, while in the post, was selected by the 33%.
- Feeling down, depressed, or hopeless: The option not at all, was selected by the 64% of the participants in the pre and by the 75% in the post.
- Feeling tired or having little energy: The 18% chose not at all in the pre, while the 33% selected this option in the post.
- Poor appetite or overeating: The volunteers selected the option not at all 45% at the beginning of the pilots, and 50% at the end.
- Feeling bad about yourself — or that you are a failure or have let yourself or your family down: The same frequency was chosen by 64% of the participants in the pre, and by the 92% in the post.
- Trouble concentrating on things, such as reading the newspaper or watching television: The 73% of the volunteers chose the option not at all in the pre, while in the post, was selected by the 85%.
- Moving or speaking so slowly that other people could have noticed. Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual: The volunteers selected the option not at all 73% at the beginning of the pilots, and 85% at the end.
- Being asked about the depression problems, the participants found that these problems were not difficult at all for them for working, taking care of

things, or getting along with other people, by the 67% in the pre, and by the 85% in the post.

Asked about possible anxiety problems, we found positive changes in:

- Worrying too much about different things: The 27% of the participants in the pre, chose not at all, while in the post, was selected by the 62%.
- Trouble relaxing: Was selected like not at all by the 36% of the volunteers in the pre, and in the post by the 77%.
- Being so restless that it is hard to sit still: The option not at all, was selected by the 55% of the participants in the pre and by the 62% in the post.
- Becoming easily annoyed or irritable: The volunteers selected the option not at all 45% at the beginning of the pilots, and 54% at the end.
- Feeling afraid, as if something awful might happen: This sentence was marked as not at all by the 73% in the pre, and by the 85% in the post.
- Being asked about the anxiety problems, the participants found that these problems were not difficult at all for them for working, taking care of things, or getting along with other people, by the 64% in the pre, and by the 85% in the post.

There weren't any positive changes in the *social relationships* section.

When assessed the general *wellbeing and their quality of life,* we found the following positive changes:

- The participants felt cheerful and in good spirits with a frequency of all the time and most of the time by the 50% in the pre and 73% in the post.
- To the sentence "I have felt calm and relaxed", the volunteers answered with the same options, the 45% of them in the pre, and 84% in the post.
- The 50% of the participants felt active and vigorous, all the time or most of the time, in the pre, and this amount increased to the 69% in the post.
- 36% of the volunteers woke up feeling fresh and rested all the time, or most of the time, in the pre, while in the post were the 46%.
- The same options were chosen by the 63% in the pre and the 77% in the post, when asked if their daily life was filled with interesting things.
- The volunteers were very satisfied or satisfied with their health with the 54% of them in the pre, while this figure increased to the 77% in the post.
- When asked "How much do you need any medical treatment to function in your daily life?" the participants chose not at all or a little, from the 82% at the beginning of the pilots to the 93% at the end.
- The 63% of the participants felt extremely or very much safe in their daily life, in the pre, while in the post, was the 69% of them.
- To the question "Do you have enough energy for everyday life?", the participants selected mostly or completely by the 54% in the pre, while in the post the figure increased to the 61%.
- When the volunteers were asked if were able to accept their bodily appearance, in the pre, the 63% chose mostly or completely, while in the post the percentage increased to the 77%.

- The same options were chosen by 45% of the participants in the pre, and by the 69% in the post when were asked "How available to you is the information that you need in your day-to-day life?"
- To the question "To what extent do you have the opportunity for leisure activities?" the participants chose the same options with the 15% in the post and none in the pre.
- The volunteers selected the options very satisfied or satisfied to the next questions:
    - Satisfaction with their sleep, 36% in the pre, and by the 46% in the post.
    - Satisfaction with their ability to perform your daily living activities, 54% in the pre, and 77% in the post.
    - Satisfaction with their capacity for work, 45% in the pre, and 69% in the post.
    - Satisfaction with themselves, 80% in the pre, and 92% in the post.
    - Satisfaction with their personal relationships, 80% in the pre, and 93% in the post.
    - Satisfaction with the support from their friends, 81% in the pre, and 92% in the post.
    - Satisfaction with the conditions of their living place, the 80% from the pre changed to the 93% in the post.
    - Satisfaction with the health services, the 45% in the pre, was increased to 61% in the post.
    - About the transport, 60% in the pre and a 67% in the post.
- To the question" How often do you have negative feelings such as blue mood, despair, anxiety, depression?" the participants selected never or seldom in the pre by the 30% while in the post increased to the 100%.

About the nutritional habits, the production participants' positive changes were the following:

- The percentage of participants increased from the pre to the post in good nutritional habits such:
    - Eat more than 5 portions of fruit and/or vegetables every day.
    - Eat more than 4 different varieties of fruit each week.
    - Opt for lean cuts of meat or remove visible fat.
    - Include some unsalted nuts and seeds in the diet.
    - Eat wholegrain cereals, with no added sugar
    - Drink plenty of fluids at regular intervals during the working day.
    - Opt for a variety of different drinks, including water, at work.
    - Drink less than 2-3 units of alcohol a day if you're a woman, or less than 3-4 units of alcohol a day if you're a man.
- The percentage of participants that avoid the following bad nutritional habits increased from the pre to the post:
    - Skip lunch more than once a week.
    - Regularly eat sugar-coated breakfast cereals or add sugar to your breakfast cereals.
    - Add salt to food during cooking.

- Regularly eat pre-prepared meals.

We continue with the *health-related physical activity* positive changes.

When the production participants were asked about the vigorous physical activities, the 73% said that didn't do any at all in the pre, while in the post, this figure lowered to the 62%.

About the moderate physical activities, the 73% said that didn't do any at all in the pre, while in the post, this figure lowered to the 46%.

TELEWORKING

The teleworkers reflected upturns in their *physical working conditions* in:

- Backache: This problem was selected with the frequency of a small part of the time or not at all with by the 44% in the pre and 67% in the post.
- Muscular pains in shoulders, neck and/or upper limbs: The same happen with this health problem. Went from 66% in the pre to the 89% in the post.
- Muscular pains in lower limbs: The volunteers chose the same options with the 55% in the pre, and the 78% at the end of the pilots.
- Headaches, eyestrain: These problems were reduced from the 78% in the pre to the 88% in the post.

Sleeping problems: There were three questions with a positive change in the participants, from the pre to the post.

- Difficulty falling asleep, the participants selected never and less often with the 77% in the pre, and in the post, the 89%.
- Waking up repeatedly during the sleep, the 66% chose less often and never, in the pre, while in the post, the figure increased to the 78%.
- Waking up with a feeling of exhaustion and fatigue, the 44% of the volunteers in the pre, changed to a 55% in the post.

About bothering problems, the participants selected the options 0 and 1 that mean that they were not bothered at all by the problem or very little, and we have selected these problems to reflect the positive changes where the percentage of these options has increased:

- WEIGHT CHANGE (gain or loss of 5 libs. or more)
- CONSTIPATION
- DIARRHEA
- HANDS TREMBLING
- STUFFY HEAD OR NOSE
- BLURRED VISION
- NOSEBLEED
- COLD OR COUGH

The positive changes in the *psychosocial working conditions* in the teleworking use case were:

When the participants were asked “How often have you felt worn out?” they selected not at all or a small part of the time in the pre by the 11%, while in the post, the 66%.

The same happens with the question “How often have you been tense?”, the same options were selected by the 22% in the pre, while in the post this figure increased to the 44%.

The same options were chosen by the 33% in the pre and by the 67% in the post, when the participants were asked “How often have you been emotionally exhausted?”.

The participants were asked two questions about work life conflict. To the question "Do you feel that your work drains so much of your energy that it has a negative effect on your private life?", the participants selected to a very small extent or to a small extent by the 33% of them in the pre, while this figure increased to the 55% in the post. The other question, "Do you feel that your work takes so much of your time that it has a negative effect on your private life?", the participants, chose the same options by the 22% in the pre, and by the 55% in the post.

In the *physical status aspect*, the positive changes were the following:

We found a positive change from the pre to the post, when the volunteers answered the question "Does your health limit you in climbing SEVERAL flights of stairs?", in the pre, the 56% selected not at all, while this percentage was increased in the post to the 78%.

When the participants were asked "During the PAST 4 WEEKS 4 did you ACCOMPLISHED LESS than you would like AS A RESULT OF YOUR PHYSICAL HEALTH?", they answered no with the 67% in the pre, but this figure increased to the 89% in the post.

The same happened with the question "During the PAST 4 WEEKS were your work or regular activities limited in the KIND of work or other activities AS A RESULT OF YOUR PHYSICAL HEALTH?", they obtained the same percentage.

There was a positive change to the question "During the PAST 4 WEEKS, how much of the time has your PHYSICAL HEALTH OR EMOTIONAL PROBLEMS interfered with your social activities (like visiting with friends, relatives, etc.)?". The 63% of the volunteers chose in the pre a little of the time or none of the time, while in the post was selected by the 77%.

In the section of *cognitive/mental status*, the participants' answers showed a positive change in the following:

To the sentence "I feel exhausted when I wake up in the morning and have to face another day of work", the participants selected the options strongly disagree or somewhat disagree by the 44% of the participants in the pre, while this percentage was increased to the 66% in the post.

Now, we will describe the main positive changes in the *affective/mental status* of the teleworking use case volunteers.

When they were asked about possible depression problems, their answers have a positive change in the following items:

- Little interest or pleasure in doing things: The 44% of the volunteers chose the option not at all in the pre, while in the post, was selected by the 67%.
- Trouble falling or staying asleep or sleeping too much: The option not at all, was selected by the 56% of the participants in the pre and by the 67% in the post.
- Feeling tired or having little energy: The 22% chose not at all in the pre, while the 33% selected this option in the post.

- Moving or speaking so slowly that other people could have noticed. Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual: The volunteers selected the option not at all 78% at the beginning of the pilots, and 89% at the end.
- Thoughts that you would be better off dead or of hurting yourself in some way: The volunteers selected the option not at all by the 89% at the beginning of the pilots, and 100% at the end.

Asked about possible anxiety problems, we found positive changes in:

- Worrying too much about different things: The 22% of the participants in the pre, chose not at all, while in the post, was selected by the 44%.
- Becoming easily annoyed or irritable: The volunteers selected the option not at all 56% at the beginning of the pilots, and 67% at the end.
- Feeling afraid, as if something awful might happen: This sentence was marked as not at all by the 78% in the pre, and by the 89% in the post.

We only found one positive change in the *social relationships* section:

- About the FRIENDHIPS, the percentage of the options selected from five to more than nine friends, increased too from 11% in the pre to 22% in the post, when the volunteers were asked about the number of friends that could call for help.

When assessed the general *wellbeing and their quality of life,* we found the following positive changes:

- The participants felt cheerful and in good spirits with a frequency of all the time and most of the time by the 56% in the pre and 67% in the post.
- 44% of the volunteers woke up feeling fresh and rested all the time, or most of the time, in the pre, while in the post were the 55%.
- The same options were chosen by the 56% in the pre and the 67% in the post, when asked if their daily life was filled with interesting things.
- To the question "How much do you enjoy life?" the 78% chose the options very much or extremely in the pre, while in the post, was chosen by the 89%.
- The same options were chosen by the 77% of the volunteers in the pre, when they were asked "To what extent do you feel your life to be meaningful?", while in the post was selected by the 88%.
- About their ability to concentrate, the 55% selected the options very much or extremely in the pre, while at the end of the pilots was chosen by the 67%.
- When the volunteers were asked if were able to accept their bodily appearance, in the pre, the 78% chose mostly or completely, while in the post the percentage increased to the 89%.
- To the question "To what extent do you have the opportunity for leisure activities?" the participants chose the same options with the 11% in the pre and the 44% in the post.

- The volunteers selected the options very satisfied or satisfied to the next questions:
    - Satisfaction with their sleep, 44% in the pre, and by the 55% in the post.
    - Satisfaction with their capacity for work, 78% in the pre, and 89% in the post.
    - Satisfaction with themselves, 77% in the pre, and 89% in the post.
    - Satisfaction with the support from their friends, 67% in the pre, and 77% in the post.
    - Satisfaction with the health services, the 66% in the pre, was increased to 78% in the post.
    - About the transport, 60% in the pre and a 67% in the post.
- To the question" How often do you have negative feelings such as blue mood, despair, anxiety, depression?" the participants selected never or seldom in the pre by the 30% while in the post increased to the 100%.

About the nutritional habits, the teleworking participants' positive changes were the following:

- The percentage of participants increased from the pre to the post in good nutritional habits such:
    - Eat more than 5 portions of fruit and/or vegetables every day.
    - Eat more than 4 different varieties of fruit each week.
    - Eat more than 4 different varieties of vegetables each week.
    - Choose baked, steamed, or grilled options when available, rather than fried foods.
    - Opt for lean cuts of meat or remove visible fat.
    - Eat any oily fish last week.
    - Regularly include pulses in the diet.
    - Avoid sugary fizzy drinks.
- The percentage of participants that avoid the following bad nutritional habits increased from the pre to the post:
    - Skip evening meals more than once a week.
    - Skip meals and snack instead on most days.
    - Base your main meals around starchy foods.
    - Regularly eat sugar-coated breakfast cereals or add sugar to the breakfast cereals.
    - Add sugar to the drinks.
    - Regularly eat cakes, sweets, chocolate, or biscuits at work.
    - Regularly eat savoury snacks at work.
    - Regularly eat pre-prepared meals.
    - Regularly eat processed meats such as ham or bacon, or smoked fish.

Next, there are explained the *health-related physical activity* positive changes.

When the teleworking participants were asked about the moderate physical activities, the 33% said that didn't do any at all in the pre, while in the post, this figure lowered to the 22%.

With the descriptions made above, comparing the three use cases, we now are going to describe which use case had more positive changes in the different areas measured:

- *Physical Working Conditions: We found more positive changes in this area in the office and teleworking use cases than in the production use case.*
- *Psychosocial working conditions: There are more positive changes in the production use case than in the other two in this regard.*
- *Physical status: The office use case is the use case that shows more positive changes compared with the other two use cases in this aspect.*
- *Cognitive/Mental status: There are more positive changes in this assessed area in the production use case.*
- *Affective/Emotional status: The use case that found more beneficial changes in this aspect was the production use case.*
- Social relationships: *We verify more changes in the office use case in this section than in the other use cases.*
- *General wellbeing & Quality of life: The production use case was the use case that counts more positive changes in this area.*
- *Health-related physical activity: The office and the production use cases, obtained more positive changes in this regard than the teleworking use case.*

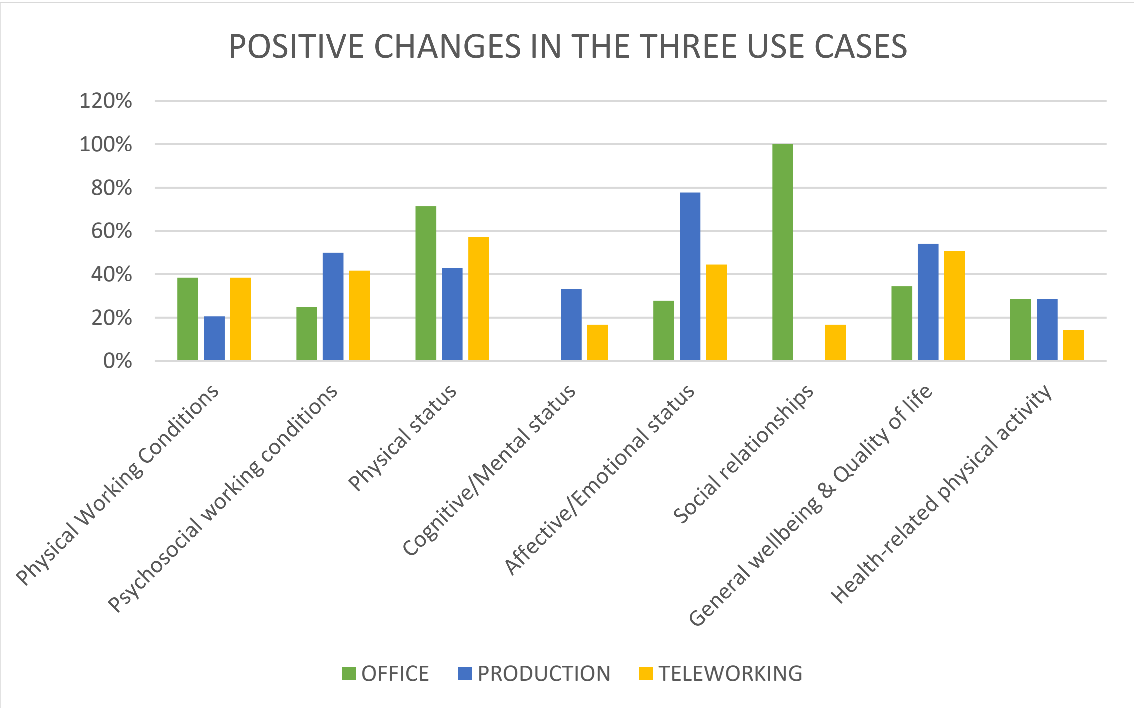


Figure 178: Positive changes in the three use cases.

Finally, we have compared the positive items of the *User experience questionnaire*, in the three use cases. The questions are rated from 1 to 7, so the higher is the score, the better the WAOW tool will be rated.

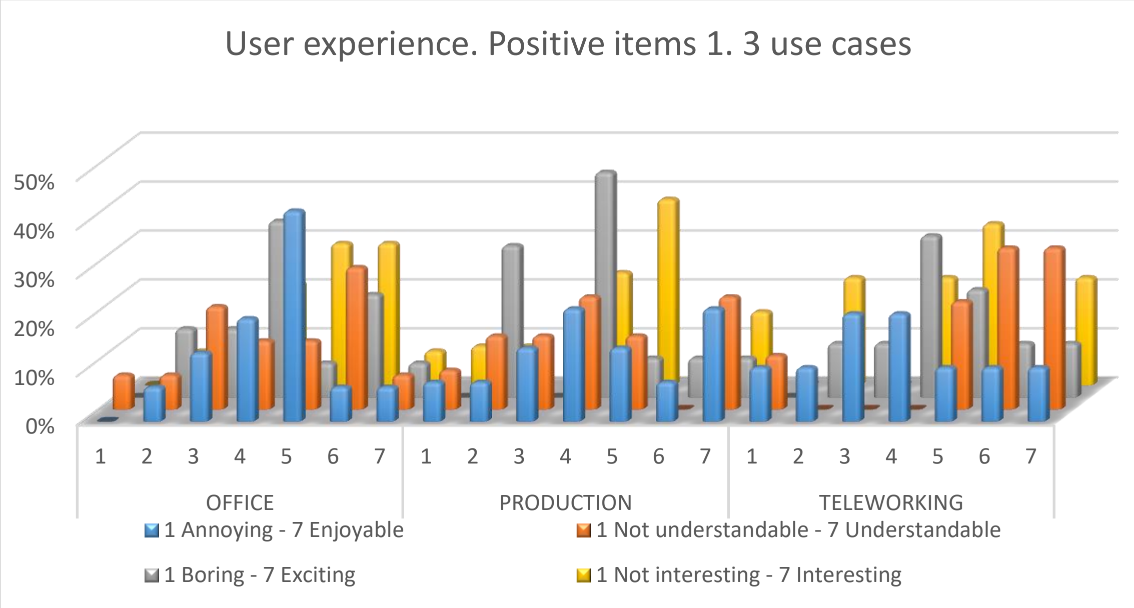


Figure 179: User experience questionnaire. Positive items 1. 3 use cases.

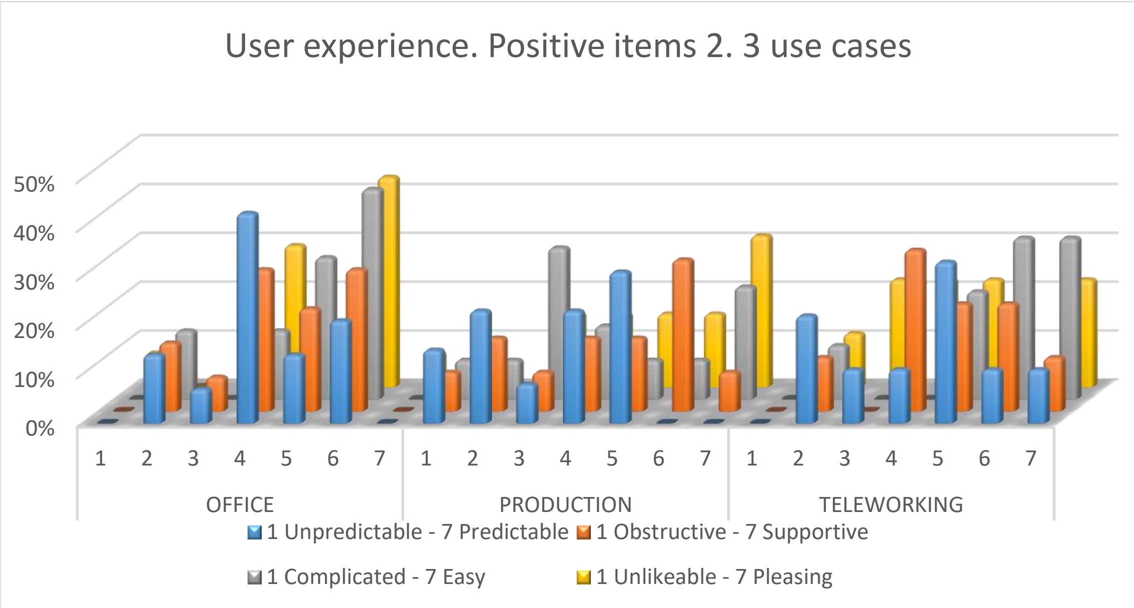


Figure 180: User experience questionnaire. Positive items 2. 3 use cases.

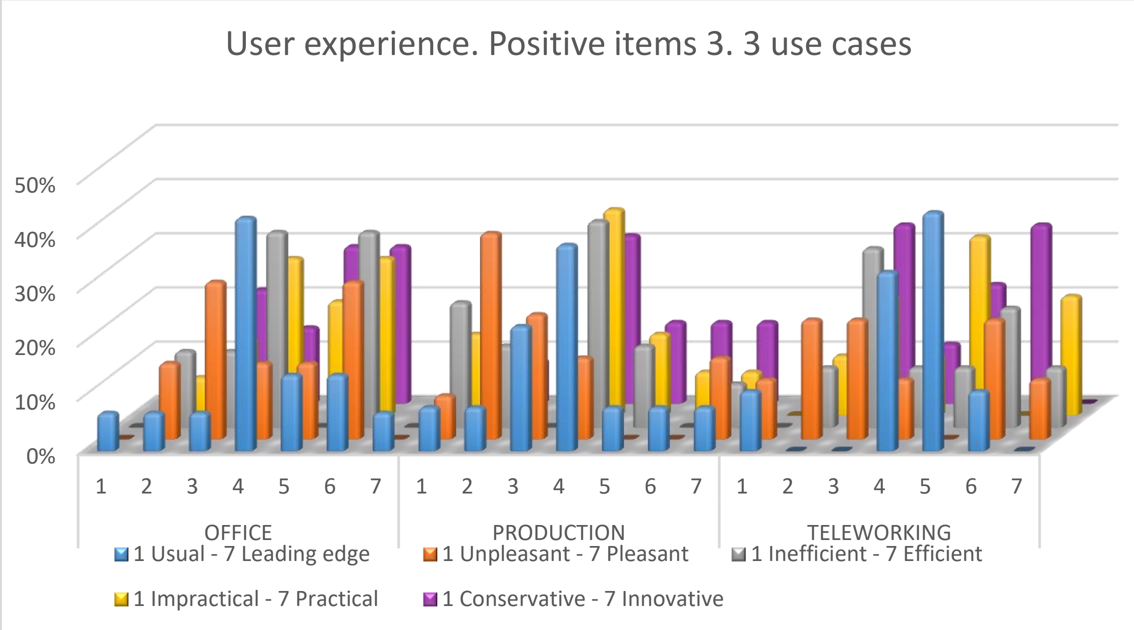


Figure 181: User experience questionnaire. Positive items 3. 3 use cases.

The office use case was the case that valued higher the following qualities of the WAOW tool: enjoyment, interest, pleasantness, and innovation.

The production use case valued higher the likeableness.

The teleworking use case valued higher: understandability, excitement, predictability, supportiveness, easiness, leading edge, efficiency, practicality.

Next there are another three graphs showing the negative items of the User experience questionnaire. In this case the lower score, means better assessment of the WAOW tool.

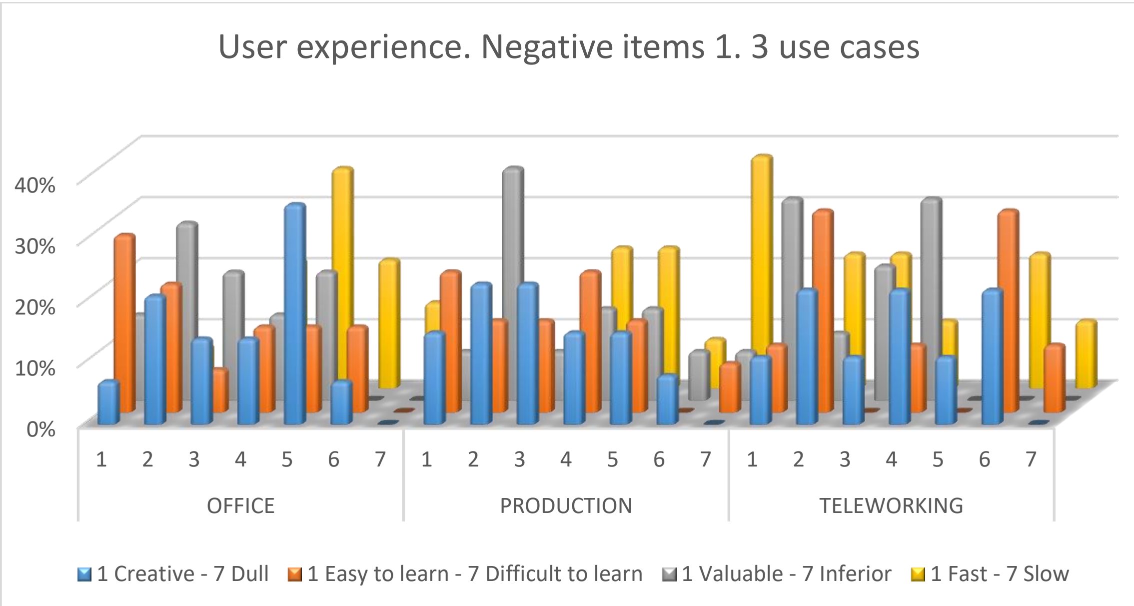


Figure 182: User experience questionnaire. Negative items 1. 3 use cases.

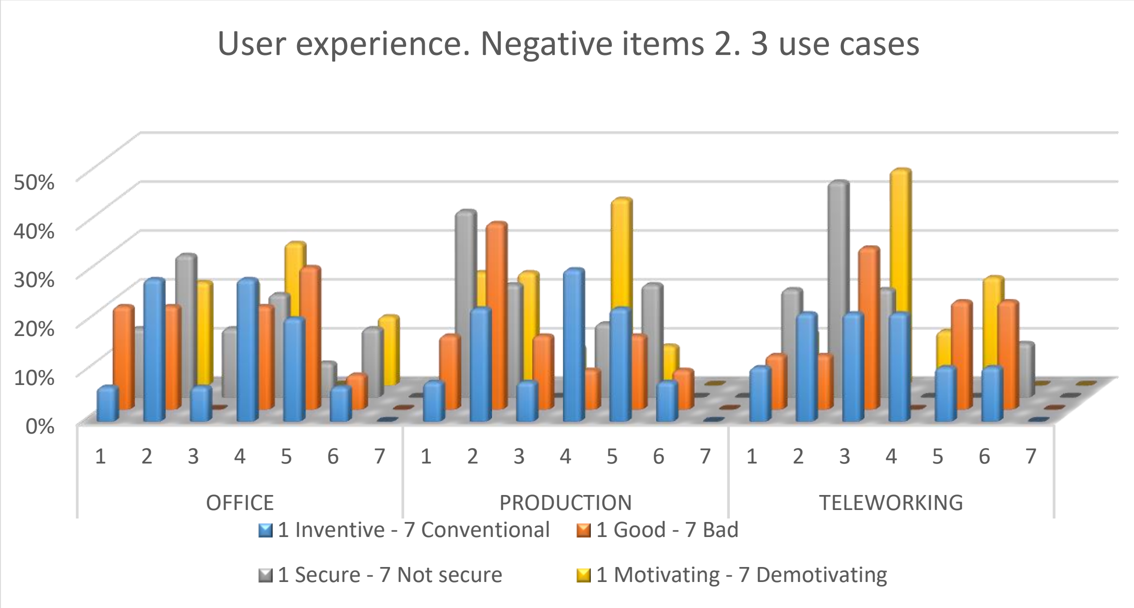


Figure 183: User experience questionnaire. Negative items 2. 3 use cases.

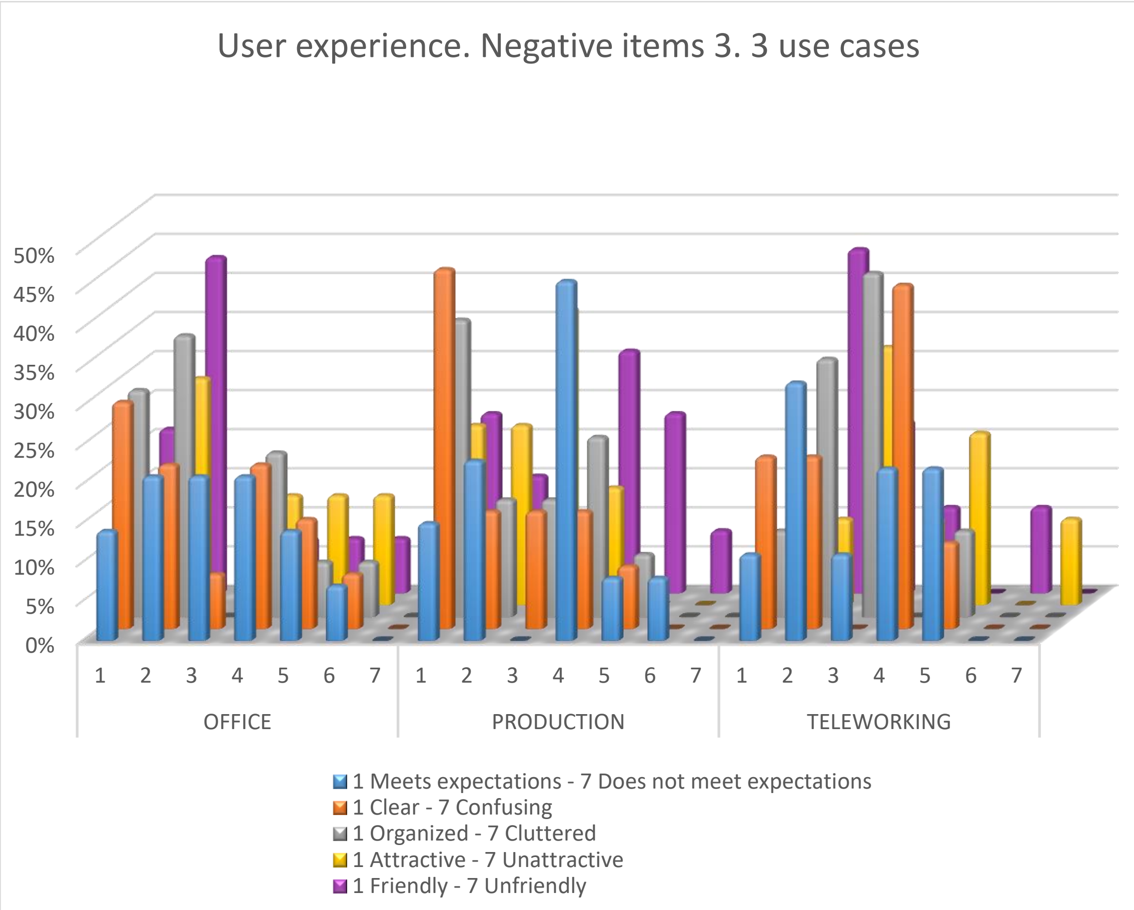


Figure 184: User experience questionnaire. Negative items 3. 3 use cases.

The office use case was the case that valued higher the following qualities of the WAOW tool: Easiness to learn, and meets expectations, The production use case valued higher: Creativeness, goodness, clearness, and attractiveness.

The teleworking use case scored higher: Value, speed, inventiveness, security, motivation, organization, and friendliness.

2.4 Tool Evaluation

2.4.1 Introduction

The WA project takes a user-centred approach based on continuous end-user involvement from design to validation. The application of human-centred development in the design of the WAOW tool enables economic and social benefits. From the very beginning of the development of the WAOW tool, relevant stakeholders were involved to best tailor the tool to the context of use. In order to continue this approach, end users were also involved to test the tool in terms of usability and acceptance and to work out possible approaches for improvement.

2.4.2 Method

The user-centred evaluation of the tool involves a two-step process during the implementation of the pilot tests. The first evaluation was carried out after the short-term tests and a second evaluation after the long-term tests in order to identify possible potential for improvement at an early stage and to be able to adapt it for long-term use.

2.4.2.1 Participants

The overall sample size of the first evaluation (after short-term tests) included N = 35 participants, in the second evaluation (after long-term tests) N = 28 user participated. In the first evaluation, a response rate of 97.22 % of all users was thus achieved; in the second survey 77.77% of all users responded. The questionnaire was sent to all WAOW tool users. Users were asked about their workplace and asked to choose from three predetermined options: Office setting, Production setting, and Work from Home/Teleworking setting. Similarly, users were asked if they worked full time or part time. Table 2 shows an overview of the sample characteristics for both short-term and long-term tests.

Table 2: Overview of sample characteristics for short-term and long-term tests

		1st Evaluation		***2nd Evaluation***	
		N	***%***	***N***	***%***
Gender	Female	22	64.1	16	57.1
	Male	13	35.9	12	42.9
	Other	0	0.0	0	0
Age	under 29	1	2.9	1	3.6
	30 - 44	6	17.1	4	14.3
	45 - 60	27	77.1	22	78.6
	Above 60	1	2.9	1	3.6
Workplace	Office	13	37.1	10	35.7
	Production	13	37.1	10	35.7
	Work from Home/Teleworking	9	25.7	8	28.6

Employment	Full Time	32	91.4	28	100
	Part Time	1	0.0	0	0
	Prefer not to say	1	2.9	0	0
	Other	1	2.9	0	0

2.4.2.2 Procedure

Data was collected twice via an online survey, once after the short-term tests in May 2022 and once after the long-term tests in June 2022. The online survey was accessible via a link which users received through an email, or through the WorkingAge App where it was integrated into the user interface. The questionnaire was available in English and Spanish, as all Greek users speak sufficiently good English.

2.4.2.3 Questionnaire

Users were briefed about their privacy information before being asked to answer questions about their demographic data. Afterwards, users were asked to rate how often they would like to use the WAOW tool at work as well as at home on a 5 Point-Likert-Scale (from *never* to *as much as possible*). If users indicated that they would like to 'never' use the tool, they were asked to explain why via an open response section. Usability was determined using statements from the SUS questionnaire where users were asked to answer statements on a scale ranging from 'strongly disagree' to 'strongly agree'. Afterwards participants were asked to rate components of the WAOW tool from 'not good at all' to 'very good'. User acceptance was determined using statements from TAM which users rated on a 5-point scale (from *strongly disagree* to *strongly agree*). Participants were also asked how they dealt with malfunctions and possible problems they may have had whilst using the tool. Users were then asked to evaluate the Interventions as well as the overall health benefits of the WAOW tool. Finally, users were asked to state the best aspect of the tool as well as suggest improvements via an open response section.

2.4.3 Results

2.4.3.1 Usability evaluation of the tool

The usability of the whole WAOW tool was surveyed twice, after short- and after long-term tests using the System Usability Scale (SUS). According to *A Practical Guide to the System Usability Scale* (Sauro, 2011), the SUS score was calculated. The usability of the WAOW tool in the short-term tests was assessed with SUS = 76.9 (*SD* = 11.09), while the usability after the long-term tests was assessed with SUS = 76.000 (*SD* = 5.463). Usability has thus improved by 15.2 points since the in-lab tests, from formerly 61.7 points in the lab tests to 76.9 in the pilot tests. According to the SUS rating scale, the WAOW tool can be rated C "*Acceptable*", with the grade between *Good* and *Excellent*. A t-test for dependent samples shows no statistically significant difference of the SUS score between male and female participants and furthermore no statistically significant difference between the three studied pilot sites. A descriptive evaluation can be found in the following Table 3.

Table 3: SUS Scores

		1st Evaluation		2nd Evaluation	
		M	**SD**	**M**	**SD**
Whole Sample		76.9	11.09	76.0	5.463
Gender	Female	75.0	8.729	76.2	6.655
	Male	80.5	9.288	75.8	3.571
Age Group	44 years and younger	79.6	7.185	79.2	6.611
	45 years and older	76.4	9.646	75.3	5.085
Workplace	Office	74.3	5.376	77.1	6.740
	Production	79.4	12.881	75.5	5.401
	Work from Home/Teleworking	77.6	6.803	75.3	4.062

2.4.3.2 Acceptability evaluation of the WAOW tool

The aim of our research was to build an acceptance model with regard to the WAOW tool usage that is based on already existing knowledge and takes context-specific factors of the interaction between human and health risk monitoring systems in an occupational as well as private setting into account. Therefore, the model was developed over four consecutive stages. First, a research model based on the literature was developed. This model took variables of the traditional technology acceptance models, such as TAM, TAM 2 and TAM 3 into account (Venkatesh et al., 2016; Venkatesh & Bala, 2008; Venkatesh & Davis, 2000). Further, the results of the in-lab tests, which were conducted within T9.2, were included. As such, the model contains context-specific factors that might be the subject for adaptation in work systems, such as perceived enjoyment, perceived safety, ethical, legal, and social implications and, on the other hand, personal characteristics such as self-efficacy, anxiety and perceptions of external control, which are considered as variables with uncertain influence on the predictors. The ELSI factors (ethical, legal, and social implications) were added based on the results achieved in T9.2 and WP7. ELSI deal, among other things, with the advantages and disadvantages of the technology. On the one hand, employees can be supported in physically demanding or monotonous tasks, on the other hand, the growth of technology can lead to job losses and a decline in human skills and knowledge. This dualism of technology leads to the need to consider ethical, legal and social implications in the development of new technology (Nelles et al., 2017). All dimensions and items derived in this way have been suitably adapted for the specific evaluation of the WAOW tool.

The following Table 4 provides an overview of the items surveyed, related to technology acceptability and ethical legal and social implications of the WAOW tool.

Table 4: Overview of items to assess acceptance

		Item	Response Option
TAM	Perceived Usefulness	1. Using the systems improves my healthy habits. 2. Using the system increases my productivity. 3. Using the system enhances my effectiveness. 4. I find the system to be useful for my health.	1 (do not agree at all) – 5 (fully agree)
	Perceived Ease of Use	1. My interaction with the system is clear and understandable 2. Interacting with the system does not require a lot of my mental effort. 3. I find the system to be easy to use. 4. I find it easy to get the system to do what I want it to do.	
	Use behaviour	I prefer to work with the system than without it.	
TAM 2	Job relevance	In my job, usage of the system is important.	
	Output quality	The quality of the output I get from the system is high.	
	Result demonstrability	I would have difficulty explaining why using the system may or may not be beneficial.	
TAM 3	Perceived enjoyment	I find using the system to be enjoyable.	
	Anxiety	The system makes me feel uncomfortable.	
	Perception of external control	I have the resources necessary to use the system.	
ELSI	Social Implications for work	I fear that I will lose the contact to my colleagues because of the system.	
	Social Implications for private life	I fear that I will lose the contact to my family and friends because of the system.	
	Legal implications (data protection)	I do not mind if the system records personal information about me.	
	Legal implications (occupational safety)	I do not mind to use the system at my workplace.	
	Legal implications (safety at home)	I do not mind to use the system at my home.	
	Perceived safety	I feel safe while using the system.	

After the first evaluation, users seemed to already accept the WAOW tool, as most TAM items were scored above three. Users scored the perception of external control (M = 3.371) and job relevance (M = 3.286) the highest. Anxiety (2.286) and output quality (M = 2.543) were scored the lowest.

Most users also seemed to agree that there were no ethical, legal, or social aspects that needed to be considered. Users indicated that they did not feel as if they would lose contact with their friends or family (M = 1.514) while they felt safe (M = 3.800) using the system.

The second evaluation makes it clear that the extended usage time leads to users rating the perceived control over the system higher, the discomfort towards the system decreases, and the enjoyment of using the system increases and furthermore ethical and social implications for work and private life are rated lower with the longer usage time. Significant differences between the measurement times were calculated, but no statistically significant differences are detectable.

The assessment of the estimated usefulness and the ease of use remain the same across the measurement points, the same applies for job relevance and quality of the output. Longer studies are necessary here to gain a more precise overview. It can be seen that the result demonstrability was rated worse in the second evaluation, this may be related to the fact that some changes could not be incorporated although the users asked for it and expected it to happen.

An overview of the items to assess user acceptance can be found in Table 5. Green fields indicate a positive trend, red fields indicate a negative trend, and orange fields indicate a constant rating (difference between first and second evaluation is < 0.1 evaluation points). The magnitude of the increase or decrease from the first to the second evaluation can be seen in the increase or decrease of the evaluation points. Depending on the dimension, an improvement can be an increase in the rating (e.g., perceived enjoyment) but also a lower rating (e.g., social implications for work should be as low as possible).

Table 5: Overview of items to assess acceptance

		1st Evaluation		***2nd Evaluation***		***Trend***
		M	***SD***	***M***	***SD***	
TAM	Perceived Usefulness	3.000	.818	2.993	1.040	-0.007
	Perceived Ease of Use	3.264	.895	3.268	0.850	0.004
	Use behaviour	2.886	.963	2.850	1.110	-0.036
TAM 2	Job relevance	3.286	1.126	3.190	0.983	-0.096
	Output quality	2.543	1.038	2.571	1.069	0.028
	Result demonstrability	2.857	1.309	2.571	1.069	-0.286
TAM 3	Perceived enjoyment	3.229	1.031	3.214	1.100	-0.015
	Anxiety	2.286	1.250	2.107	1.065	-0.179
	Perception of external control	3.371	1.285	3.679	1.188	0.308
ELSI	Social Implications for work	1.600	.976	1.357	0.826	-0.243
	Social Implications for private life	1.514	.950	1.143	0.356	-0.371
	Legal implications (data protection)	3.629	1.416	3.857	1.044	0.228

Legal implications (occupational safety)	3.771	1.262	4.000	0.981	0.229
Perceived safety	3.800	.964	3.857	1.208	0.057

2.4.3.3 Evaluation of the tool's individual functions and sensors

Users were asked to rate how often they would like to use specific sensors as well as the app from the WAOW tool whilst at work and at home. Users were asked to answer these questions once after the short-term tests and once after the long-term tests. Sensor usage at work was rated very highly with almost all users stating that they would like to use the sensors and functions of the tool 'as much as possible'. For the usage at home, users preferred the activity tracker, the scale as well as the questionnaires and other features of the app, whilst sensors such as the environmental sensor and the noiseBox were rated lower for at home use. The results can be seen in Figure 185 and Figure 186.

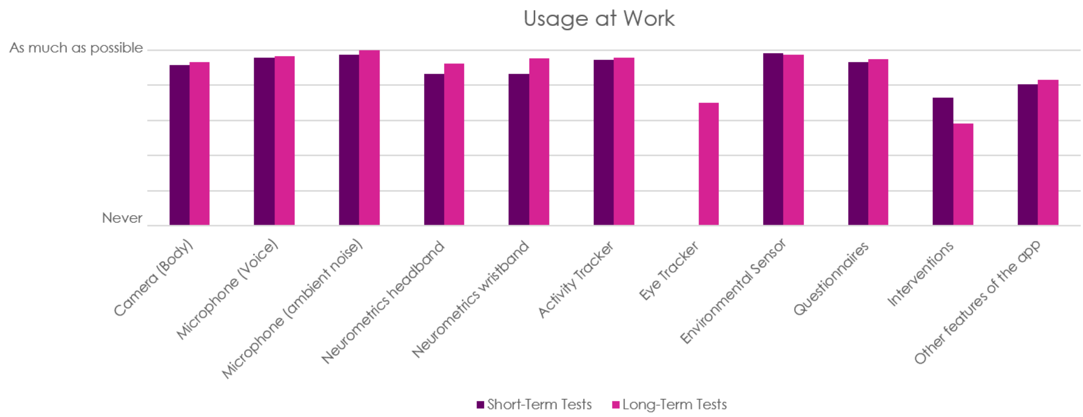


Figure 185: Results of average user responses for sensor usage at work in short- and long-term tests (some sensors were only used in the long-term tests and are therefore not within short-term test usage).

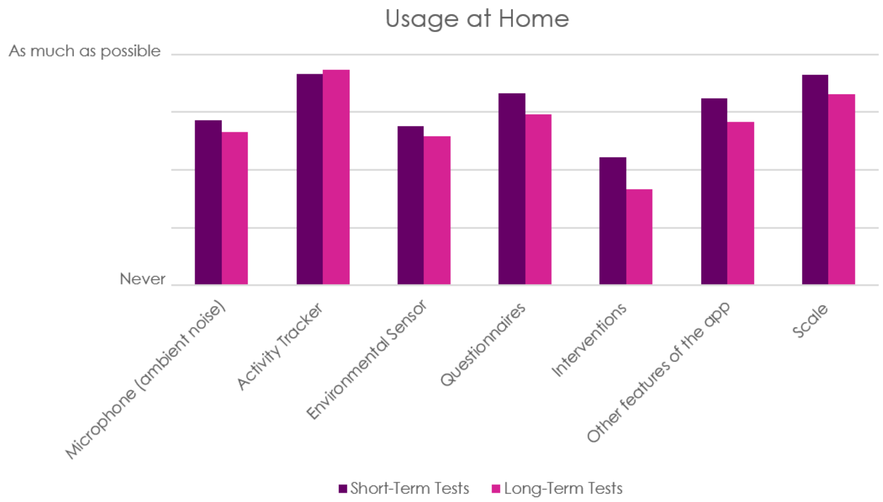


Figure 186: Results of average user responses for sensor usage at home in short- and long-term tests.

To find out why users did not want to use specific sensors, they were asked to indicate their reasoning via an open-answer question. Users indicated that they either didn't know that they could use specific sensors, that the sensor didn't work properly or that they had privacy concerns.

After the short-term tests, users were asked to state what they thought was the best aspect of the WAOW tool, via an open-response question. Figure 187 shows user responses to that question. Most users mentioned that the Health Benefits (64%) were the best aspect of the WAOW tool. The second most mentioned aspect were the sensors which 20% of users answered as being the best aspect of the Tool.

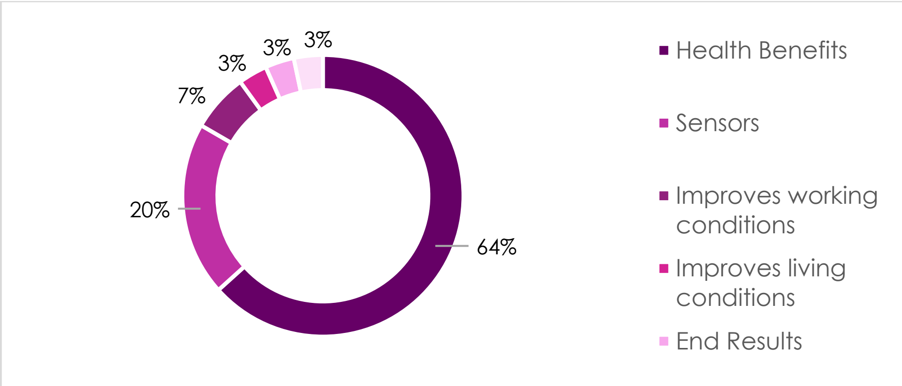


Figure 187: Open response question of best aspect of WAOW tool after Short Term Tests

Users were also asked to state what they thought was the best aspect of the WAOW tool after the long-term tests. This was done in order to see changes in user behaviour after using the WAOW tool for a longer period of time. After the long-term tests, more users stated that they thought the Sensors were the best aspect of the Tool (50%) followed by the health benefits of the Tool (32%). After the short-term tests, users stated that the collected data was one of the best aspects of the tool, however, after the long-term tests this was no longer mentioned by users. The results can be seen in Figure 188.

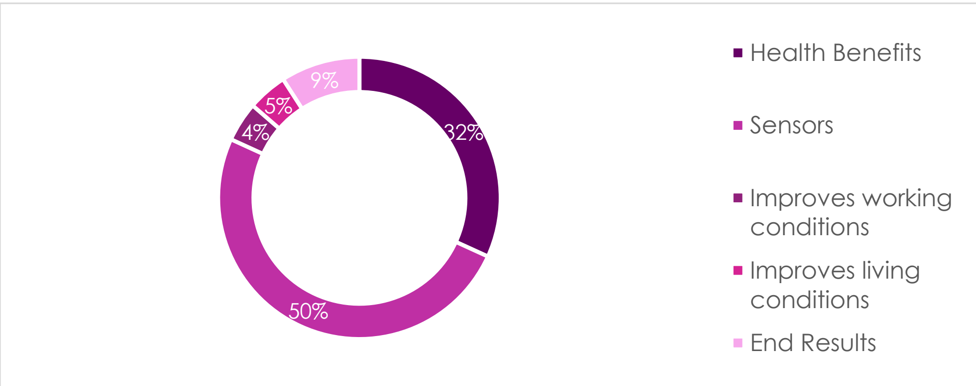


Figure 188: Open response question of best aspect of the WAOW tool after Long Term Tests (LTT)

The questionnaire also asked users to rate predetermined positive effects that the WAOW tool could possibly aid with, once after the short-term tests and once after the long-term tests. After the short-term tests, users indicated that they believed the WAOW tool would have the biggest positive influence on physical state of health (M = 3.2). However, users also indicated that they did not believe that the WAOW tool could have a large positive effect on the handling of physically demanding tasks (M = 2.1). However, after the long-term tests, users rated the predetermined positive effects much higher than they had previously done. Positive effects such as the overall well-being at work (M = 3.1) and the handling of physical (M = 2.3) and cognitive (M = 2.4) demanding tasks received an increase in user support after the long-term tests. This shows that users see the usefulness of the WAOW tool after using it for a longer period of time. The results can be seen in Figure 189.

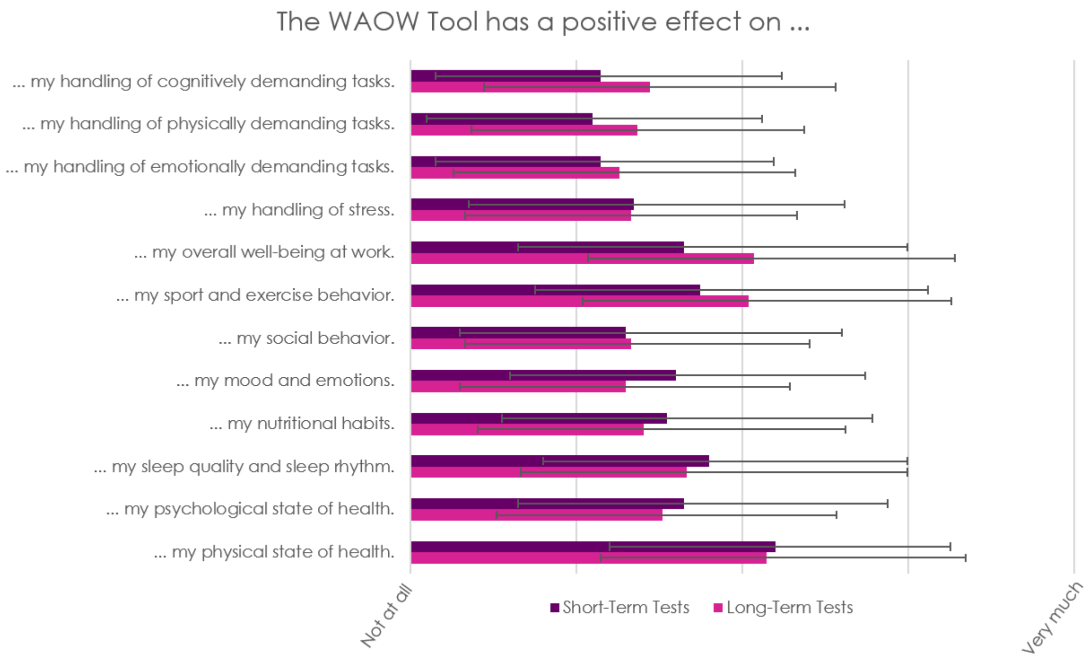


Figure 189: Average user evaluation of possible positive effects caused by the WAOW tool

2.4.3.4 Evaluation of Interventions

In a first step, users were asked about the interventions in general. We wanted to know what they thought of the interventions suggested by the WAOW tool in the case of an identified health risk, e.g. when high stress is identified, and further we wanted to know how they rated interventions that offer general tips and support and thus serve as prevention. Both questions were asked after the short-term tests and long-term tests. Despite the fact that the long-term tests were very short due to the Corona pandemic and thus the interventions were probably not able to bring about actual health improvements, there is a positive trend in the evaluation (Figure 190).

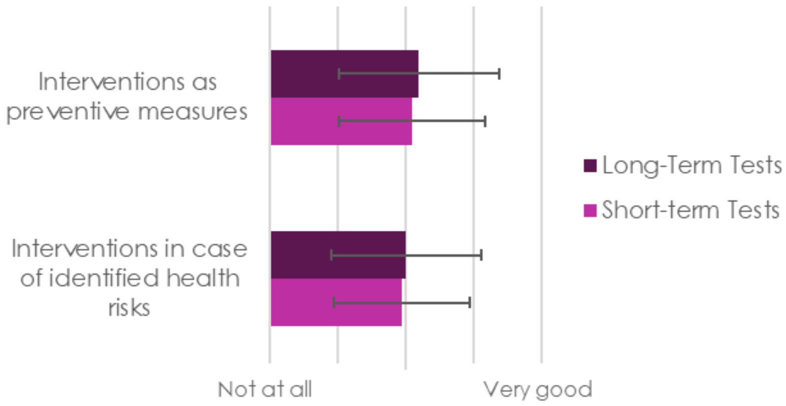


Figure 190: Best aspect of interventions after short-term tests according to users

The questionnaire asked users to evaluate the interventions via an open response question. The questionnaire asked participants to state what they thought was the best aspect of the interventions, as well as if they had any ideas for improvements.

Figure 191 show user responses on what they believe is the best aspect of the WAOW tool interventions. Most users agreed that the accuracy of the sensors (43%) was the best aspect, followed by the reminders of the tool (24%).

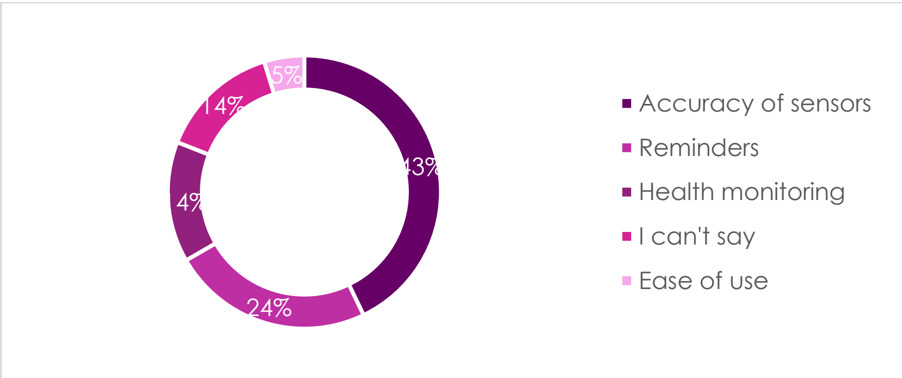


Figure 191: Best aspect of interventions after long-term tests according to users

Furthermore, users were asked to indicate any suggestions for improvement. Figure 192 shows that users wanted interventions for the WAOW tool.

After the long-term tests, more than half the users agreed that they were content with the interventions (54%). However, a few suggestions stated that they would like more personalized interventions (14%), a higher frequency of interventions (14%), as well as more relevant interventions (11%) and more user-friendly interventions (4%).

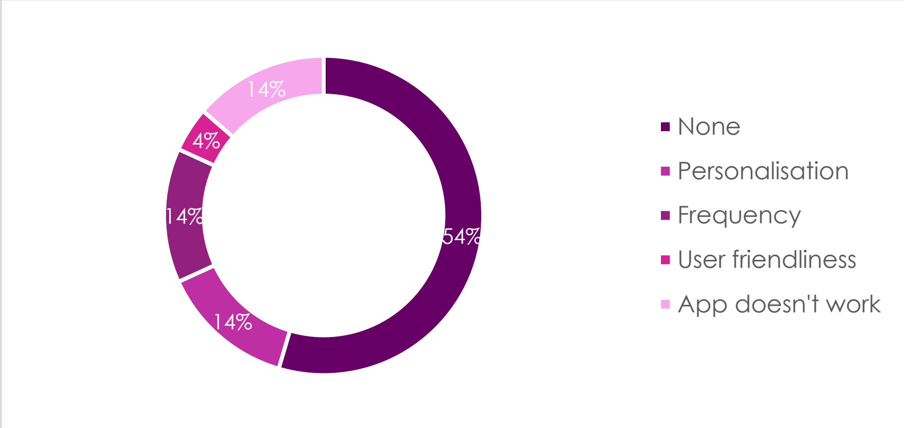


Figure 192: Suggested improvements for interventions after long-term tests

2.4.3.5 Issues with the WAOW tool

In order to determine possible issues with the WAOW tool after the short-term tests, users were asked to state problems they had whilst using the tool as well as how they dealt with these problems. Figure 193 shows user responses to problems they had whilst using the Tool. Users seemed to agree that the App wasn't working quiet properly (33%) as well as having connection problems (22%).

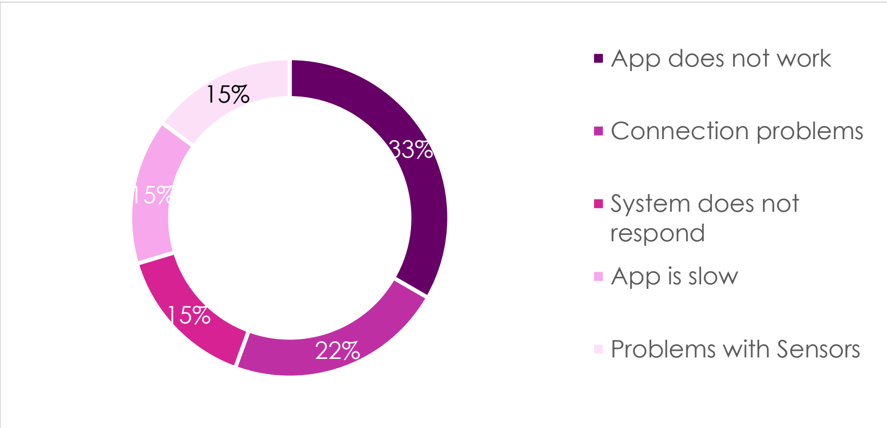


Figure 193: Open response question of problems users had whilst using the Tool after Short Term Tests

Users were then asked how they dealt with those problems and given three choices on how to obtain help. Figure 194 shows user responses to that question. After the short-term tests, most users indicated that if they had problems, they asked friends/colleagues for help (M = 3.8). However, after the long-term tests' users seemed to have fewer problems with the WAOW tool and were split between needing help from the developers/researchers (M = 2.5) and colleagues/friends (M = 2.5). In all cases, the less help was taken up, a sign that a learning effect has set in, and the users have learned to deal with the tool.

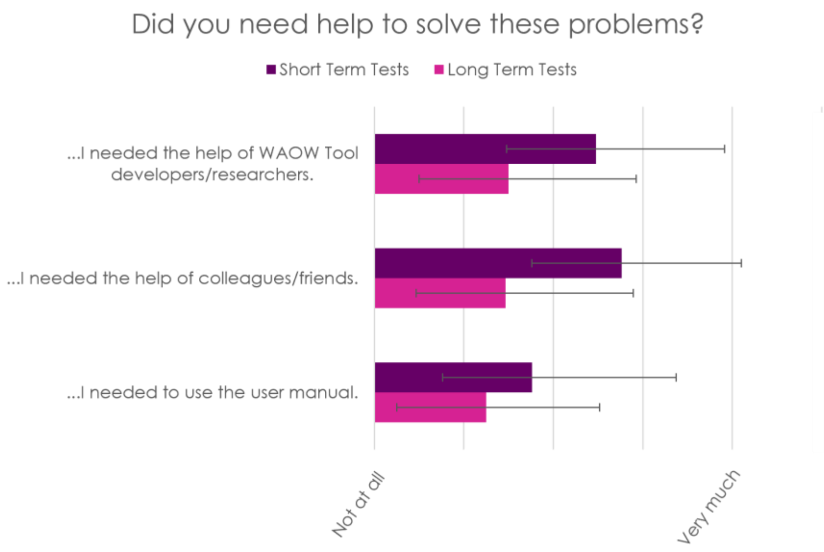


Figure 194: User responses on how they dealt with problems whilst using the WAOW tool

2.4.3.6 *Further Development*

In order to identify aspects that should be further developed in the future, users were asked to indicate which aspects they would most like to see developed. Most users mentioned that they would want the App to be improved (36%), which is no surprise since users mentioned they had the most problems whilst using the App. The second most wanted improvement were the Interventions with 18% of users mentioning this. Figure 195 shows what improvements users wanted for the Tool.

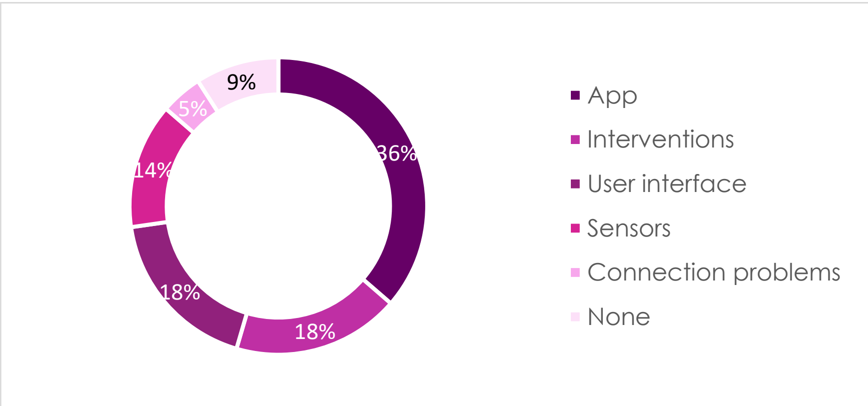


Figure 195: Open response question of user wanted improvements for the WAOW tool after Short Term Tests

2.4.4 Discussion

The aim of this part of the evaluation protocol of the WAOW tool was to be able to assess usability and acceptance in the context of short- and long-term tests with real users in the real working environment. Building on the in-lab tests, similar factors were recorded in order to evaluate the development over time and to be able to react adaptively to results and make possible adjustments within the framework of a user-centred development. The evaluation during the pilot test was carried out in two phases, on the one hand during the short-term test and further during the long-term test. Despite the fact that the long-term tests could not be carried out as planned (especially not as long as planned) due to various challenges (e.g. pandemic, semiconductor crisis), there is a positive trend in the evaluation with regard to usability and acceptance.

First of all, it is to be mentioned positively how much the assessment of usability has improved from laboratory tests to field tests. Usability has improved by 15.2 SUS-points since the in-lab tests, from formerly 61.7 points in the lab tests to 76.9 in the pilot tests. According to the SUS rating scale, the WAOW tool can be rated C "*Acceptable*", with the grade between *Good* and *Excellent*. This development is very positive and shows that the in-lab tests and associated user studies were of great value and that the feedback gained on user satisfaction could be incorporated into the WAOW tool and thus brought about this significant improvement within the in-company tests. In contrast, there is no major change between short- and long-term tests, but this can be explained by the rather short period of use, in which only some of the desired changes could actually be implemented.

Another positive aspect is that there are no statistically significant differences between the groups of interest examined, neither with regard to gender differences in the evaluation, nor in age-related evaluations, and furthermore nor in the evaluation of the individual use cases. This shows that the adaptive solution of the WAOW tool to cover all the use cases analysed, as well as the target user group evaluated, is working.

With regard to acceptance and influencing factors, it is particularly evident that social and legal implications play a much less important role than assumed after the in-lab tests. Neither social implications nor concerns about data protection and data security are perceived as seriously as they appeared to be in the in-lab tests. The measures developed based on the in-lab test data, which focused in particular on a clear communication and information flow with introductory sessions, related material and explanations, thus seem to have been received and accepted by the users. In addition, all these factors have improved over the period of use from short- to long-term testing. This means that initial doubts and worries can be reduced through adequate communication and support, and in addition that even a short use of the tool makes users feel more comfortable and confident in using it.

The evaluation of the individual components of the tool shows a high level of agreement at both survey times; the users would like to use all sensors and functions almost all the time. This evaluation again shows a very positive picture and the acceptance of the users towards the system and the functions. Even though the eye tracker is rated lower in this evaluation, acceptance is still very high. The reasons for the lower rating are discussed in chapter 6, summarised the users see great potential in the functions of the eye tracker, but would like a contactless and wireless device for better flexibility, a remote eye tracker could be a practical solution here.

The evaluation also deals with problems with the tool, here the learnability of the tool should be checked. In fact, after the long-term tests, it became apparent that the help offers had to be accepted much less frequently and that the users could help themselves, which speaks for a positive statement regarding the learnability.

Future work related to the WAOW tool should focus on the WAOW app - the interface of the tool and thus a core element. Unfortunately, some problems were noted by users, as some of them noticed app crashes or connection problems. Despite this, a positive trend can be seen in the evaluation by users. Although the in-company tests turned out to be much shorter than expected, even the short period of use led to improvements in terms of usability and acceptance.

2.5 WAOW tool measurements: Gesture Recognition

2.5.1 Introduction

The Facial Recognition and Authentication and Gesture Based Interaction component is a communication component included in the WAOW tool to facilitate the user's interaction with the system. It provides an alternative pathway for the user to start and stop the sensors from running, instead of doing so through the settings menu of the app, they can perform an appropriate gesture. The gestures are performed through finger enumeration, 1 finger is encoded to mean "stop/pause sensors" and 2 fingers mean "start/resume sensors". Only authorized users are allowed to interact with gestures, which is achieved by first performing a facial recognition and authentication step. As described in the evaluation protocol (D9.1), an evaluation of the short-term tests takes place, a further evaluation of the long-term tests is not necessary, as this submodule is not intended to trigger any health- or social-related effects.

2.5.2 Method

To validate the facial recognition and authentication and gesture interaction component, the following KPIs are defined:

1. **Setup**: the camera is pre-configured to connect to the WorkingAge network. When powered on, its status light will blink red and when connection is established it will switch to solid green.
2. **KPI1**: Add User functionality
    a. Upon first use, the user takes a picture of themselves to create their profile, as prompted by the WAOW app
    b. The edge-cloud server acknowledges new user
3. **KPI4**: Edge cloud sends high level data to mobile app
    a. Upon user performing the appropriate gesture, the messages regarding performed gesture are sent to the Mobile App.

2.5.3 Results – Short Term Tests

1) Test procedure

The goal of the single-day tests was to confirm the correct setup of the camera and test the communication functionality between the Facial/Gesture Recognition component and the Mobile App via ZeroMQ system messaging service (e.g. the publish and subscribe services) in order to ensure that the registration procedure has been successfully completed (KPI1). Also, we had to ensure that the gestures performed through each user have been properly sent to the Mobile App (KPI4).

This registration procedure is only required when a user is added for the first time to the WAOW tool. Whenever a new user registers to the WAOW tool, the App will publish a message, via ZeroMQ, with topic: "adduser/photo"; Payload: UserPseudoID, SensorGroupID, RSA 4096 public key, Base64-encoded, encrypted photo of the worker. An example of JSON payload for this message is:


```
{
     "userpseudoid": " Ubbd328ec_7e27_4526_a814_2189aec608b6 ",
     "sensorgroupid": " S8771a305-ba89-4eb1-b6e4-d9983a565f10",
     "rsa4096publickey": "ssh-rsa
AAAAB3NzaC1yc2…GgtShbs9649r/Loufhl…"
     "photo":"/9j/4AAQSkZJRgABAgAAAQABAAD…pHje8vrprm9Hk2lv5sD\\n7llVI
"
}
```


When the user registration message is received, the Facial/Gesture Recognition component answers with a message to the App, to validate the registration, with topic: the UserPseudoID the message is directed to, concatenated with "addeduser" and separated by the "/" character; for example: Ubbd328ec-7e27-4526-a814-2189aec608b6 /addeduser; and payload a JSON message composed by sender: hostname or IP of the device that acknowledged the registration. An example of JSON payload for this message is:


```
{
     "sender": "computer1.workingage.eu"
}
```


Below we present the logfiles from the Mobile App as they were generated during the user registration procedure:


```
22/02/2022 14:04:08 Sent: [proxy:zeromqproxy.workingage.eu,
adduser/Sa1ca1f9d-13fa-448c-80fd-f1e77373bf7f
{"rsa4096publickey":"ssh-rsa … GLc\u003d generated-
key","sensorgroupid":"Sa1ca1f9d-13fa-448c-80fd-f1e77373bf7f",
"userpseudoid":"U8aaef6e4_b81f_4662_8cee_9d10a8962605"}]
22/02/2022 14:04:08 Sent: [adduser/photo]
```

```
22/02/2022 14:04:10 Received: [proxy:zeromqproxy.workingage.eu,
U8aaef6e4_b81f_4662_8cee_9d10a8962605/addeduser{"sender":
"computer1.workingage.eu"}]
22/02/2022 14:04:10 Received decrypted:
[U8aaef6e4_b81f_4662_8cee_9d10a8962605/addeduser{"sender":
"computer1.workingage.eu"}]
22/02/2022 14:04:10 Registered in sensor: computer1.workingage.eu
```

Whenever a gesture is recognized, the Facial/Gesture Recognition component publishes a message, via ZeroMQ, to a specific topic (topic is the UserPseudoID the message is directed to, concatenated with "send.gesture" and separated by the "/" character; for example: U550e8400-e29b-41d4-a716-446655440000/send.gesture) and payload a JSON containing the recognized gesture action as a number (If "gesture" == 1 the user requests to stop recording. The App will publish a stop message; If "gesture" == 2, the user request for start recording. The App will publish a start message.) An example of JSON payload for this message is:

```
{
    "gesture": 1
}
```

The message will be received by the Mobile App subscribed to the UserPseudoID specified into the topic, as a regular message sent by the Facial/Gesture component, including the payload. Below you can see the logfiles as they generated from the Mobile App during some successfully performed gestures by a registered user:

```
28/02/2022 06:59:30   Received: [proxy:zeromqproxy.workingage.eu,
Ub2bbabc9-a128-471d-ab71-bb1cf242f099/send.gesture{"gesture": 1}]
28/02/2022 06:59:30   Received decrypted: [Ub2bbabc9-a128-471d-ab71-
bb1cf242f099/send.gesture{"gesture": 1}]
28/02/2021 06:59:31   Received: [proxy:zeromqproxy.workingage.eu,
Ub2bbabc9-a128-471d-ab71-bb1cf242f099/send.gesture{"gesture": 2}]
28/02/2022 06:59:31   Received decrypted: [Ub2bbabc9-a128-471d-ab71-
bb1cf242f099/send.gesture{"gesture": 2}]
28/02/2022 06:59:32   Received: [proxy:zeromqproxy.workingage.eu,
Ub2bbabc9-a128-471d-ab71-bb1cf242f099/send.gesture{"gesture": 1}]
28/02/2022 06:59:32   Received decrypted: [Ub2bbabc9-a128-471d-ab71-
bb1cf242f099/send.gesture{"gesture": 1}]
```

2) Expected results

The user should be correctly identified among all users present in the database. The performed gestures should be correctly identified by the system, as well.

The subsequent discussion summarizes the performance of the Facial/Gesture component, in terms of precision/recall/ f1-scores scores, which can help measure the accuracy of the system. Finally, the min/max/mean latency values for each gesture performed are computed, as they generated from each user during the teleworker pilot.

First, we compute the confusion matrix in order to evaluate the accuracy of our system. By definition a confusion matrix C is such that $C_{A,B}$ is equal to the number of observations known to be in group ‘A’ and predicted to be in group ‘B’. In the case of the face/gesture component there are two classes, 1: Stop (Negative) and 2: Start (Positive). So, in our binary classification we have 4 different cases:

- True Negative (TN): when a case was negative and predicted negative
- True Positive (TP): when a case was positive and predicted positive
- False Negative (FN): when a case was positive but predicted negative
- False Positive (FP): when a case was negative but predicted positive

Also, we can define $C_{1,1}$ as the count of TN, $C_{1,2}$ as the count of FN, $C_{2,1}$ as the count of TP and $C_{2,2}$ as the count of FP. The results are summarized in Figure 196, showing $C_{1,1}$: 8 cases; $C_{1,2}$: 1 cases; $C_{2,2}$: 11 cases; $C_{2,1}$: 0 cases.

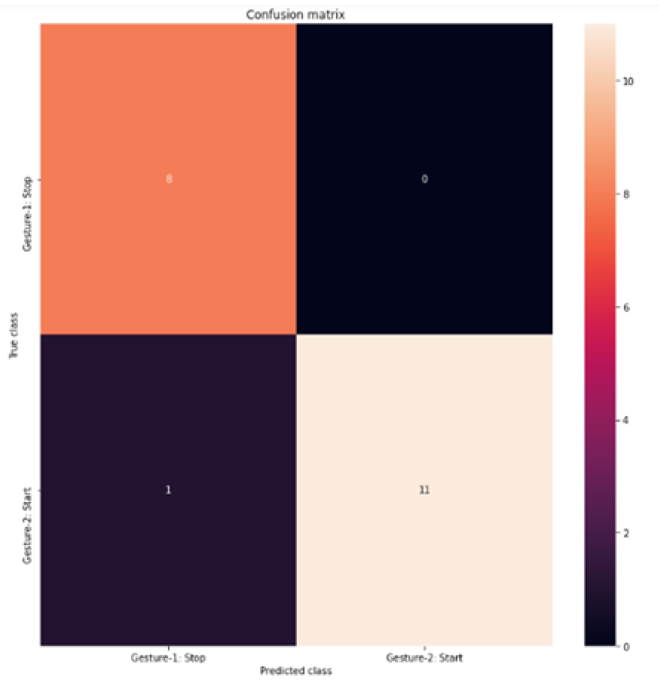


Figure 196: Confusion Matrix of the predicted "start"/"stop" cases

Additionally, we report some results to estimate the quality of the predictions from the Facial/Gesture Recognition component. We use the precision, recall and F1 Score, which are used in machine learning as performance evaluation metrics, to evaluate the accuracy of our system. These are defined as:

- Precision (% of your predictions were correct) = TP/(TP + FP)
- F1 Score (% of positive predictions were correct) = 2*(Recall * Precision) /(Recall + Precision)
- Recall (% of the positive cases that were correctly identified) = TP/(TP+FN)
- Latency = Time duration (End time – start time) during the gesture process

Table 6: Performance metrics for face/gesture component

Gestures	Precision	Recall	F1-score	total	Measurements
1	0.89	1.00	0.94	8	***Latency(ms)*** Min: 135.088000 Max: 615.985000 Mean:348.505667
2	1.00	0.92	0.96	12	**Accuracy:** 95%

The performance metrics for the face/gesture component are reported in Table 6, with the component succeeding quite well in all metrics.

2.5.4 Discussion and Outlook

In general, with computer vision systems, there are several parameters, such as distance, light exposure, angle of view, etc, that can affect the accuracy of the system. One of the most significant tasks in computer vision is to check these parameters and see the system behaviour for different values. The camera field of view should have constant lighting conditions and be well-illuminated. Direct sunlight exposure or high contrast between user and background should be avoided.

For the volunteers in the pilot study, the cameras were placed and calibrated to fulfil most of these requirements, and additionally some time and distance thresholds were set in order to achieve as noiseless background as possible. However, there is still room for improvements specifically in cases where there is a lot 'environmental' noise which affects significant to the accuracy of the whole system.

Finally, the users were successfully trained to correctly perform the gestures in order to achieve the best possible results. The component was reported as intuitive and easy to use.

2.6 WAOW tool measurements: Posture Recognition

2.6.1 Introduction

A pose estimation module has been developed by ITCL providing useful information for monitoring ergonomic habits. In order to provide an unobtrusive platform, the module consists of a single RGB camera located at a suitable viewpoint i.e. 2D images are used for pseudo-3D pose estimation. Objective risk assessments based on RULA and REBA have been implemented and interpreted to provide ergonomic information to the Decision Support System. To this end, the positions of the joints of the worker/user are estimated using different deep learning subtasks.

The Body Posture sensor provides the app with high-level information about the ergonomic risk of the user. This information is divided into 4 categories: None, if no user is detected, Low, Mid and High. These values have been validated by an expert in the In-Lab tests.

2.6.2 Method

Deliverable D9.1 Evaluation protocol establishes a series of metrics for the testing of reliability of components, including the Body Posture sensor. Since no additional validation or calibration processes are applicable to this sensor, the following test method was applied to the short-term tests. A description of participant can be found at the beginning of the document for each use case.

2.6.3 Results – Short Term Tests

Once the in-lab tests had been completed and all the modules were integrated in the WA tool, the short-term tests were carried out. These tests have been helpful in validating the integration of the system in a possible real environment and being able to make modifications for the long-term tests.

The connection between the cameras and the edge cloud was made through a raspberry device, which had to be configured so that the transmission of information was carried out securely. Each of the cameras was configured by ITCL and validated individually prior to the start of the short-term tests.

The next aspect to validate in the short-term test was the correct process of user addition. In all three pilots, this process was monitored to ensure that it was carried out correctly. Although, due to network problems, some users had to repeat the process, but in the end all users were correctly registered in the Body Pose Edge cloud.

Once the users had been registered, the correct positioning of the cameras had to be checked. In the case of office and teleworking, it was possible to set up a good location for the cameras. In the production environment, it was somewhat more complex, due to the limitations of the environment and security protocols.

The D9.1 Evaluation Protocol document states that the cameras should be mounted on the left side of the worker, at about 2 meters distance and a height of 1.5 meters. In the short-term tests, due to space limitations of the environments, this was not possible for a number of workplaces, which had effect on the results. In the teleworking case, many users did not have enough space on the left side and had to put it on the right. In the office setting, something similar happened; the cameras were re-placed in order to avoid them being in the middle of corridors or capturing other employees not participating in the pilots. It meant that in some places the camera placement was not perfect, but it could work. In the production use case, the cameras had to be fixed to prevent movement and falls, and thus comply with safety protocols. This caused a major impediment to the camera positioning requirements, producing occlusions or that the whole body of the user was not visible in the images captured by the camera.

Despite the impediments, each workplace of the three locations (pilots) (office, teleworking and production) were analyzed in order to establish the best possible location for the cameras.

As mentioned in the previous paragraphs, to validate the correct functioning of the system, the protocol described in D9.1 Evaluation Protocol has been followed. Thus, for each of the users, the following points were checked:

- Camera and raspberry configuration and connection.
- Edge cloud up and running.
- User registered at edge cloud.
- User sends start to the edge cloud.
- Check at app logs or edge cloud logs that high level info regarding ergonomic assessment were generated.

The last point is the most difficult to verify. As commented before there are cases, especially in the production environment, where the camera did not have a good viewpoint and were not able to correctly estimate the user's position. This problem was increased by network problems. With high network traffic load and low coverage points, the speed of image transmission was affected. Despite this, body risk analysis had been carried out in most cases, but with a much lower frequency than initially estimated.

2.6.4 Results – Long Term Tests

Once the short-term test had been completed, the system was validated and installed in the different pilot locations. The cameras were working properly and the communication with the server was working as well, so the system was ready for the long-term tests.

That was the starting point, but there were different problems throughout the tests. Firstly, due to an incident at EXUS pilot, the long-term tests were not carried out with the Body Pose component. In addition, the information collected during the short-term tests was lost.

As for the production pilot (Grupo Antolin), this was the most challenging one. There were no suitable vantage points for the cameras due to limitations of the hosted infrastructure in the environments which could interfere with the workers' work tasks or jeopardize job security, so the positions used did not meet the requirements for which the system was designed. This resulted in a lower frequency in the output of the system. In addition to this, workers do not have a fixed workplace, so they can only use the system when they are in their original position. This varied from day to day and meant that a reduced number of data was collected by the system, especially when this situation happened frequently. In addition, a great number of the production workers had problems using the application; some stopped the sensors unintentionally, causing start messages not to be sent to the sensors (Figure 197).

Figure 197: Camera moved by user (left) and camera correctly positioned (right)

Despite all these problems, the Body Pose system worked properly in these trial tests with a minimum number of correct conditions, as can be seen below, launching different interventions. There are large differences between the numbers of interventions per user, as in many positions the camera viewpoint was not adequate. But when the point of view was correct, the system worked correctly (Figure 198).

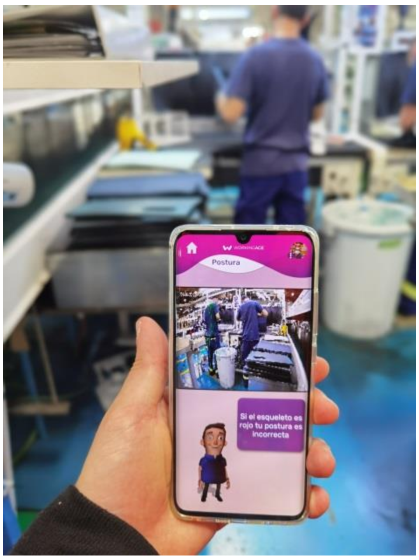


Figure 198: Ergonomic risk analysis of the worker in manufacturing environment (screenshot of the app)

Regarding the office pilot (MUTUA), it was possible to make a much better installation with a better viewpoint for the cameras. Thanks to MUTUA's collaboration, these positions could be adjusted and reviewed on a frequent basis. There were some problems in the registration of users, which led to a time when no information was generated. In addition, due to lack of network coverage, connection to the camera was occasionally lost. Once this was solved, a lot of information could be collected. As can be seen, the results were much better than in the production pilot (Figure 199).

Figure 199: Ergonomic risk analysis of the working in office environment. (Screenshots of the app)

In Table 7 you can see the number of interventions by each of the pilots. As mentioned above, there is no data available of the EXUS pilot, due to the incident that occurred. The data was sufficient, although we will see it detailed by user later on.

Table 7: Number of interventions by each of the pilots for posture recognition system

Pilot	Nº of interventions
Office (Mutua)	29958
Factory (Antolin)	11615

Below is the number of interventions per user, with 13 users at office pilot (MUTUA) and 10 at production pilot (Antolin). In the office case, there were actually 13 users, so we can see that data was collected from all of them. On the other hand, in the production case, we had 13 users, but information was only collected from 10. This is due to the problems mentioned below (Table 8).

Table 8: Number of interventions per user based on posture recognition sensor

User_id	Nº of interventions
Office (MUTUA)	
U041bed1d_9a41_45e1_84ca_[…]	889
U0b93951f_c781_4597_ba16_[…]	2
U2ee3f66a_2b57_43c7_9311_[…]	2
U346cdb20_b565_494c_a4d3_[…]	8792
U5f27b8f9_9646_415c_b17f_[…]	3918
U63c32158_23f0_476f_b7bf_[…]	2433
U85070592_37e9_41db_bab2_[…]	2364
Ua23382f9_554b_4c4f_a205_[…]	6134
Uc364fb98_4a54_47dd_84a8_[…]	2
Ud8c6722c_0743_41e9_908c_[…]	55
Ued6bdce0_6d09_4938_9ac0_[…]	504
Uf03ff3a1_e790_44b6_98bf_[…]	57
Uf1b92dc5_33cd_4a93_8e2e_[…]	4806
Factory (Grupo Antolin)	
U13d98d27_5092_40e7_b20d_[…]	1898
U178d8826_1208_4df7_aaee_[…]	300
U2c484e7f_e1ad_43b5_b102_[…]	7733
U48b886b5_61c6_4247_a77b_[…]	100
U4b5e16fc_3173_4635_9fe1_[…]	86
U51303b0e_b5bf_4e07_bb87_[…]	45
U6af6309f_98eb_48e4_b867_[…]	1272
U87bfea76_2591_4a4a_8da9_[…]	24
U8a3e7473_57a2_4bcf_83f4_[…]	156
Uf8606357_fdaf_4989_95dc_[…]	1

As mentioned above, there are quite a few differences from one user to another, this is due to the conditions of the workplace. In some cases, the visibility of the user is poor (part of the body is occluded). It may also be due to problems with the app, or the network. This means that the server that captured the images did not receive the order to start capturing the images.

However, we can conclude that there are a large number of users who did not receive much information, those with less than 10 interventions, for example. The problem may have been due to the use of the mobile phone, or the camera may have been displaced or occluded, or any of those problems mentioned in the previous paragraphs.

Below you can see a series of graphs for each of the pilots, in order to draw conclusions from the long-term test data. Firstly, you can see the number of interventions per hour. Figure 200 shows the number of interventions throughout the day. In this graph, we see on the top the number of interventions and on the bottom the level of risk. A higher risk value means a worse position of the user and therefore a higher ergonomic risk (x-axis describes the time of the day, the y-axis the number of interventions at this time).

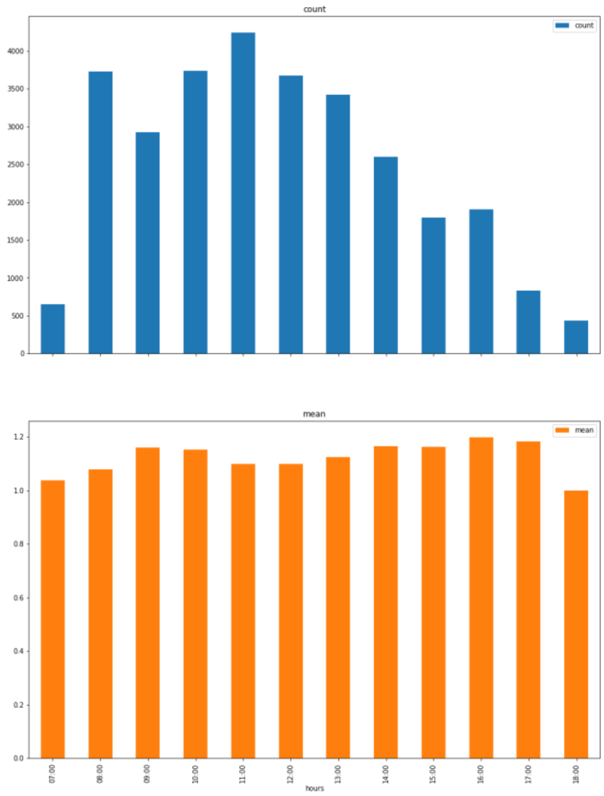


Figure 200: Interventions by hours at office pilot. Count of interventions (Top), Mean risk of the intervention (Bottom)

In the office case, there are a higher number of interventions in the first hours of the day. However, if we look at the average value of risk, we can see that as the day progresses, the risk becomes greater. This means that users have a worse posture as the day goes on (x-axis describes the time of the day, the y-axis the number of interventions at this time).

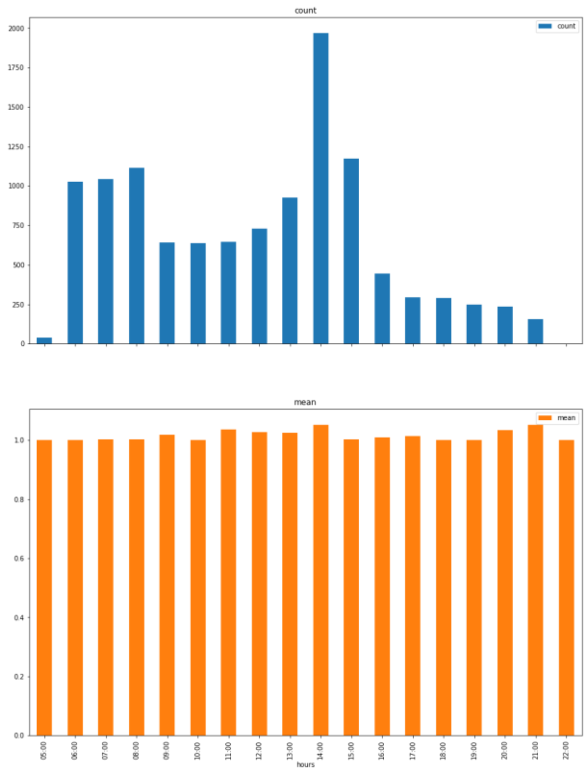


Figure 201: Interventions by hours at production pilot. Count of interventions (Top), Mean risk of the intervention (Bottom)

In the production case, the graph is completely different because the workers have different shifts, a morning or an afternoon one (Figure 201). This means that the hours during which interventions were given are more extensive. Although certain patterns can be seen, in the middle of the graph (end of the tomorrow shift) the number of warnings increases, as does the risk of warnings. In the afternoon shift, the number of warnings is low, but when the last hour approaches, the risk level of warnings is high, as was the case in the office shift.

Another analysis was carried out by grouping the interventions by days of the week. In this case we can see that Thursdays have been the days with the highest number of interventions, although Mondays have had a higher risk (x-axis describes the weekday, the y-axis the number of interventions at that day).

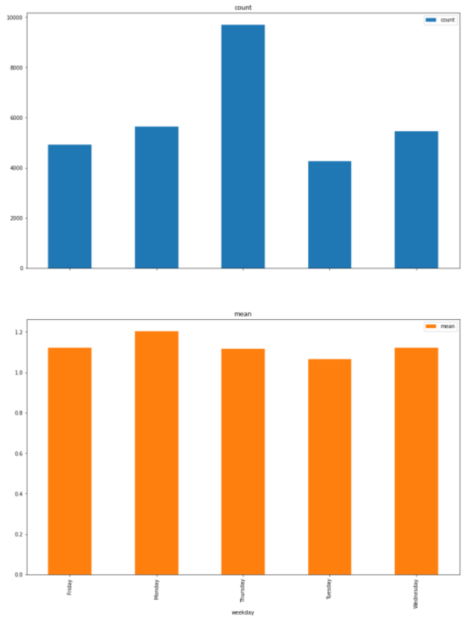


Figure 202: Interventions by weekdays at Office pilot. Count of interventions (Top), Mean risk of the intervention (bottom)

In the production case, a higher number of interventions are seen in the middle of the week (Tuesday/Wednesday/Thursday), although no significant differences in risk are seen (Figure 203) (x-axis describes the weekday, the y-axis the number of interventions at that day).

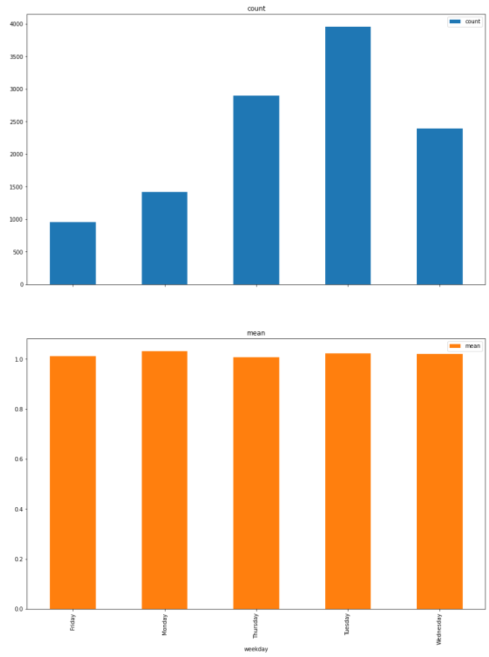


Figure 203: Interventions by weekdays at production pilot. Count of interventions (Top), Mean risk of the intervention (bottom)

The following graphs also show the number of warnings per day and the mean risk level of the warnings. But grouped by days, this way it is possible to see how users have evolved over the pilots.

Firstly, we can look at the office pilot. In this case, during the first days, there were few interventions. This is because users were registering gradually, so at the beginning there would only be a few active users, and as the pilot progressed, all of them. In this case, the times of highest activity coincide with the times of highest risk (Figure 204) (x-axis describes the day of usage, the y-axis the number of interventions at that day).

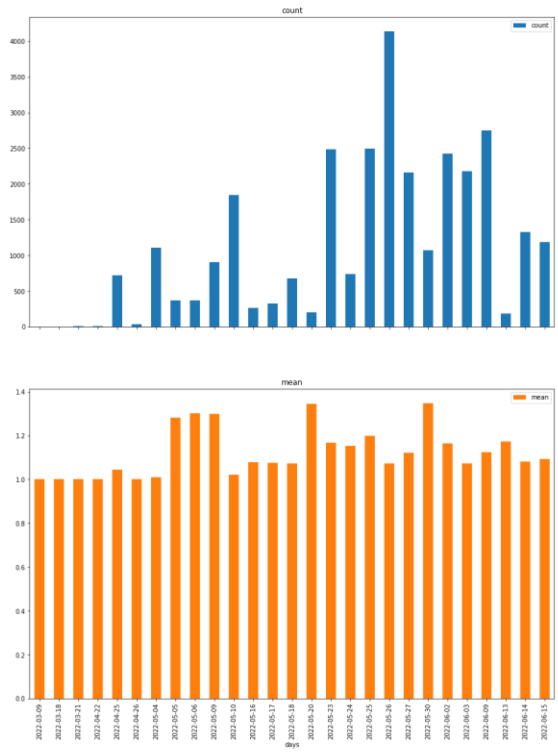


Figure 204: Interventions by days at Office pilot. Count of interventions (Top), Mean risk of the intervention (Bottom)

In the production pilot, this is different because all users were registered at the same time. However, it can be seen that from May 20th onwards, activity decreases notably. This may be due to users changing the location of their workplace and ceasing to use the system, communication failures or other problems. It is possible to see in the graph with the average risk level per day that the values have been generally low, with a certain peak around May 26th (Figure 205) (x-axis describes the day of usage, the y-axis the number of interventions at that day).

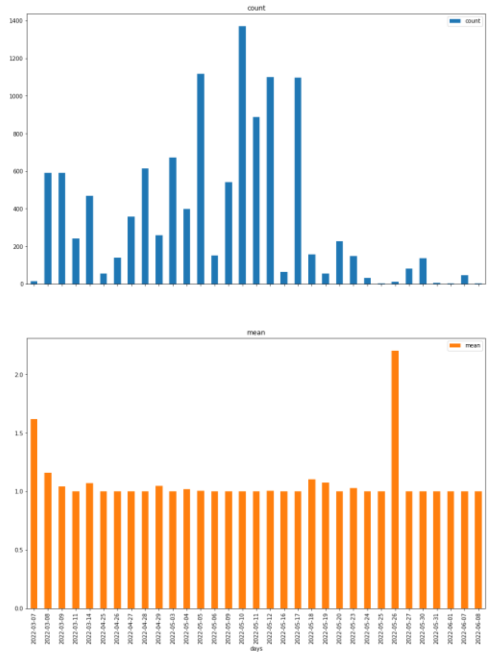


Figure 205: Interventions by days at production pilot. Count of interventions (Top), Mean risk of the intervention (Bottom)

2.6.5 Discussion and Outlook

In conclusion, we can say that the Body Pose component has worked well, but with certain limitations. The main problem is that in the pilots it was not possible to comply with the camera positioning requirements. This is a problem because the model was trained on a set of assumptions. To avoid this being a problem, one option is to fine-tune the application to suit each case adapting to the working environment and to a different camera viewpoint. In this way, even if the requirements cannot be met, the system will be able to detect the person, their joints and thus estimate the ergonomic risk in a much more reliable way. It has been a first version, but with the learning gained, the development of an ergonomic risk analysis tool can be carried out.

2.7 WAOW tool measurements: Facial Affect Analysis

2.7.1 Introduction & Methods

A number of facial affect analysis algorithms have been developed by UCAM, to monitor user's display of facial action units (AU) and dimensional facial affect (arousal and valence).

We developed three different software and algorithms for facial action unit detection: 1) ResNet-50 based, 2) AULA-Caps (Churamani et al., 2020), and 3)Graph-based (Luo et al., 2022). Each of these provide a real-time evaluation of the users' facial gestures in terms of 12 facial action units (AU1, AU2, AU4, AU6, AU7, AU10, AU12, AU14, AU15, AU17, AU23 and AU24).

1) The ResNet-50 system provides a baseline system trained and evaluated on BP4D dataset that works in real-time on a frame-by-frame basis. This subsystem is designed in two versions, as a standalone solution and as a version that works in combination with the WAOW tool.

2) The Action Unit Lifecycle-Aware Capsule Network (AULA-Caps) (Churamani et al., 2021) is a novel algorithm we developed based on spatial and spatio-temporal features encoding the temporal evolution of facial AU activations, using both frame and sequence-level features. AULA-Caps exists as a standalone solution.

3) The novel graph-based algorithm proposes an AU relationship modelling approach that deep learns a unique graph to explicitly describe the relationship between each pair of AUs of the target facial display (Luo et al., 2022). The PyTorch code is open access: https://github.com/CVI-SZU/ME-GraphAU. The graph-based AU detection has been evaluated as a standalone solution.

We developed two different algorithms for inferring facial affect in terms of arousal and valence prediction: 1) ResNet-50 based and 2) Graph based.

1) The ResNet-50 system provides a baseline system trained and evaluated on AffectNet dataset that works in real-time on a frame-by-frame basis and infers the arousal and valence level of the captured facial display. After acquiring facial video frames, the subsystem continuously returns a valence and an arousal prediction for every 1 min. This ResNet-50 system is designed in two versions, as a standalone solution and as a version that works in combination with the WAOW tool. This version was integrated into the WAOW tool and was tested via short-term tests – we refer to this subsystem the Integrated Facial Affect Analysis Subsystem (IFAAS).

2) The novel graph-based algorithm is built upon the graph-based AU detection algorithm described above in point (3), and it infers facial affect in terms of periodic arousal and valence prediction. After acquiring facial video data, the subsystem continuously returns a valence and an arousal prediction for every 1 min. It relies on a new real-time periodical dimensional affect (arousal and valence) estimation framework (based on a well-trained graph neural network), where arousal and valence status of the user are independently categorized into three classes (Low, Medium, and High). The testing of the software and the evaluation of the algorithms for predicting user´s dimensional affect status is conducted on all video data (from four different sites, namely AUD and RWTH in Germany, BS in Italy and UCAM in the UK) collected during the in-lab study. Graph-based algorithm works standalone and was tested on in-lab facial data from 4 partner sites – we refer to this subsystem as Graph-based Facial Affect Analysis Subsystem (GFAAS).

2.7.2 IFAAS Short-term tests

EXUS site

IFAAS code was implemented in a container (docker) and was integrated and deployed on the EXUS/UCAM server. The implementation required one docker container per user.

Tests: Short-term tests were conducted on EXUS site in two stages: 1) with a single user and 2) with multiple / all users. Single user tests confirmed that the IFAAS subsystem was working with all services running smoothly. The multi-user tests were conducted by starting all dockers for all users. It was confirmed that they all initialized simultaneously indicating that they all worked, however, the user registration process would fail due to insufficient memory. It was determined that any single user docker would run, therefore it was not a coding error. After multiple and iterative evaluations, it was found that 3 IFAAS containers could run simultaneously, before consuming all GPU memory. Also note that during this process EXUS stopped the gesture component, as both IFAAS and Gesture Subsystem shared the resources of the same server and GPU. The GPU resource consumption was monitored at the same time, and each IFAAS docker required ~2-3GB of memory on the GPU (8GB total) on that specific EXUS/UCAM server machine. IFAAS, similarly to the Gesture Subsystem, processes live video stream (as opposed to e.g., the BodyPose Subsystem that could process a snapshot and periodically, such as 1 frame per minute) at ~25-30 frames per second which requires large GPU requirements for each docker.

Data: The data received from three cameras working simultaneously was stored on the EXUS/UCAM server on EXUS site. The subsystem outputs were checked and confirmed that valence and arousal predictions were obtained, and values stored in relevant files. The statuses of the cameras were also checked and confirmed via log files of the containers. However, accessing the data on the FTP server proved problematic despite investigations by the EXUS team. Before the investigations could be concluded, the UCAM/EXUS server at the Exus site was damaged due to a flood and could not be brought back online by the time UCAM integration person's contract finished in the first half of May 2022. Later checks showed that all UCAM code and data was lost on the server due to the flood incident.

Additional investigations related to the saving of the data files showed that there was an issue with WAOW App – if the data is more than certain size (>5 MB), only blank text files were being saved. This issue was reported and was planned to be resolved in the updated version of the app, however by little this did not take place by the time UCAM integration person's contract finished in the first half of May 2022.

Mutua site

IFAAS code was implemented in a container (docker) and was integrated and deployed on the EXUS/UCAM server first located at ITCL and then moved to Mutua. The implementation required one docker container per user.

Tests: IFAAS code was integrated and configured successfully, and it was tested successfully with the ITCL team. There were a couple of issues related to intermittent internet access and GPU drivers that was vital for the docker image to run on that site's server. This issue was reported and discussions were made with Exus, however things did not get resolved by the time UCAM integration person's contract finished in the first half of May 2022.

2.7.3 Evaluation and Results of GFAAS

Dataset: The dataset contains data from a total of 55 participants, with 7 participants coming from the AUD site, 16 participants coming from the BS site, 20 participants coming from the RWTH site, and 12 participants coming from the UCAM site. It contains 935 clips (each clip corresponds to a sub-task, and 17 sub-tasks were recorded for each participant), where 565 of them are annotated using the joint protocol agreed upon for the 4-site in-lab study.

Evaluation protocol: As the in-lab data is collected from four different sites, we conduct the leave-one-site-out validation protocol for models' training and evaluation, i.e., at each time, we use all clips from three sites (three different partner's data) to train a model, and evaluate the trained model on the remaining one site (one partner's data).

Metric: The classification accuracy is employed as the measurement to evaluate facial affect recognition.

1. **Baseline systems:** we implement three standard deep learning-based short video-level modelling approaches to provide a benchmark for the task of facial affect recognition in work-like settings:
   Baseline 1: Given a short face video, the first baseline starts with generating frame-level affect predictions for all frames, which are then combined to output periodical affect prediction. In particular, we individually employed two frame-level facial analysis models, i.e., a ResNet-50 that is pre-trained for facial expression recognition (i.e., pre-trained on the FER 2013 dataset (FER, 2013) and our novel GraphAU model (to be presented at IJCAI'22 at the end of July 2022) (Luo et al., 2022) that is pre-trained for facial action unit (AU) recognition (i.e., pre-trained on the BP4D dataset (Zhang et al., 2014). We individually use the latent feature output by the second-last fully connected layer of the ResNet-50, as well as the 12 AU predictions generated by our novel GraphAU model, as the frame-level facial features. Then, we individually apply a multi-layer perceptron (MLP) on each of them to provide frame-level valence and arousal predictions, which is trained by re-using the clip-level self-reported valence/arousal scores as the frame-level label. To obtain periodical (clip-level) affect predictions, we combine

all frame-level predictions of the target clip with the following widely-used strategies: **(i)** using the mode prediction of all frame-level predictions as the periodical affect predictions; **(ii)** applying a Long-short-term-memory Network (LSTM) to combine all frame-level predictions; and **(iii)** applying spectral encoding algorithm (Song et al., 2020) (which has been shown to encode dynamic cues) to produce a spectral heatmap from all frame-level predictions, which is then fed to a 1D-CNN to generate periodical affect predictions.

Baseline 2: The second baseline also applies the same two pre-trained models used in baseline 1 to provide frame-level facial features. Differently from the baseline 1, we employ three long-term modelling strategies to combine all frame-level facial features of the clip as the clip-level (periodical) affect representation: **(i)** averaging all frame-level facial features; **(ii)** applying LSTM to process all frame-level facial features; and **(iii)** spectral encoding all frame-level facial features. These clip-level affect representations are then fed to either an MLP (for (i)) or 1D-CNN (for (ii) and (iii)) to generate clip-level affect predictions.

Baseline 3: The third baseline applies a spatio-temporal CNN (Temporal Pyramid Network (TPN)) to process the facial sequence. In particular, we first divide each clip into several segments, where each consists of 160 frames, and down-sample each segment to 32 frames. We then feed the cropped face sequence (32 frames) to TPN for affect classification. If a clip contains multiple segments, then the clip-level predictions are obtained by averaging all segment-level predictions.

2. **Leave-one-site-out cross-validation results:** Table 8 and Table 9 show the valence and arousal classification results achieved by all baseline systems for all sub-tasks (the models are trained using all clips in the training set regardless of the task type). The listed results are the average result of four folds. It can be seen that the majority of the baselines achieved over 51% classification accuracy, with the GraphAU(P)-SE system achieving the best valence classification result (54.50%) and the ResNet(P)-SE system achieving the best arousal classification result (53.71%). Meanwhile, we found that if we only use the mode prediction of all frame-level predictions, both valence and arousal classification results are clearly worse than the other systems, i.e., the two corresponding systems only achieved less than 50% valence and arousal classification accuracy. These results indicate that: (i) according to the Figure 206, the long-term modelling for either frame-level predictions or features is a crucial step to achieve more reliable periodical arousal/valence predictions; (ii) the frame-level facial analysis (AU recognition/facial expression recognition) models that are pre-trained using the lab-based facial datasets can still extract human affect-informative facial features from facial displays triggered by work-like tasks, as their features frequently provide more than 50% accuracy for both tasks (the chance-level accuracy should be around 33% for three class classification); and (iii) directly pairing workers' facial sequences with clip-level affect labels

to train spatio-temporal models does not provide superior results, which further validates that frame-level facial analysis models pre-trained on facial datasets acquired in naturalistic settings are beneficial for facial affect analysis in work-like settings.

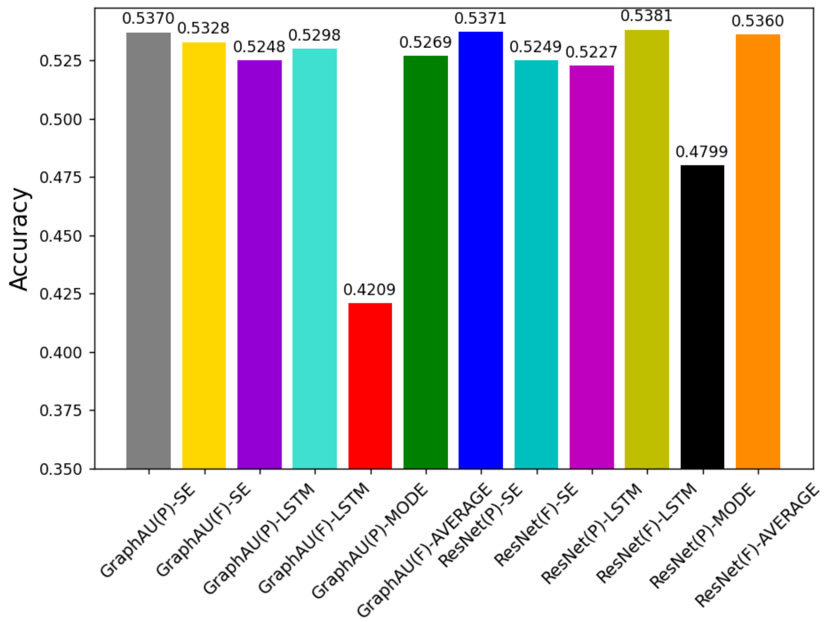


Figure 206: The average valence and arousal classification accuracy of all baseline models.

Table 9 shows the results achieved for worker's valence recognition, where the name of each method is formatted as frame level facial feature-long term model, where P and F represent the frame-level prediction and facial features, respectively. For example, ResNet(P)-SE denote the system that applies ResNet facial features to make frame-level affect predictions, and then using spectral encoding algorithm to summarise all frame-level valence/ predictions as the clip-level valence/arousal prediction.

Table 9: Results achieved for worker's valence recognition

Model	NBB	NBE	NBH	NbS	DB	DE	DH	DS	WEB	WEN	WEP	Mean
GraphAU (P)-SE	0.5943	0.5929	0.6114	0.5543	0.6564	0.5424	0.5091	0.5673	0.6729	0.3700	0.3204	0.5450
GraphAU (F)-SE	0.5743	0.5929	0.5914	0.6143	0.6764	0.5424	0.4491	0.5091	0.6129	0.2800	0.3840	0.5297
GraphAU (P)-LSTM	0.5943	0.6129	0.5714	0.5543	0.6564	0.5091	0.5291	0.5491	0.6729	0.3200	0.3040	0.5339
GraphAU (F)- LSTM	0.5943	0.6129	0.5714	0.5543	0.6564	0.5091	0.5291	0.5491	0.6729	0.3700	0.2640	0.5349
GraphAU (P)-MD	0.5100	0.4086	0.5271	0.5143	0.5345	0.3673	0.2255	0.3236	0.3871	0.3400	0.2800	0.4016
GraphAU (F)-AG	0.6143	0.5929	0.5714	0.5543	0.6164	0.4891	0.5091	0.5291	0.6329	0.3700	0.3200	0.5272
ResNet (P)-SE	0.5286	0.6214	0.5357	0.5443	0.7021	0.4883	0.5805	0.4862	0.6814	0.4800	0.2600	0.5371
ResNet (F)-SE	0.5971	0.6386	0.5743	0.5829	0.6449	0.4912	0.5948	0.5234	0.6986	0.3600	0.2600	0.5423

ResNet (P)-LSTM	0.5800	0.6386	0.5743	0.5657	0.6478	0.5177	0.5777	0.5234	0.6814	0.3600	0.2600	0.5388
ResNet (F)-LSTM	0.5800	0.6386	0.5743	0.5657	0.6478	0.5005	0.5948	0.5234	0.6986	0.3600	0.2600	0.5403
ResNet (P)-MD	0.5900	0.5662	0.6287	0.4747	0.4882	0.4135	0.4200	0.4383	0.5513	0.2469	0.3948	0.4739
ResNet (F)-AG	0.5800	0.6386	0.5743	0.5657	0.7021	0.5005	0.5777	0.5405	0.6814	0.3600	0.2600	0.5437

Table 10: The results achieved for worker's arousal recognition.

Model	NBB	NBE	NBH	NbS	DB	DE	DH	DS	WEB	WEN	WEP	Mean
GraphAU (P)-SE	0.7571	0.5929	0.4671	0.3057	0.6345	0.4756	0.3818	0.3673	0.6986	0.6500	0.4880	0.5290
GraphAU (F)-SE	0.7171	0.6129	0.5671	0.3257	0.6345	0.4793	0.3836	0.3473	0.6786	0.6200	0.5280	0.5358
GraphAU (P)-LSTM	0.7371	0.6129	0.5271	0.3057	0.6745	0.4018	0.3218	0.3255	0.6586	0.6200	0.4880	0.5157
GraphAU (F)- LSTM	0.7371	0.6729	0.5471	0.3057	0.6527	0.4236	0.3218	0.3255	0.6786	0.6200	0.4880	0.5248
GraphAU (P)-MD	0.5914	0.6357	0.4271	0.2043	0.4291	0.4291	0.3873	0.2873	0.5757	0.5700	0.3040	0.4401
GraphAU (F)- AG	0.7571	0.6529	0.5271	0.3257	0.6945	0.4018	0.3218	0.3255	0.6786	0.6200	0.4880	0.5266
ResNet (P)-SE	0.7657	0.6600	0.5300	0.2743	0.6148	0.4536	0.4434	0.3587	0.7071	0.5600	0.5400	0.5371
ResNet (F)-SE	0.7143	0.6014	0.4929	0.4314	0.5795	0.5353	0.4213	0.3369	0.5900	0.4600	0.4200	0.5075
ResNet (P)-LSTM	0.7100	0.5400	0.4914	0.3129	0.6603	0.3961	0.3561	0.3569	0.6700	0.6200	0.4600	0.5067
ResNet (F)-LSTM	0.6700	0.6214	0.5129	0.3929	0.6366	0.5252	0.4452	0.4005	0.7100	0.5600	0.4200	0.5359
ResNet (P)-MD	0.6857	0.5614	0.4357	0.2400	0.5917	0.4179	0.3608	0.3055	0.6671	0.6000	0.4800	0.4860
ResNet (F)- AG	0.7114	0.6443	0.5329	0.3114	0.6584	0.4616	0.4016	0.3587	0.6700	0.5800	0.4800	0.5282
TPN	0.7389	0.6477	0.5459	0.2167	0.5609	0.4294	0.3306	0.3467	0.6928	0.5536	0.4897	0.5048

3. **Ablation studies:** we evaluated the influences of the task type and recording site as follows:

3.1 Task type: Figure 207 reports the average results achieved by all baselines for each task. It is clear that valence can generally be better predicted using facial behaviours under different working conditions, with 52.07% average recognition accuracy compared to arousal predictions (51.45%). Specifically, the facial behaviours triggered by eight sub-tasks allow the model to achieve over 50% recognition accuracy, where the valence prediction results of three baseline tasks are superior to other sub-tasks. Based on this result, we hypothesize that different subjects display affect in different intensity (arousal) when undertaking the same task (even though their facial behaviour may be similar) due to their person-specific attributes (e.g., background, personality, etc.). The person-specific attributes particularly impact the positive and negative memory tasks, triggering the display of different levels of arousal. Facial behaviours displayed during baseline sub-tasks are also very informative for predicting subjects' arousal, i.e., the arousal classification results achieved for the baseline sub-tasks of N-Back, Operation and Video Conference have more than 9.5%, 17.2% and 7.5% absolute accuracy improvements over the results of other sub-tasks, respectively. Moreover, for both N-back and Operation tasks, the arousal classification accuracies achieved on sub-tasks that require less workload are clearly higher than the results achieved on more difficult tasks. This finding suggests that human facial displays triggered by high workload are not reliable for inferring worker's arousal state.

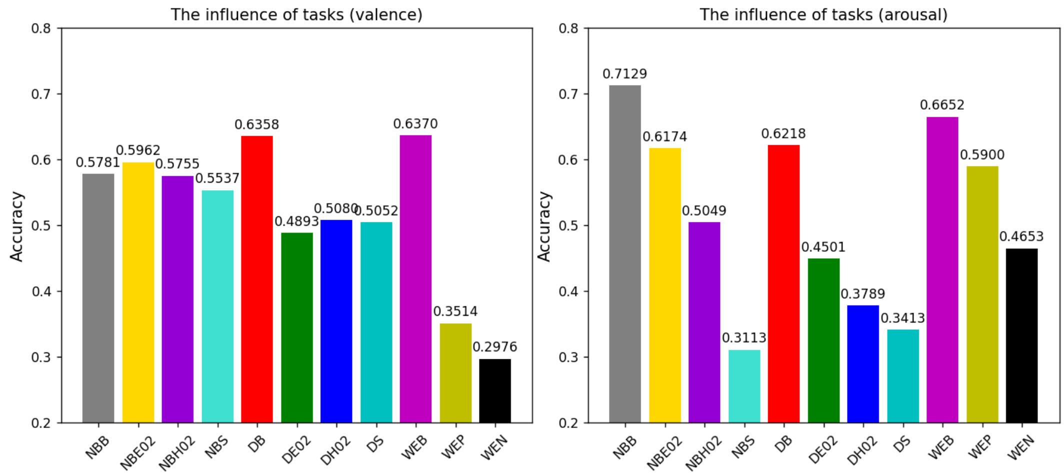


Figure 207: The influence of different tasks on valence and arousal prediction.

3.2 Recording site: We also explore the differences in affect classification for different sites. Table 11 displays the leave-one-site-out four-fold cross-validation results. It is clear that the data collected at different sites impacts the valence classification results, with more than 10% accuracy difference between the lowest (0.4803 (AUD)) and the highest (0.5890 (RWTH)). These results indicate that people at different sites may display different facial behaviours when expressing valence. On the other hand, the performance variations for arousal classification are much smaller, indicating that the relationship between arousal and workers' facial behaviours are more stable as compared to valence.

Table 11: The results of the four-fold cross-validation results achieved by our best models (GraphAU(P)-SE for valence, ResNet(P)-SE for arousal).

Validation set	**AUD**	**BS**	**RWTH**	**UCAM**
Valence	0.4803	0.5351	0.5890	0.5764
Arousal	0.5455	0.5766	0.5545	0.5048

2.7.4 Discussion and Outlook

The short-term tests showed that the facial affect analysis subsystem was successfully integrated to the WAOW tool (the IFAAS subsystem) and was confirmed to be working and producing valence and arousal predictions. However, a number of challenges (as discussed in 2.7.2) were faced and not all of them were resolved in time. We have learned valuable lessons from this process that will inform our future works.

The evaluation of the new algorithm (GFAAS) on the in-lab data from four sites show that facial behaviours are informative for inferring workers' periodical valence and arousal states under all tasks, but the performances are highly dependent on the *task type* and *setting*. Among all evaluated systems, we would choose the *GraphAU(F)-AVERAGE* system as the preferred subsystem, as it achieved good performance on both valence and arousal classification and does not suffer from heavy post-processing steps. However, since this version was not evaluated with short-term and long-term in-company tests, further investigations should be undertaken to support these findings.

Media hype illustrates emotion recognition to be a solved research problem with accuracy levels going beyond 90%. However, this is only true for strictly posed and strictly controlled environments and for the recognition of exaggerated facial expressions of basic emotions. The reality is instead illustrated with various Challenge events the research community has been organising since early 2010s. For instance, the winner of the AVEC'19 Cross-cultural Emotion Sub-challenge (Zhao et al., 2019) reported the best concordance correlation coefficient (CCC) performance of 0.4 and 0.471 for arousal and valence on the challenge testing set, by training multimodal prediction models and smoothing the results by averaging the predictions within a fixed window. Such results demonstrate that emotion recognition in naturalistic settings is still a challenging problem, open for future research, and fusing multimodal features helps improving the prediction results. The implications of this for the future research and improvements on the WAOW Tool is to aim to fuse the features from different sensors, including eye gaze, audio, physiology, and face to predict user affect and mental states.

Finally, it is important to note that the openly available facial affect datasets used for algorithm training generally contain displays of young and middle-aged adults. Affect recognition models trained on these do not generalize well on other user groups such as the elderly or the children. As WorkingAge aims to create smart working environments for all ages, in parallel to the abovementioned work, we at UCAM also (i) undertook an empirical investigation of bias and fairness for facial expression recognition (Xu et al., 2020) which has received over 60 citations already; (ii) created a conceptual guide with recommendations for fairness in facial affective signal processing (Cheong et al, 2021); (iii) investigated the application of Continual Learning paradigms for bias mitigation in facial expression recognition and facial action unit detection (Churamani et al., 2022), and (iv) contributed to a Nature Scientific Report (Hupont et al., 2022) with colleagues from the European Commission on '*The* *landscape of facial processing applications in the context of the European AI Act and the development of trustworthy systems'* by providing critical and technical expertise on the relevant topics. We are confident that these efforts constitute important milestones in ensuring that future affect recognition systems are built and applied in an ethical and fair manner.

2.8 WAOW tool measurements: Neurometrics

2.8.1 Introduction

The Neurometrics Subsystem was used for the online evaluation of the users' mental workload, stress, and emotional state. The testing of the software and hardware and especially the algorithms for the calculation of the above-mentioned user´s mental states were already performed during the In-Lab studies (please refer to Deliverable 9.3). The Neurometrics Subsystem included two wearable devices to perform the neurophysiological measurements: the Muse S (Interaxon, Canada) and the Empatica E4 (Empatica, Italy). In the context of the pilot studies, the applicability and usability of such wearable sensors were assessed. In particular, the main objectives of the Short-Term tests consisted in:

- Assessing the correct functionality of the Neurometrics Subsystem in the context of the entire WAOW tool and its correct interaction with the Decision Support System (DSS).
- Performing the algorithms calibration per each user involved in the tests. This operation was required because we were not allowed for making any *a priori* assumptions on the involved users and no-calibration data related to the involved users were available in the pre-test phase.

2.8.2 Method

The Neurometrics sensors included in the WAOW tool consisted of the Empatica E4, a wristband for gathering the Electrodermal Activity (EDA) and the Photoplethysmographic (PPG) signals for extracting the Skin Conductance Level (SCL) and the Heart Rate (HR), and the Muse S, a headband collecting the Electroencephalographic (EEG) signal for extracting the Eye Blinks Rate (EBR). The SCL was involved in the stress evaluation (Borghini et al., 2020), while the combination of the SCL and the HR corresponded to the Emotional Index (EI) for the emotional state evaluation (Vecchiato et al., 2014). The EBR parameter was used as a correlate of the mental workload (Charles & Nixon, 2019). Such mental states were evaluated along all the users' working day.

1) Test procedure

- Turn on the Empatica E4, the Muse S and the Intel NUC.
- The user must wear the Empatica E4 and the Muse S correctly (please see the Mounting Instructions in D9.1).

- After few seconds the Empatica E4 and the Muse S will pair automatically to the Windows Intel NUC.

Regarding subjective data collection to validate such measurements, five times per day each worker was presented with a pop-up window on the WAOW mobile app that asked them to respond to five questions related to their cognitive load, stress, and affective states during the previous 10 minutes (provided in the In-Company tests - ISA scale Quantitative Feedback Request Form).
During the short-term phase of the experiments the above-mentioned neurophysiological data were collected in order to calibrate the algorithms for the mental and emotional states classification. In particular, the high-level information about the user's mental states required a per-user calibration. Such a calibration is needed to estimate the appropriate classification thresholds for each user. The relation between the neurophysiological features considered and the correspondent mental and emotional states are described below:

- As the EBR and mental workload are inversely related, every time the user's EBR is below the classification threshold, the mental workload is labelled as High. On the contrary, when the user's EBR is above the classification threshold, the mental workload is labelled as Low.
- The SCL and the stress are directly related. Therefore, every time the user's SCL is above the classification threshold, the stress level is labelled as High. Otherwise, the stress level is labelled as Low.
- The EI is directly related with the emotional state. If the EI values were lower than their corresponding mad values, they will be set to zero. This aspect was considered to define the *Neutral* state. The EI can therefore assess if the user is experiencing a, *Positive*, or *Negative* overall emotional status.

For each of the three neurophysiological parameters, the *EBR*, *SCL* and *HR*, the classification threshold was computed as follows:

$$Threshold = 25^{th}\ percentile(X) + std(X)$$

where, the *X* is the distribution of the values considered and is acquired during the calibration phase, the *25th percentile* is the number between the smallest number (minimum) and the median of the dataset, and *std* is the standard deviation of the distribution.

2.8.3 Results – Short Term Tests

Using the Neurometrics Subsystem, 18 measurements were recorded on six persons among the three In-Company sites (3 in Mutua Universal, 2 in EXUS, 1 in Grupo Antolin), with a total measurement duration of over 15 hours. The environmental conditions as well as the user requirements were tested using the D9.1 evaluation protocol.

The Neurometrics sensors were used during the measurements for as long as possible during each working day, downtime occurred when breaks were taken, or the sensors needed a battery charge. The classification thresholds, and the classification performance (accuracy, specificity, and sensitivity) related to each parameter, i.e., the EBR, HR and SCL, computed for each participant are reported in Table 12:

Table 12: Classification thresholds related to each neurphysiological parameter evaluated through the Neurometrics sensors during the short-term tests. Related to each parameter the classification performances (accuracy, specificity and sensitivity) are reported.

Subject ID	EBR threshold (eyeblinks per second)	EBR classification performance	HR threshold (BPM)	HR classification performance	SCL threshold (µS)	SCL classification performance
Mutua – Subj1	0.27	Accuracy = 81%	81.27	Accuracy = 80%	2.11	Accuracy = 80%
		Specificity = 94%		Specificity = 92%		Specificity = 92%
		Sensitivity = 66%		Sensitivity = 65%		Sensitivity = 65%
Mutua – Subj2	0.24	Accuracy = 80%	77.54	Accuracy = 78%	1.88	Accuracy = 79%
		Specificity = 92%		Specificity = 90%		Specificity = 91%
		Sensitivity = 67%		Sensitivity = 64%		Sensitivity = 64%
Mutua – Subj3	0.25	Accuracy = 83%	72.81	Accuracy = 84%	2.56	Accuracy = 85%
		Specificity = 94%		Specificity = 95%		Specificity = 96%
		Sensitivity = 69%		Sensitivity = 70%		Sensitivity = 70%
EXUS – Subj1	0.33	Accuracy = 84%	88.51	Accuracy = 82%	0.96	Accuracy = 81%
		Specificity = 94%		Specificity = 94%		Specificity = 93%
		Sensitivity = 70%		Sensitivity = 68%		Sensitivity = 67%
EXUS – Subj2	0.31	Accuracy = 80%	84.82	Accuracy = 80%	1.58	Accuracy = 80%
		Specificity = 93%		Specificity = 92%		Specificity = 92%
		Sensitivity = 67%		Sensitivity = 68%		Sensitivity = 68%
Grupo Antolin – Subj1	0.48	Accuracy = 79%	100.51	Accuracy = 77%	3.27	Accuracy = 79%
		Specificity = 91%		Specificity = 90%		Specificity = 91%
		Sensitivity = 65%		Sensitivity = 63%		Sensitivity = 64%

2.8.4 Results – Long-term Tests

Using the Neurometrics Subsystem, 48 measurements were recorded on six persons among the three In-Company sites during the long-term tests, with a total measurement duration of over 40 hours.

A Pearson's correlation analysis was performed to compare the classification outcomes of the Neurometrics Subsystem with the outcomes provided by the other Subsystems involved in the cognitive and physical state evaluation of the participants. In particular, the emotional state classifications performed through the EI derived by the Neurometrics Subsystem were compared with the classification outcomes of the Voice Analysis Subsystem. Four participants were selected to be included in the Pearson's correlation analysis, as they were monitored with both the Neurometrics and Voice Analysis Subsystem along the same working days. The results are reported in Table 13:

Table 13: Pearson's correlations between the emotional state classifications provided by the Voice Analysis Subsystem and the ones provided by the Neurometrics Subsystem.

Subject ID	**Audio vs. Emotional state correlation**	**Time period**
Mutua – Subj1	**r = 0.69; p = 0.0002**	from 07/06 to 08/06 (09:30 – 10:30) and (14:10 – 15:10)
Mutua – Subj2	**r = 0.73; p = 0.0001**	from 07/06 to 08/06 (09:45 – 10:45) and (14:15 – 15:15)
Mutua – Subj3	**r = 0.59; p = 0.01**	from 07/06 to 08/06 (09:50 – 10:50) and (14:00 – 15:00)
EXUS – Subj1	**r = 0.77; p = 0.0002**	from 06/06 to 07/06 (11:55 – 12:55) and (15:40 – 16:40)

Similarly, the Pearson's correlation analysis was performed between the Body Pose Subsystem classification outcomes and the mental states classifications provided by the Neurometrics Subsystem. In particular, the neck position classifications were compared with the mental workload classifications and the lower back position classifications were compared with the stress classifications, since different previous works in literature (Shahidi et al., 2013; X. Wang et al., 2022) assessed such relationships. The results are reported in Table 14:

Table 14: Pearson's correlations between the mental states classifications provided by the Neurometrics Subsystem and the body poses classifications provided by the Body Pose Subsystem.

Subject ID	**Neck position vs. Mental workload - correlation**	**Lower Back position vs. Stress - correlation**	**Time period**
Mutua – Subj1	**r = 0.57; p = 0.02**	**r = 0.67; p = 0.005**	from 04/06 to 08/06 (09:30 –

			10:30) and (14:10 – 15:10)
Mutua – Subj2	**r = 0.64**; **p = 0.01**	**r = 0.55**; **p = 0.01**	from 04/06 to 08/06 (09:45 – 10:45) and (14:15 – 15:15)
Mutua – Subj3	**r = 0.60**; **p = 0.01**	**r = 0.59**; **p = 0.01**	from 04/06 to 08/06 (09:50 – 10:50) and (14:00 – 15:00)
Grupo Antolin – Subj1	**r = 0.53**; **p = 0.04**	**r = 0.64**; **p = 0.008**	from 03/06 to 07/06 (07:05 – 08:00) and (14:10 – 15:00)

Finally, to validate the mental states classifications performed through the Neurometrics Subsystem, a Pearson's correlation analysis was performed between the normalized parameters representing the mental workload (Figure 208),stress (Figure 209), emotional state (Figure 210) and the respective subjective evaluation provided through the ISA scales by the participants. The results are reported in Table 15 below:

Table 15: Pearson's correlations between the mental states classifications provided by the Neurometrics Subsystem and the respective subjective evaluations provided by the participants through the ISA scales.

Subject ID	**Mental workload vs. ISA - correlation**	**Stress vs. ISA - correlation**	**Emotional state vs. ISA - correlation**
Mutua – Subj1	**r = 0.68; p = 0.01**	**r = 0.65; p = 0.01**	**r = 0.78; p = 0.0001**
Mutua – Subj2	**r = 0.65; p = 0.02**	**r = 0.56; p = 0.03**	**r = 0.57; p = 0.02**
Mutua – Subj3	**r = 0.73; p = 0.0009**	**r = 0.76; p = 0.0005**	**r = 0.75; p = 0.0008**
Grupo Antolin – Subj1	**r = 0.61; p = 0.01**	**r = 0.66; p = 0.02**	**r = 0.57; p = 0.03**

The following figures (Figure 208 - Figure 210) represent the time dynamic of each mental and emotional state evaluated through the Neurometrics Subsystem and the corresponding ISA score. In particular, the orange bars correspond the time dynamics of the neurophysiological parameters representing the above-mentioned mental and emotional states, while the blue lines represent the time dynamics of the ISA subjective scores.

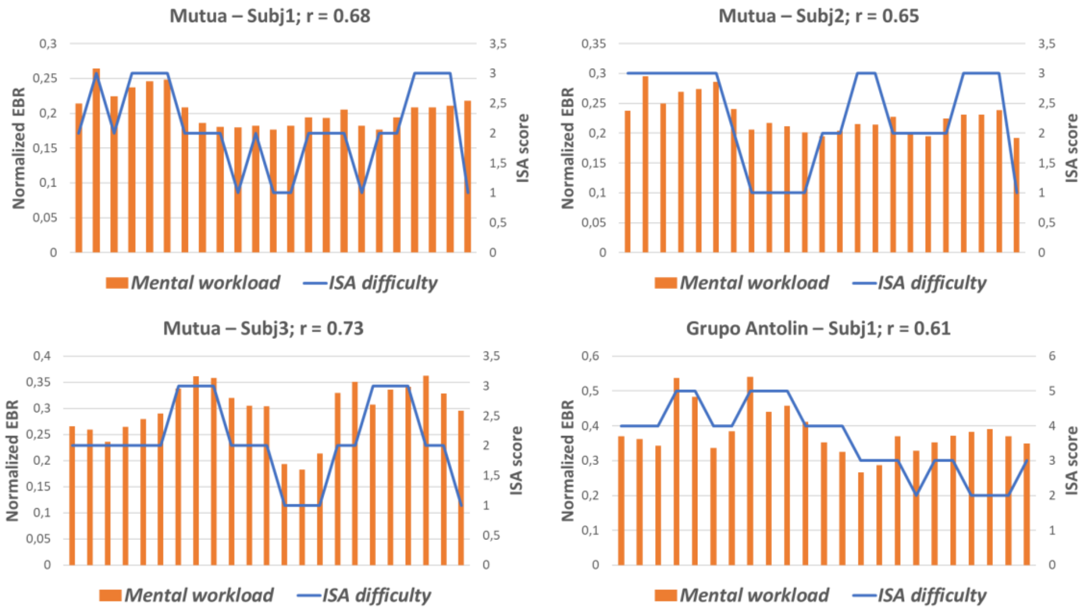


Figure 208. Time dynamics of the mental workload index and the ISA difficulty score along the working day for each monitored participant.

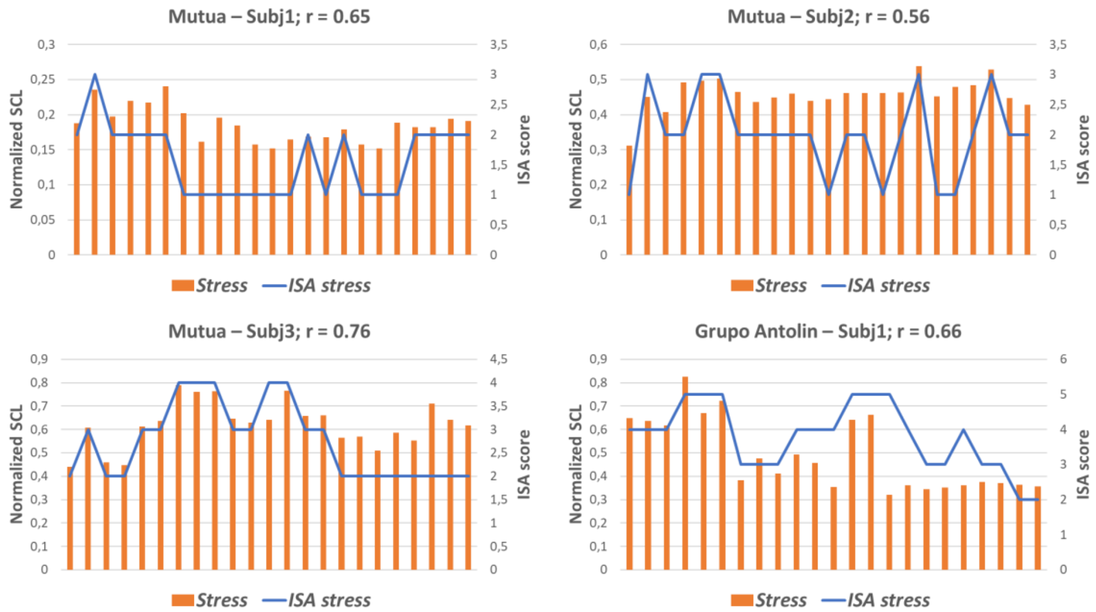


Figure 209. Time dynamics of the stress index and the ISA stress score along the working day for each monitored participant.

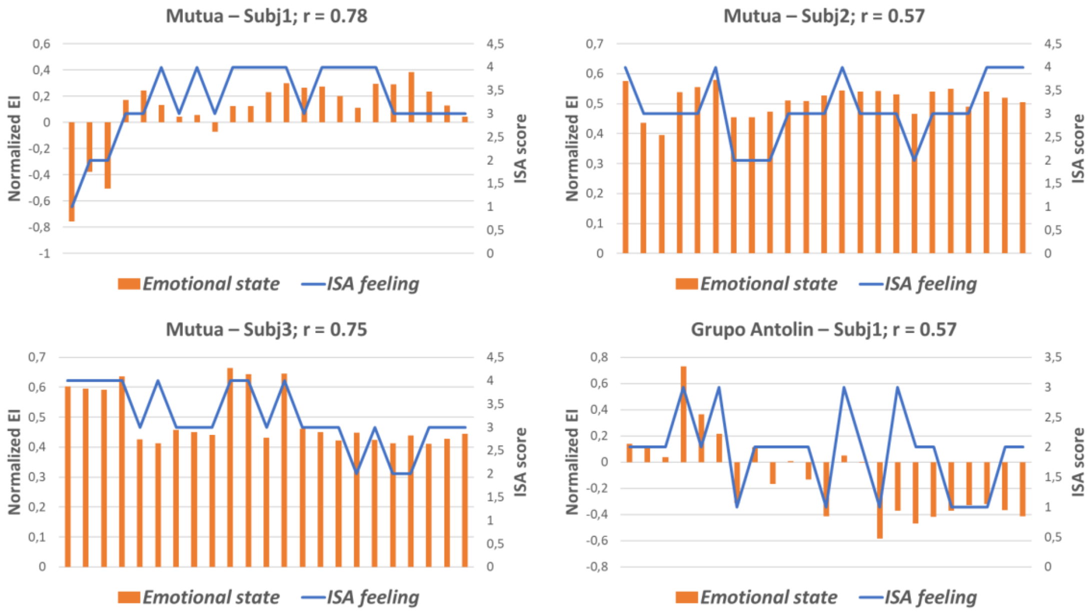


Figure 210. Time dynamics of the emotional index and the ISA feeling score along the working day for each monitored participant.

2.8.5 Summary and Outlook

The results related to the Neurometrics Subsystem are very promising, especially considering the limitations and impacts due to the pandemic and semi-conductor crisis. Indeed, such two factors relevantly delayed the overall technical development of the different Subsystems and they did not allow for properly testing the entire software segment, which plays a crucial role in the Neurometrics Subsystem functioning.

Despite this aspect, the short-term tests allowed us to properly calibrate the mental and emotional states classification algorithms for the participants involved in the In-Company phase among the three testing sites. Additionally, the sensors acceptance analysis (please see Paragraph 2.4) confirmed the positive approach towards the wearable sensors integrated in the Neurometrics Subsystem by the participants involved in all the three In-Company sites.

Regarding the long-term tests, the Pearson's correlations between the neurophysiological data used for the mental workload, stress and emotional state estimations and the subjective data are positive, high, and significant for all the monitored participants during the In-Company phase. Furthermore, the additional Pearson's correlation analysis performed to enhance the effectiveness of the monitoring capacity of the different Subsystem included in the WAOW tool, i.e., the comparisons between the mental states classifications provided by the Neurometrics, the body pose classifications and the emotional state classifications provided by the Voice Analysis, resulted to be positively and significantly correlated for all the monitored participants along the Long Term tests in the context of the In-Company phase.

Finally, it is crucial to note the advantages associated to the use of neurophysiological measurements for the mental and emotional states evaluation in operational working environments, with the respect to the subjective evaluations. The Neurometrics Subsystem allows to evaluate the workers' mental workload, stress, and emotional state with a high temporal resolution, without interrupting the workers for completing any subjective questionnaires. In fact, during the In-Company phase it was demonstrated that the Neurometrics Subsystem is fully compatible with an online mental and emotional states evaluation on workers while they are performing ordinary activities in their real working environment.

2.9 WAOW tool measurements: Eye Tracking

2.9.1 Introduction

The Eye Tracking Subsystem is used for real-time evaluation of the mental strain of the user. The testing of the software and hardware and especially the algorithms for the calculation of the user´s mental strain status has already taken place in the in-lab studies. In the pilot studies, the applicability in the field is examined in particular. Among other things, the practical usability in relation to the concrete workstations is examined and a comparison of the measurement accuracy in the field with the measurements previously carried out in the laboratory is examined to be able to adjust the system if necessary.

The goal of the short-term field tests is therefore twofold:

- On the one hand, the practical application within the workplaces is investigated,
- On the other hand, the measuring accuracy is considered.

Figure 211: Use of the Eye Tracker at work

2.9.2 Method

Following the specifications given in D9.1 Evaluation protocol, we divided the evaluation in two parts: short term tests and long-term tests, whereas the first point is carried out at the premises of RWTH Aachen University and the second point is done within the WorkingAge long-term:

1. **Short-Term Tests:** Applicability and improvement of functionality when used at real workplaces.
2. **Long-Term Tests:**
   Usage within the WAOW tools long-term tests:
   a. A pre-selection of the workplaces and users is carried out remotely by telephone/mail, to find the most suitable usage context.
   b. A concrete workplace analysis is carried out on site before the eye tracking measurements are carried out within the WorkingAge long-term test. This is decided to increase the flexibility during the Covid-19 pandemic.
   c. The recorded measurement is checked for accuracy. For this purpose, sensor measurements will be compared to subjective measurements of the WAOW tool. Besides, influencing variables resulting from point 1 are set in relation to the measurements to identify possible interfering variables.

2.9.3 Results – Short Term Tests

2.9.3.1 Eye tracking subsystem tests

Using the eye-tracking subsystem, 13 measurements were recorded on four people at the office workplaces at RWTH Aachen University, with a total measurement duration of over 35 hours. The environmental conditions as well as the user requirements were tested using the D9.1 evaluation protocol.

The eye tracker was used during the measurements for as long as possible during each working day, downtime occurred when breaks were taken or meetings where participants could not or did not want to wear the eye tracker. Thus, the recording of the participants could be paused at any time while using the eye tracker. The recorded data from the eye tracker was saved and analysed. Every 2 hours, the subjective strain was recorded with the Nasa-TLX (Hart & Staveland, 1988) in order to compare it with the eye-tracking data.

We started to check the plausibility of data measures, following the D9.1 evaluation protocol. The criteria defined could be fulfilled for all measurements, in all cases the data are analysed by the eye-tracking subsystem as expected (Table 16).

Table 16: Plausibility check - short-term-test

Sensor	Pupil Core Pupil Labs Eye Tracker
Edge Server	RWTH Windows laptop
System checks	Not relevant
Subsequent evaluation	
Plausibility check of test data: • Pupil values between 1mm and 8mm • Confidence level >= 0.8 • Comparison of the strain levels with those of the in-lab study	✓
Plausibility check of measured data: • Pupil values between 1mm and 8mm • Confidence level >= 0.8	✓

The following Table 17 shows the mean values of all measurements performed. Four different users used the eye tracking subsystem during their normal working day for at least one day and up to 6 days. The length of the measurement varied from 33 minutes to 8:43 hours (M = 2:55:17 hr). As already described, all values are within the expected range (>1mm; <8 mm) of pupil size. The accuracy of the measurement (confidence level) is always above 0.8, as expected, on average with M = 0.981 (SD = 0.037) above the threshold.

Table 17: Measured pupil data - short-term-test

User	Date	Mean (Pupil diameter)		Mean (Confidence level)	
		Left	Right	Left	Right
1	Feb 04	2.833	3.290	0.991	0.996

2	Feb 15	2.558	2.865	0.959	0.961
2	Feb 16	2.737	2.812	0.950	0.940
2	April 26	2.944	2.977	0.958	0.922
3	April 26	3.111	2.959	0.996	0.998
1	May 05	2.067	2.708	0.993	0.996
1	May 06	1.964	2.565	0.986	0.998
1	May 09	2.303	3.368	0.987	0.995
1	May 12	2.688	3.406	0.983	0.994
4	May 12	3.395	3.322	0.988	0.995
4	May 13	3.145	2.816	0.988	0.997
1	May 16	2.763	3.414	0.995	0.997

On average, a total of M = 139.92 strain levels were calculated, for the most part the minute-by-minute calculation worked. User were asked to answer the same questionnaire as included in the WAOW tool, to assess the subjective strain level in order to be able to compare objective and subjective data. Figure 212 gives an overview of the comparison of objective strain levels, taken by the eye tracking subsystem and subjective strain level data, assessed via the questionnaire (the average strain level value calculated over the entire period of use for each user). The eye tracking subsystem outputs mental strain in three levels, from low mental strain to medium to high mental strain. The subjective evaluation is collected via the Nasa TLX questionnaire, where the total value is converted into a corresponding high level value as described in D3.2, where four levels of strain can be reached, starting with low strain, over medium-low strain to medium-high strain and high strain.

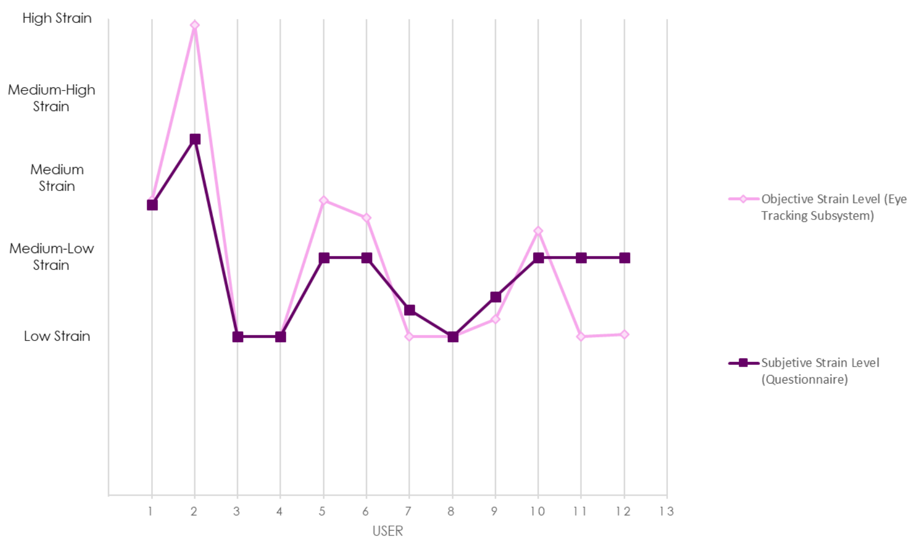


Figure 212: Comparison of strain level data

For a more detailed view, a full evaluation of one user is shown below (Figure 213). Starting at 10:45 am, the user used the eye tracking subsystem until 4:45 pm. The subsystem was not used continuously due to a lunch break and several meetings, so no pupil data was measured during this time and therefore no strain levels were calculated. Based on the pupil dilatation (upper figure), the strain level (middle figure) is calculated. In addition, the subjects were asked to indicate their subjective strain with the help of a questionnaire (bottom figure).

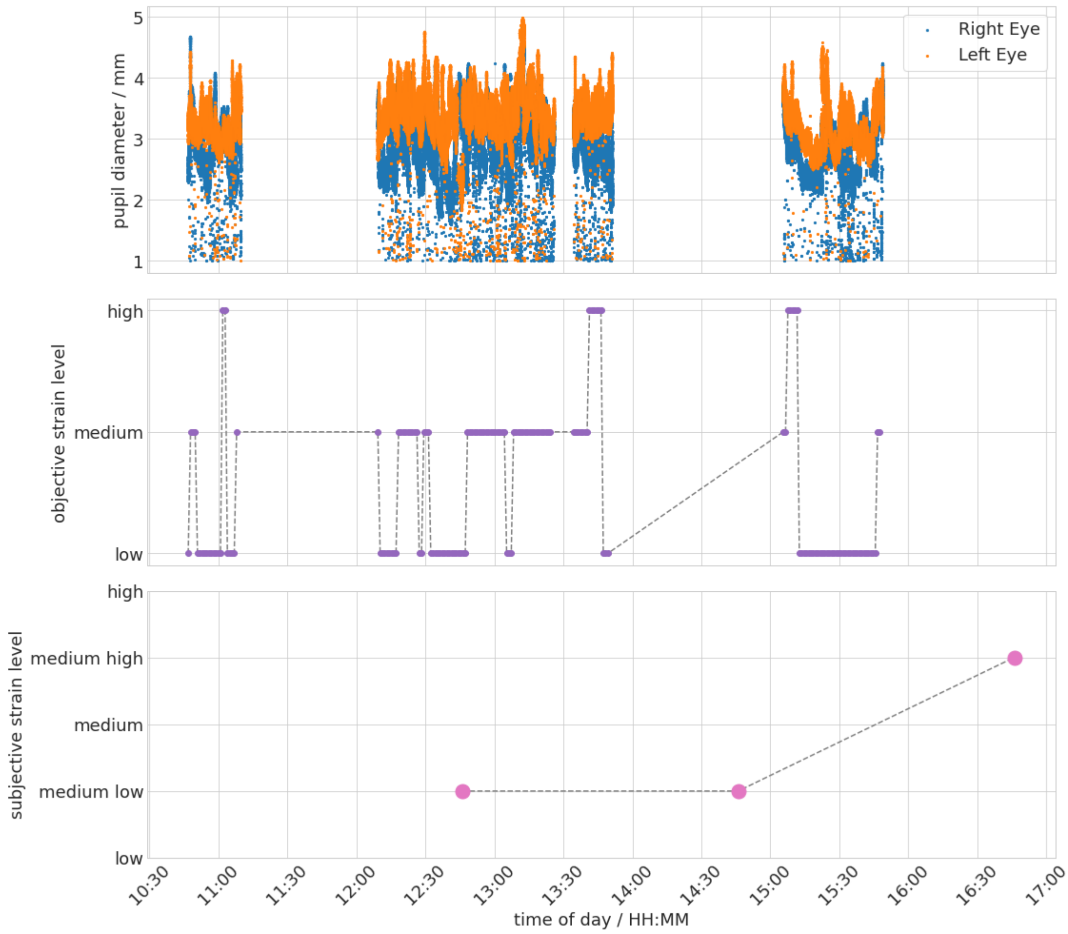


Figure 213: Evaluation of one exemplary user

2.9.3.2 Eye tracking subsystem acceptance tests

Due to the shortening of the pilot test within the WorkingAge project, the eye tracker could not be tested to a full extend in the use cases as planned. To ensure that full-scale tests could still be carried out, various acceptance tests based on the overall usability and acceptance assessment of the WAOW tool adapted for the eye tracking subsystem were carried out with the users of the short-term tests (description of usability & acceptance tests can be seen in chapter 2.4).

Participants seemed to agree that there were no ethical, legal or social aspects that needed to be considered when using the eye tracking subsystem. Users indicated that they did not feel as if they would lose contact to their colleagues as well that they do not see issues related to their data protection. However, with regard to safety in the workplace, users noted that the eye tracker's cable posed a safety risk.

The evaluation of the estimated benefit is highly rated, the users emphasise that they are often under pressure due to their office work and feel cognitively overloaded and therefore see the eye tracking subsystem as a useful support in their everyday work. The ease of use is rated better with further use. While the users initially felt overwhelmed and were not sure how to operate the system, they were able to learn quickly and carry it out independently. Despite this, the users admit that they would not want to wear such a system at the workplace all the time, but only in phases of high stress because the test persons find both the cable and the glasses themselves impractical for the long term usage. However, the users see a great improvement in the use of a remote eye tracker, which has no cable and is not worn as glasses, but has to be placed in front of the screen on the desk. This could be a method to overcome the disadvantages of the eye tracking subsystem and should be investigated in further studies.

When looking at the results of the eye tracking acceptance test, it should be noted that the participants were young and technically as well as scientifically experienced. A transferability of the results to the older target group must first be ensured. A full overview of the results is given in Table 18.

Table 18: Eye tracking subsystem acceptance ratings

		Evaluation	
		M	***SD***
TAM	Perceived Usefulness	3.750	0.500
	Perceived Ease of Use	3.750	0.500
	Use behaviour	2.250	0.500
TAM 2	Job relevance	4.250	0.500
TAM 3	Perceived enjoyment	2.000	0.000
	Anxiety	1.750	0.500
	Perception of external control	3.500	0.577
ELSI	Social Implications for work	1.250	0.500
	Legal implications (data protection)	3.750	0.500
	Legal implications (occupational safety)	3.250	0.500

2.9.4 Results – Long Term Tests

2.9.4.1 Eye-Tracking Subsystem – Evaluation of Strain Level

During the long-term tests, five users were measured at the MUTUA pilot site, three men and two women, all over 45 years old, all with brown eyes, working at an office workstation in an environment with uniform lighting conditions. Due to connectivity issues with the WAOW tool, the measurements of the first user have not been recorded. Therefore, the measurements of the following four users are listed below. The four users used the eye tracking subsystem during their normal working day for around 70 minutes. All values are within the expected range (>1mm; <8 mm) of pupil size. The accuracy of the measurement (confidence level) is always above 0.8, as expected, on average with M = 0.810 (SD = 0.009) above the threshold, however, lower than in the short-term tests.

Again, it becomes clear that objective and subjective data correspond; the mental strain assessed by the eye tracking subsystem is also recognized as such by the user. The eye tracking subsystem outputs mental strain in three levels, from low mental strain (1) to medium (2) and high mental strain (3). The subjective evaluation is collected via the Nasa TLX questionnaire, where the total value is converted into a corresponding high level value as described in D3.2, where four levels of strain can be reached, starting with low strain (1), over medium-low strain (1.5) to medium-high strain (2.5) and high strain (3). The following Table 19 gives an overview of the measured strain levels, averaged over the time of the measurement.

Table 19: Objective and subjective strain evaluation

	Objective Strain Level				Subjective Strain Level
User	via pupil diameter		confidence level		via questionnaire
	M	SD	M	SD	
2	1.667	0.951	0.804	0.006	2.5
3	1.000	0.000	0.803	0.013	1.5
4	1.000	0.000	0.823	0.037	1.5
5	1.523	0.888	0.810	0.016	2.5

The following Figure 214 shows the strain levels measured by the eye tracking subsystem over the time of the measurement. A low strain level corresponds to a value of 1, a medium strain level to a value of 2, and a high strain level to a value of 3. User 2 and User 3 show a high strain level at different times. However, for user 2 in all cases for less than 5 minutes, an intervention of the WAOW tool should not be triggered by this. User 5 shows a high strain level for more than 5 minutes in two cases, so an intervention should be triggered.

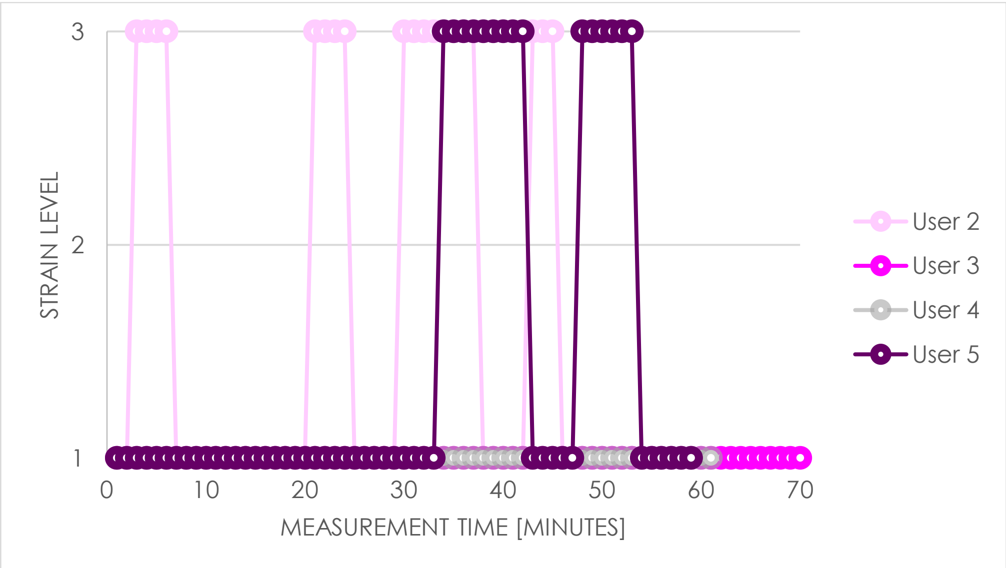


Figure 214: Plot of the calculated strain level of the eye tracking subsystem over the time of the measurements

2.9.5 Discussion and Outlook

The overall results can be considered satisfying, especially if taking into account, that age-related effects due to the Corona pandemic could not be considered in the laboratory study and thus could only be included in the pilot test (see also D9.3).

Nevertheless, the Short-Term tests show better results compared to the Long-Term tests in terms of the probability of the measured values, which might be related to the age-related changes of the pupil, this has to be investigated in further studies.

Short and long-term tests showed that the eye tracking subsystem is capable of measuring individual mental strain based on the pupil dilation and evaluating it in real time. From the collected subjective evaluations, we were also able to determine that, contrary to expectations, the use of the eye tracking subsystem causes fewer issues during actual use and is accepted by the users. The short-term tests, which could be carried out on a larger scale than initially planned due to the Corona pandemic as they were conducted in premises on side, made it clear that the users were very quickly able to use the eye tracking subsystem independently and without problems and required little to no support. Analysis of the data collected revealed an overall growing trend of user approval. A longer test period and more frequent use of the eye tracking subsystem contributed to the acceptance of the system and a lower assessment of the limitations for their own work.

However, there are some problems, especially with the fact that the eye tracker is connected to the laptop via a cable and that it is a pair of glasses that has to be worn. A solution here would be a remote eye tracker, so that the disadvantages of the subsystem could be compensated. Further studies are necessary for this.

2.10 WAOW tool measurements: Voice Analysis

2.10.1 Introduction

The voice analysis tool takes care of identifying the emotional status of the user from the speech (voice). Users wear Bluetooth microphones, the sensors, which are used to collect (record) audio clips. Audio clips are short (< 10 seconds) speech segments that are analysed by the voice analysis server. The voice analysis server uses a classifier (a machine learning model) to label (classify) the audio clips. There are two deployments of the voice server, one for Spanish (used in the Mutua pilots) and one for Greek (used in the EXUS pilots).

Following the results of in-lab tests, we are using an ensemble model composed by the combination of two classifiers provided by PoliMI and Audeering respectively. Moreover, after the results of in-lab tests, the model is used as a binary classifier to distinguish between "positive" and "negative" emotions. The original models distinguished among four emotions ("happiness", "anger", "sadness", and "neutral"), the labels were merged to improve the overall results ({“happiness”, “neutral”} become “positive” and {“anger” and “sadness”} become “negative”), for further details refer to D9.3, Section 5.4.

2.10.2 Method

Following the specifications given D9.1 Evaluation protocol, we divided the evaluation in three parts: short term tests and long-term tests.

During the short-term tests, we wanted to analyse to which extent the voice analysis module was capable of predicting the correct emotion in a real scenario. To do so, we prepared a subset of the audio clips used in the in-lab tests as references and we asked the users to take part in individual recording sessions based on those clips. The audio clips were examples of speech in different emotions (we used the four original emotion labels to label the clips, then we reduced to two classes during the evaluation):

- In EXUS we used 20 reference clips (5 for each of the four emotions).
- In Mutua we used 21 reference clips (5 for each emotion apart from “happiness”, which used 6; we did so because single happiness clips were shorter than others and we wanted the same number of recordings).

The users had to listen to the aforementioned reference audio clips and repeat the content trying to enact the emotion perceived in the clip itself. In this way we could have a ground truth to assess the quality of the model (similarly to what we did for the in-lab tests). In the EXUS pilots we managed to take part remotely to the recording sessions, to guide the users and keep track of the information about recorded clips and predicted emotions synchronously. In the Mutua pilots this remote supervision was not possible, so users were asked to keep track of the day and time they did the recording in order to help gather the information about recorded clips and predicted emotions asynchronously. In both cases users were asked to repeat the same clip twice (or thrice in some cases) in order to gather more evaluation samples. Once everything was collected, we computed some standard classification metrics using the predictions of the clips recorded by the users and the labels associated with the corresponding reference clip.

During the long-term tests, we wanted to verify that the voice analysis module was capable of managing the workload generated by the multiple users using the service at the same time and to collect the high-level information predicted by the voice analysis module in a longer period of time. In particular, the latter is useful to assess the overall effectiveness of the WAOW tool and for posterior simulations and analyses of the WAOW tool (data can be fed to the DSS, core of the WAOW tool, to simulate the users and see how the tool responds).

2.10.3 Results – Short Term Tests

The results of the short-term tests are in line with (or even better than, in some cases) what we observed during the laboratory tests. During the recording sessions we gathered samples from both the pilots the voice analysis tool was employed:

- In EXUS we collected 197 samples (49 of "happiness", 49 of "anger", 52 of "sadness", and 47 of "neutral") from 5 different users.
- In Mutua we collected 284 samples (82 of "happiness", 66 of "anger", 68 of "sadness", and 68 of "neutral") from 7 different users.

(Note that in some cases clips were not recorded correctly and couldn't be used in the evaluation, the reported numbers refer to the clips we managed to use).

As premised, to evaluate the models, we collected some standard metrics to evaluate binary classifiers goodness, in particular we computed Accuracy, Precision, Recall, F1 and AUC of the ROC curve. The first four metrics were computed after finding the optimal classification threshold the optimised the ROC curve. The computed scores were the following (Table 20):

Table 20: Computed scores for accuracy, precision, recall, F1 and AUC

	Accuracy	Precision	Recall	F1	AUC	Threshold
EXUS	61%	61%	61%	61%	64%	0.57
Mutua	64%	64%	64%	63%	65%	0.54

Additionally, in the last column, we provided the optimal classification threshold identified during the evaluation.

The overall results can be considered satisfying, each of the metrics reached a score > 60%. In particular the AUC of the ROC curve is a good indicator of the fact the model is behaving correctly on both positive and negative emotion classes.

Results in EXUS on Greek are better than those in the laboratory, this is probably due to the supervision in the recordings and/or the higher number of samples. Results in Mutua on Spanish instead are very close to those in the laboratory, this is probably due to the fact that clips were collected correctly despite the asynchronous approach.

2.10.4 Results – Long Term Tests

During the long-term tests, we observed that the system managed to keep up with the workload generated from all the users. By the end of the short-term tests, we managed to collect around 100k samples of high-level data (this means that the voice analysis tool labelled around 100k audio clips collected from the users through the microphones):

- In EXUS we collected 77828 data points, from 8 different users.
- In Mutua we collected 38835 data points, from 7 different users.

All high-level samples registered during the long-term test period have been organised into an Excel file (one sheet per pilot). The tables in each sheet have the following columns:

- Sensor group ID: The ID of the sensor sending the data to the voice analysis tool.
- User pseudo ID: The ID of the user associated to the sensor.
- Date: the date when the clip was recorded expressed as yyyy-mm-dd.
- Time: the time of the day the clip was recorded expressed as hh:mm:ss.
- Value: the emotion label ("Positive" or "Negative") associated to the audio clip by the voice analysis tool.
- Probability: the probability score ($p \in [0,1] \subseteq \mathbb{R}$) associated to the label predicted by the voice analysis tool.

In the following Figure 215, we report a plot of the cumulative sentiment detected during the long-term tests.

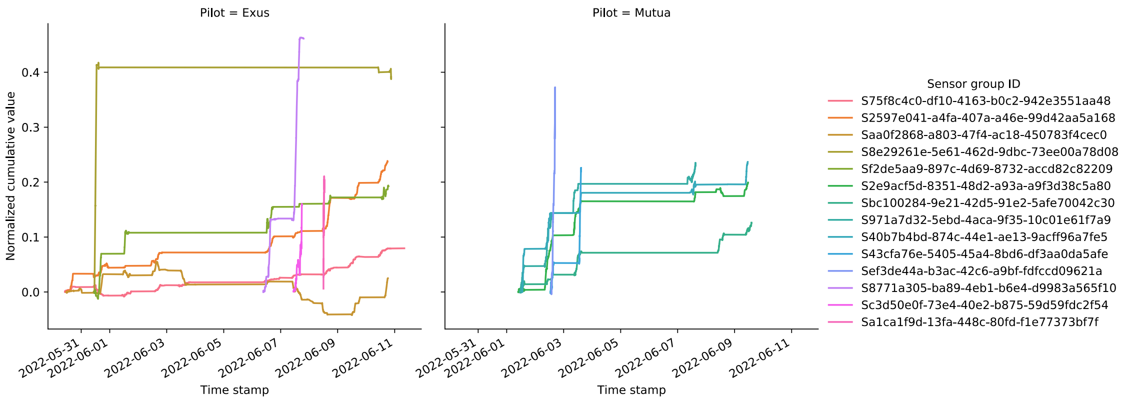


Figure 215: Plot of cumulative sentiment detected during long term tests

In the plots we associated a value of 1 to "positive" sentiment and a value of -1 to "negative sentiment", we normalised the scores on the number of data points (per user) and we computed the cumulative sentiment during the long term tests time lapse (per user). Despite we would need a longer testing period and a more frequent use of the voice analysis tool for a confident claim, we can see a growing trend for almost all users, meaning that in the time lapse of the high level data collection the number of registered positive sentiments is higher than that of negative sentiments (again, per user).

2.10.5 Summary and Outlook

The overall results can be considered satisfying, especially if taking into account that the languages the voice analysis tool worked on (Spanish and Greek) are considered under resourced (especially taking into account that in MUTUA two languages are spoken: Spanish and Catalan). In fact, languages like English present way more resources (like labelled data sets) that allow to train more robust models.

The short terms tests showed not only good results in both pilots, but the results were either in line with the in-lab ones (like in Mutua for Spanish) or better (like in EXUS for Greek). As a side note we point out that the users were not professional speakers or actors and their ability to enact emotions may be limited, thus the actual performances of the model may be better when the voice analysis tool is used over real (in the sense of not enacted) samples.

The long-term tests showed that the model is capable of sustaining the workload generated by multiple users. At the end of the tests, we collected many high-level data points across multiple days and from multiple users that will help simulations to further assess the overall WAOW tool quality. From the collected data analysis, we noticed an overall growing trend of positive sentiment in the users, a longer test time and the more frequent use of the voice analysis tool from the user will help get better insights of the voice analysis tool effectiveness. These results need to be confirmed also through analysis of other sensors and the responses to the questionnaires.

2.11WAOW tool measurements: NoiseBox

2.11.1 Introduction

The NoiseBox has been built by audEERING to monitor the surrounding acoustics. It measures noise level and segments voiced part of the acoustics and sends the voiced parts to the voice-analysis server. It is built based on a raspberry-pi hardware, and connects to microphone. The NoiseBox has been used in the short- and long-term tests.

2.11.2 Method

In the NoiseBox a machine-leaning model is integrated to recognize the Voice Activity, Voice Overlap (i.e., two or more people talk at the same time). The model is based on Long-Short-Term Memory Neural Network (LSTM-NN) and has the F1-measure of 0.94.

Beside Voice Activity Detection and Voice Overlap Detection, the level of noise is also computed and captured as Low (<60dB), Medium (60dB-80dB) and High (>80dB).

2.11.3 Results – Short-term Tests

During the short-term tests, NoiseBox was used to capture the voiced segments and send the segments to the voice-server for the analysis mentioned in the previous section. We monitored carefully the segments, and the NoiseBox was able to ignore unvoiced segments, to capture all voiced segments and to send them to the voice-server.

In parallel, also the level of noise was monitored and the NoiseBox result were perfectly matching to the noise level during the experiment (e.g., once the participant was talking, the noise level was changing from Low to Medium or High).

2.11.4 Results – Long-term Tests

During the long-term tests, the NoiseBox was performing its tasks to segment voice parts of acoustic signals, send them to the server, and in parallel monitor the noise level in the environment.

As can be seen in Figure 216, the NoiseBox could successfully capture more than 115K voice segments and send them to the voice server (for emotion analysis) in the EXUS and Mutua sites.

Moreover, the noise level values could be captured successfully during the test. Figure 216 depicts an example of the noise values for one participant in one of the sites during about 18 hours of operation.

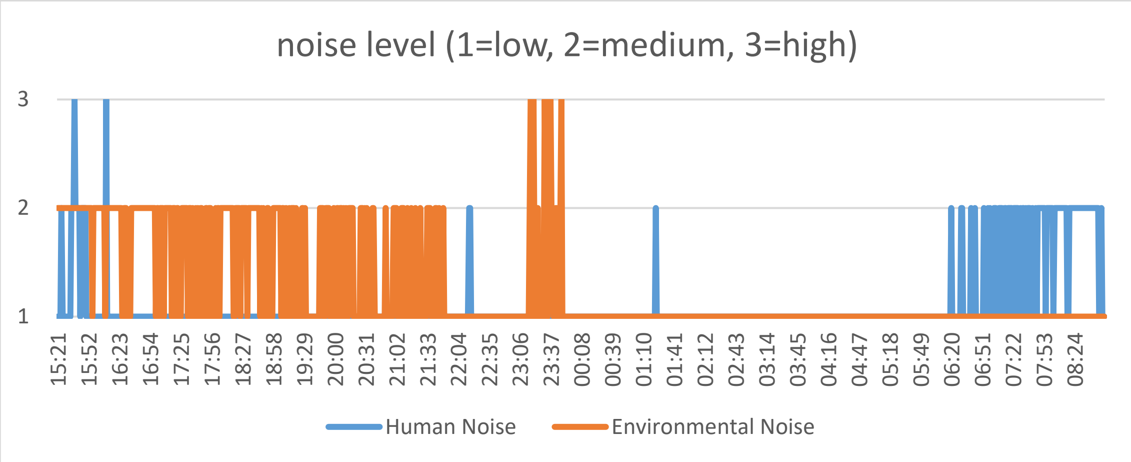


Figure 216: Noise levels during about 18 hours of operation of one of the NoiseBox in EXUS site.

2.11.5 Summary and Outlook

AUD developed 40 NoiseBoxes from scratch (design, 3d-printing, montage, software installation, programming, test, automatic deployment) for the WorkingAge project and deployed them on the three sites. There was some hardware failure (broken SD-Card, operating-system failure) during both development and short- and long-term testing. However, we could manage to fix the problems in-time to fulfil the project goals. In the WorkingAge project, the NoiseBox could provide a possibility to monitor user's emotion during work environment. Although for this project NoiseBoxes were developed to monitor noise and voice, their functionality can be extended to embed other machine learning models such as anomaly detection or security and surveillance.

2.12 WAOW tool measurements: Activity Tracker

2.12.1 Introduction

The Activity Tracker is a commercial device, a smartband (MiBand 4), this sensor provides the app with information on physical activity, steps and distance walked, calories burned, heart rate and hours of sleep.

2.12.2 Method

The following method to check the operation of the activity tracker was defined in D9.1:

- Put the smartband on your wrist
- Connect the sensor to the app via Bluetooth.
- Configure app to read sensor values.
- Check that the app receives physical activity, steps and distance walked, calories burnt, heart rate and hours of sleep values (Some data such as hours of sleep or distance travelled need to use the smartband for a while to collect data). Checking it is simple since all these data will appear in their respective cards in the app (steps, sleep and heart rate).

2.12.3 Results – Short Term Tests

The activity trackers were tested together with the app following the method specified in D9.1, and all of them worked correctly, they showed the physical activity in the device itself and in the WAOW Tool app, so the Bluetooth communication worked perfectly and it was verified in the log that the data about physical activity, steps and distance walked, calories burned, heart rate and hours of sleep arrived at the DSS.

2.12.4 Results – Long Term Tests

During the long-term tests, DSS decisions and user notifications related to physical activity have also been monitored to help determine whether data sent from the sensor to the WAOW app were correctly relayed to the WAOW tool (Table 21).

Table 21: DSS decisions and user notifications related to physical activity

Pilot	Type of pilot	Users Registered	Sensors with data in the DSS logs	Percentage
Grupo Antolín	Production	13	11	84
Mutua Universal	Office	14	12	100
EXUS	Telework	9	7	77

These percentages are from demonstrable data, in reality if we check them with the number of backups that were recovered from the apps we have a 100% success rate, that is, the margin of error is due to errors in the app, or communication with the server, but the activity tracker worked properly with the app in all cases since users saw their physical activity data on their mobiles.

The DSS records were checked, and it was found that the activity trackers had been sending information throughout the pilot and that the DSS released several advice or goals achieved.

2.12.5 Summary and Outlook

The activity tracker is a commercial device, on which we use reverse engineering to read the data it collects, so it was expected that unless there was a faulty device, it should not give any error. All users were able to pair them with the WAOW tool app (some did it with help). There was only one complaint from users, the app was configured in such a way that it shows the hours slept during the 24 hours of a specific day and several users said that it would make more sense to show what is slept in a night even if the hours belong to two different days, this does not affect the decision making of the DSS at all since it is only the way in which the data was displayed to the user, but we will take it into account if we develop something similar again.

2.13 WAOW tool measurements: Scale

2.13.1 Introduction

The Scale is a commercial device “Xiaomi Mi Body Composition Scale 2” this sensor provides the app with information on user weight and BMI. The sensor connects to the WAOW Tool app and the phone via Bluetooth.

2.13.2 Method

The following method to check the operation of the activity tracker was defined in D9.1:

- Connect the sensor to the app via Bluetooth.
- Configure app to read sensor values.
- Get on the scale barefoot

Check that the app receives weight and BMI. Checking it is simple since all these data will appear in the card “weight” in the app.

2.13.3 Results – Short Term Tests

The scales were tested together with the app following the method specified in D9.1, and all of them worked correctly, they showed the user's weight and BMI in the device itself (just the weight) and in the WAOW Tool app, so the bluetooth communication worked perfectly and it was verified in the log that that the data about weight and BMI arrived at the DSS.

2.13.4 Results – Long Term Tests

During the long-term tests, DSS decisions and user notifications related to scale data have also been monitored to help determine whether data sent from the sensor to the WAOW app were correctly relayed to the WAOW tool.

Table 22: DSS decisions and user notifications related to scale data

Pilot	Type of pilot	Users Registered	Sensors with data in the DSS logs	Percentage
Grupo Antolín	Production	13	9	69

Mutua Universal	Office	14	0	0
EXUS	Telework	9	3	33

The percentages in this sensor are a bit low, in fact in the Mutua pilot no data was collected, this was not a sensor failure, the scale is a commercial device that the volunteers had in their homes, it has its own official app, many users used this official app instead of the WAOW Tool app, others did not pair the scale to the app (by having to do it at home or they did not know or simply did not do it), in addition, the scale requires the user to use it every so often, they were asked to use it at least once a week, but some didn't use it at all.

For users who followed the instructions for use, it worked perfectly and communicated and interacted with the DSS.

2.13.5 Summary and Outlook

Although the sensor worked correctly and fulfilled its purpose, the fact that a user must make the decision to use it and requires an action on their part, instead of something automatic like a wearable, complicates continued use over time as it has been demonstrated during the long-term tests. In future projects or studies this conclusion should be taken into account.

2.14 WAOW tool measurements: Environmental Sensor

2.14.1 Introduction

The Environmental Sensor is a custom device specifically developed for the WA project that monitors four environmental variables: illumination (lux), temperature (°C), relative humidity (%), and CO2 concentration (ppm). It has been designed as a non-invasive, unattended, and transparent device for the user, who only has to connect the sensor with the WAOW tool (Smartphone) via Bluetooth and charge the battery whenever its level is low.

This sensor generates information at two levels: the low-level information, consisting in the conditions and pre-processed numerical values of the environmental variables, and the high-level information consisting in warnings and alarms triggered by the analysis of the low-level information, according to the evolution of the numerical values over predefined and configurable periods of time.

2.14.2 Method

Deliverable D9.1 Evaluation protocol establishes a series of metrics and Key Performance Indicators for the testing of reliability of components, including the sensors and, therefore, the Environmental Sensor. Since no additional validation or calibration processes are applicable to this sensor, the following test method was devised and has been applied to the short-term tests.

1) Test procedure

- Connect the sensor to the app via Bluetooth.
- Configure the app to read the sensor values.
- Check that the app receives Illumination, Temperature, Relative Humidity, and CO_2 values.
- Check that none of the received values is abnormal or non-plausible:

Parameter	**Invalid value criteria**
Temperature	Less than 0 or greater than 50
Humidity	Less than 0 or greater than 100
Lux	Less than 0 or greater than 6553
CO_2	Less or equal to 0 or greater than 5000

2) Expected results

The sensor device must connect to the app and receive Illumination, Temperature, Relative Humidity, and CO_2 data.

3) Addressing issues

In case the sensor is not connecting to the smartphone or values are not within normal ranges, the devices must be connected to a USB charger to ensure the issue is not related to low battery. If the problem persists or read sensor values are abnormal, the device must be replaced with another unit and sent to ITCL for diagnostic tests.

Also, the following KPIs have been used to validate the installation and operation of each Environmental Sensor Deployed in the short-term and long-term tests, as described in the System Setup Instructions:

KPI #33: The environmental sensor is installed.

KPI #34: The environmental sensor is ON.

KPI #35: The environmental sensor is connected to the WAOW tool.

KPI #36: The environmental sensor is sending data to the WAOW tool.

It is important to remark that, in the WAOW ecosystem, the Environmental Sensor does not provide the user with direct information or guidelines. Neither does the user have access to the low- or high-level information that the sensor produces through any graphical user interface (GUI). Instead, the data, both high and low level, gathered by the sensor are relayed via the WAOW app to the DSS to be incorporated into the decision-making process as part of the evaluable inputs and the guidelines and warnings that are issued to the user.

With these design constraints in mind, two validation mechanisms were implemented in the WAOW system. The first one enables the display of environmental sensor logs in a dedicated sub-menu of the app, where low-level data gathered by the sensor can be checked. These allows researchers and users to check whether the sensors are working properly, and even diagnose possible malfunction of some of its sensing components (e.g., if values are not consistent with actual environmental conditions). The second one regards the high level information produced by the DSS, that logs information regarding the environmental conditions, so that researches can check if the DSS is receiving information from the Environmental Sensor and if this information is being processed properly.

2.14.3 Results – Short Term Tests

All Environmental Sensors passed the test method described above, rendering the expected results: they connected to the app and sent Illumination, Temperature, Relative Humidity, and CO_2 data.

A total of 39 sensors were made available for the pilots, XXX of them were deployed as detailed in Table 23 below:

Table 23: Available sensors for the pilots

Pilot	Type of pilot	Available Sensors	Deployed Sensors
Grupo Antolín	Production	13	13
Mutua Universal	Office	15	11
EXUS	Telework	11	11

KPI #33: The environmental sensor is installed: a total of 34 Environmental Sensor units were successfully deployed.

KPI #34: The environmental sensor is ON: all units deployed turned on. One unit suffered damage due to misuse of the charging port in the Mutua pilot and had to be replaced by a new one, solving the issue.

KPI #35: The environmental sensor is connected to the WAOW tool. 100% of deployed environmental sensors connected to the WAOW App successfully.

KPI #36: The environmental sensor is sending data to the WAOW tool. 100% of the deployed environmental sensor sent data to the WAOW App successfully.

2.14.4 Results – Long Term Tests

The test procedure described in D9.1 has been applied to evaluate the sensors performance during the log-term tests, as well. Results were similar to the ones observed during the short-term tests:

KPI #33: The environmental sensor is installed: the 34 Environmental Sensor units deployed during the short-term tests continued to be used by the volunteers during the long-term tests.

KPI #34: The environmental sensor is ON: all units deployed turned on. Another unit suffered damage due to misuse in the Grupo Antolin pilot. The sensor was inadvertently dropped on the floor and the charging port was damaged. Replacing it by a spare one solved the issue.

KPI #35: The environmental sensor is connected to the WAOW tool. 100% of deployed environmental sensors connected to the WAOW App successfully. No Bluetooth range or communication issues have been identified.

KPI #36: The environmental sensor is sending data to the WAOW tool. 100% of the deployed environmental sensor sent data to the WAOW App successfully.

According to these data, four different performance categories have been evaluated: Installation (KPI 33), Power supply (KPI 34), Bluetooth Connection (KPI 35), and Data sending (KPI 36). Only two of the sensors (5.55% of the total) experienced malfunction during the tests, both short and long term, and in both cases, it was due to misuse. In both cases, the malfunction was associated to the Power supply category, and they had to be replaced to solve the issue, since the damage was beyond repair. This situation triggered a hardware design analysis to determine whether the design of the sensor or the enclosure could be optimised to reinforce a possible weak spot. As a conclusion, the footprint of the USB charging connector has been modified to provide enhanced mechanical resistance to shocks and prevent this situation in the sensors that may be produced from now on.

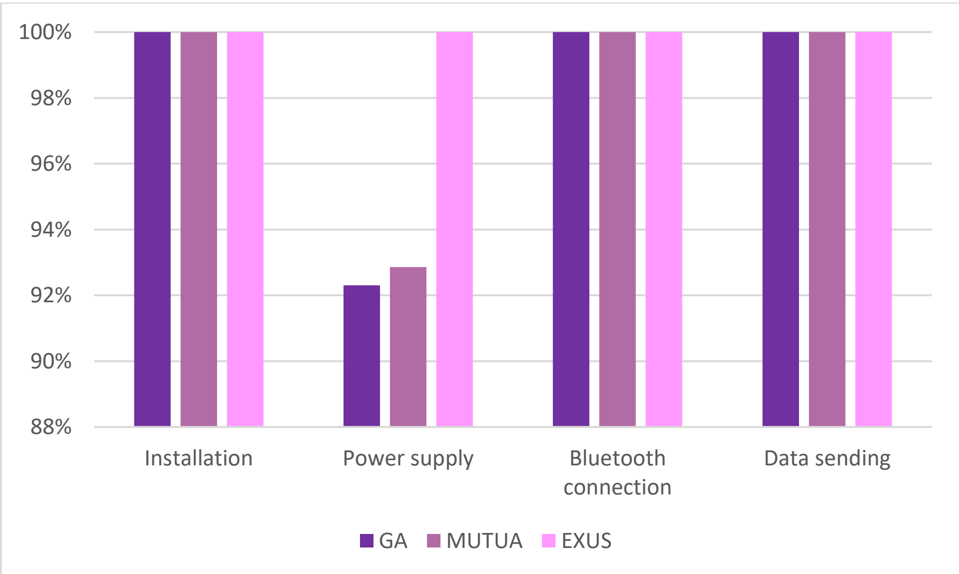


Figure 217: Percentage of successful performance of environmental sensors per pilot

During the long-term tests, DSS decisions and user notifications related to environmental conditions have also been monitored to help determine whether data sent from the sensor to the WAOW app were correctly relayed to the WAOW tool (Table 24).

Table 24: DSS decisions and user notifications related to environmental conditions

Pilot	Type of pilot	Users Registered	Sensors with data in the DSS logs	Percentage
Grupo Antolín	Production	13	9	69
Mutua Universal	Office	14	11	92
EXUS	Telework	9	7	78

These numbers show that 80% of the sensors had their information successfully relayed to the DSS by the WAOW system to be incorporated into the decision-making process and the generation of recommendations to the users even though 100% of the deployed sensors were able to connect to the smartphone and send information to the WAOW App. It must be noted, however, that the process delivering data from the environmental sensor to the DSS is totally independent from the sensor, the only communication process in which the sensor is involved is the sending of data to the app via Bluetooth. Lack of data in the DSS could be caused by unexpected events during the pilots such as incorrect user registration, sensors not being turned on or charged, sensors being disconnected from the Smartphone, network issues, etc. so this is not a factor considered relevant for the validation of the Environmental Sensor.

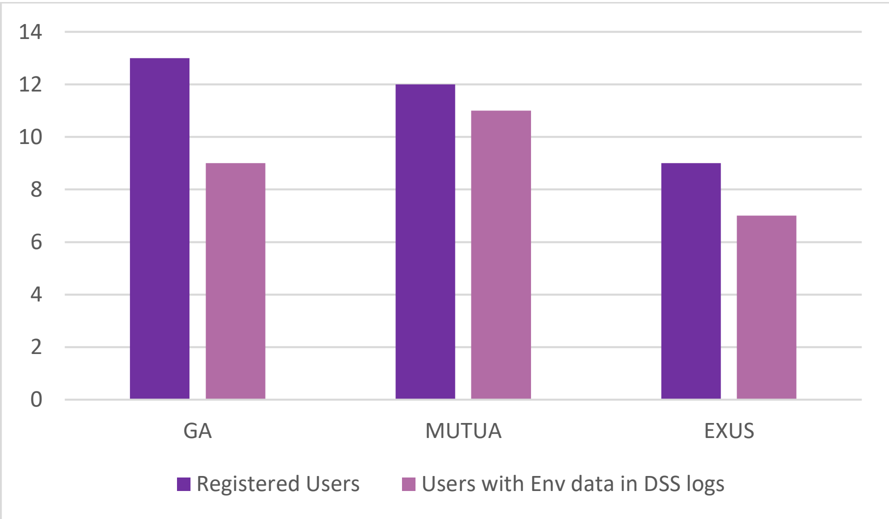


Figure 218: Registered users vs. environmental data per user in DSS logs.

2.14.5 Summary and Outlook

The overall results of the evaluation of the Environmental Sensor performance during the pilots are satisfying, having met the expectations and the criteria established for validation. 100% of the sensors deployed operated as expected, and only two of them had to be replaced due to accidental situations or misuse.

Even so, the replacement units worked perfectly fine and solved the issues. The Environmental Sensor has proved to be an unintrusive, easy to use device that requires little intervention from the user. Any issues with relaying data to the WAOW ecosystem that may have been identified are independent of the sensor design or performance so the viability of the sensor as a stand-alone environmental measuring unit has also been validated.

The deployment of sensors in pilot environments has also produced valuable information on the performance of the Environmental Sensor, which has already led to the optimisation of the technical requirements specification, functionality, and hardware and firmware designs, focusing mainly on the mechanical resistance aspect of the device.

The Environmental Sensor is now ready to be produced in short to medium series and to be deployed in third party environments where it can be easily integrated through its Bluetooth connection and its communication protocol. It can be integrated with little to no adaptations required in a wide range of applications, thanks to its comprehensive, flexible control and communication protocol and it non-invasive nature. Additionally, due to the impact of the Covid-19 pandemic all over the world, workplace environmental metrics, and especially CO2 concentration, have become crucial. CO2 concentration can be used to determine when ventilation is required or if it has occurred. Ventilation being one of the major factors to avoid the transmission of airborne pathogens in indoor spaces, this implies that devices such as the Environmental Sensor can play a key role, extending their use from workplace comfort monitoring to health safety monitoring.

2.15 WAOW tool measurements: Questionnaires

2.15.1 Sleep Quality

2.15.1.1 Introduction

As described in D3.1, the participants, were asked at the beginning of the day using a smiley scale based in an item from the SF-36 questionnaire. Figure 219 shows the four smileys faces, that rate the sleep from Very good to Very bad.

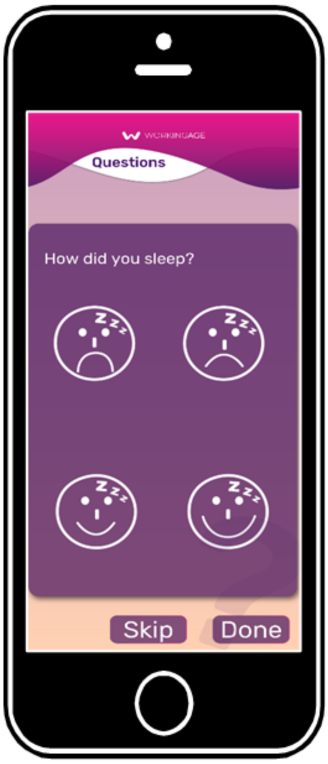


Figure 219: Sleep quality smiley scale. SF-36 item.

2.15.1.2 Results

Unfortunately, the volunteers did not fill in the questionnaires regularly, so the data received is not enough to be shown in scientifically valid way.

Below we make a representation of the few answers received by the participants.

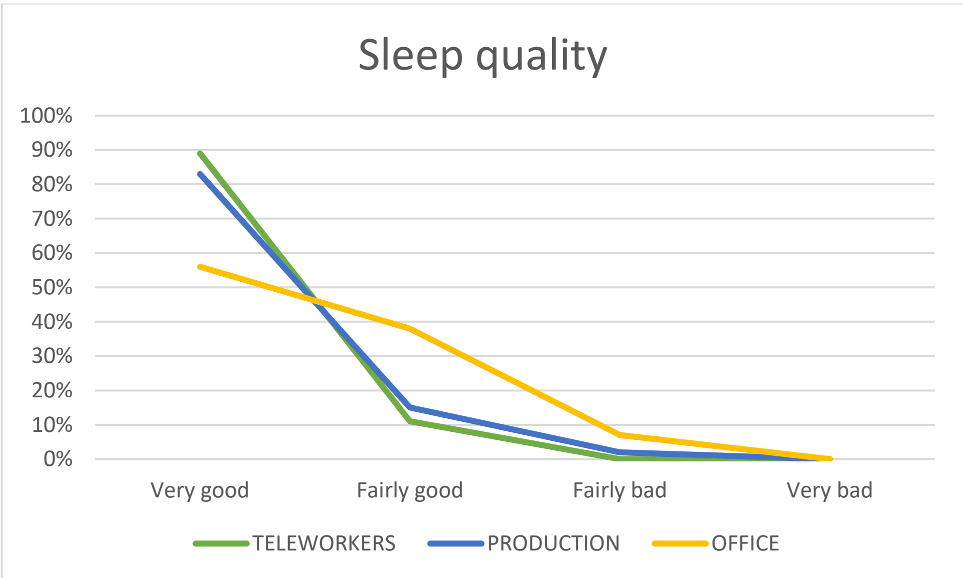


Figure 220: Sleep quality

2.15.2 Mood

2.15.2.1 Introduction

As described in D3.2 the overall mood of each participant was assessed using a smiley scale to be answered once a day by all users. The scale used is based on the five-level smiley scale of Jäger (2004) ranging from 1 "bad" to 5 "very good" mood.

The Smiley Scale is used for subjective assessment emotional aspects. The users of the WAOW tool are asked to rate their mood once a day. The questionnaire is integrated into the WAOW tool and can be answered on the provided smartphone. The representation is realized as shown in Figure 221. The Smileys are displayed and by selecting one the current mood can be entered. Users will be reminded every morning to reply to the questionnaire by means of a notification, the reply is voluntary; it can be skipped if desired.

Figure 221: Smiley Scale following Jäger (2004).

2.15.2.2 Results

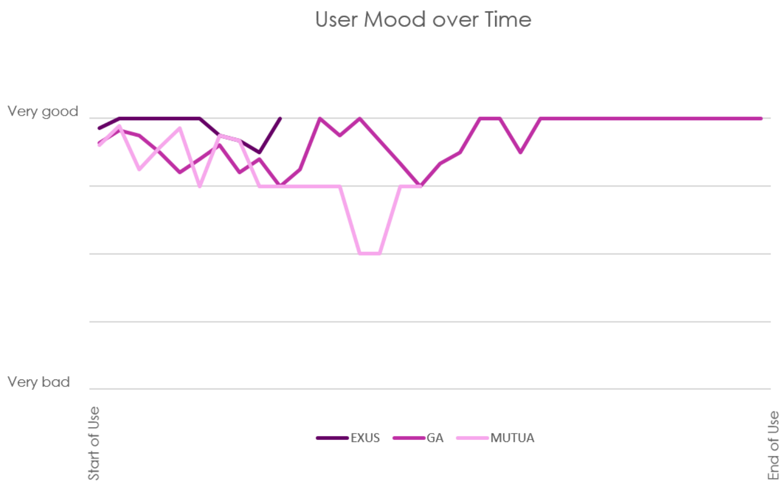


Figure 222 shows user mood over time, as indicated by the user via the WAOW Tool App. Most users indicated that they were in a “good” or “very good” mood. As can be seen in the Figure, there was one MUTUA Employee who indicated that his mood was “neutral” at the beginning of his workday. Unfortunately, the test persons did not fill in the questionnaires regularly, which is why a development related to tool use cannot be calculated scientifically valid.

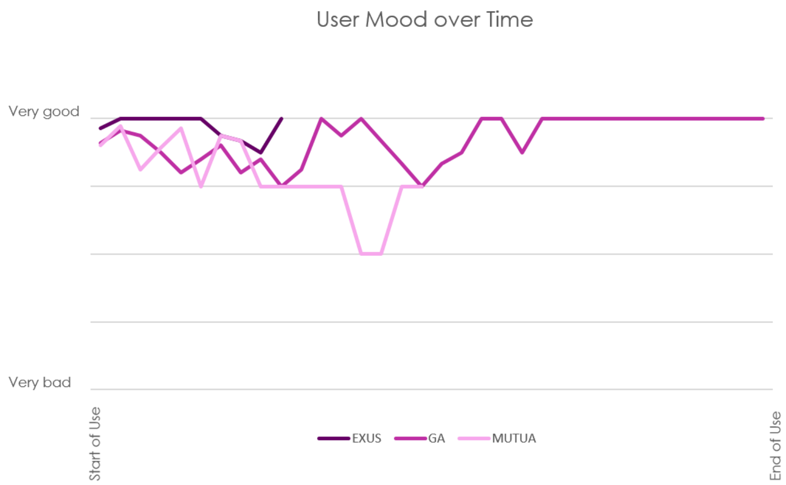


Figure 222: Average user mood over time of using the WAOW Tool

2.15.3 Work Ability

2.15.3.1 Introduction

As described in D3.1 the work ability of each participant was assessed using three questions based from the Work Ability Index (WAI) (2007) items WAI1 and WAI2. These questions will be used to assess the progression of self-perceived ability to work. The WAI1 item will be asked to the user once a month to get a general follow-up measurement, while the two questions of item WA2 will be asked once at the end of each working day.

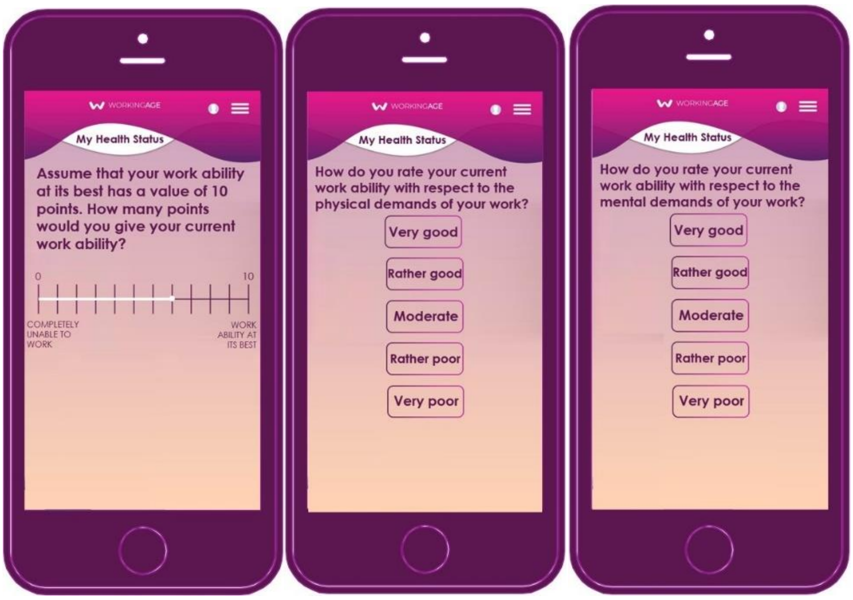


Figure 223: Questions to assess the subjective work ability of the user.

The questions are integrated into the WAOW tool and can be answered on the provided smartphone. The users will find these questions with the frequency explained before. The reply is not mandatory, they can answer later or skip the questions if they want to.

2.15.3.3 Results

Again, we didn't find enough data from the participants, to show any relevant result.

2.15.4 Subjective strain assessment

2.15.4.1 Introduction

As described in D3.2 the NASA TLX questionnaire was used for subjective assessment of mental and emotional work-related aspects. The users of the WAOW tool were asked to rate their working day once a day, always at the end of their working day, using the scales of the NASA TLX. This questionnaire consists of six sub-scales representing independent groups of variables: Mental, physical and temporal demands, frustration, effort and performance. It can be assumed that a combination of these dimensions represents the "workload" that most people experience when performing most tasks. These dimensions have been selected after a comprehensive analysis of the primary factors that define the subjective experience of workload for different individuals performing a variety of activities ranging from simple laboratory tasks to flying an airplane (Hart, 2006). The questionnaire is integrated into the WAOW tool and can be answered daily on the provided smartphone. The representation is realized as shown in Figure 224. This means that the single scales are displayed one after another and a slider is provided for answering. A more detailed description of the scales can be called up on request. Users will be reminded (at the end of the working day) to reply to the questionnaire by means of a notification, the reply is voluntary; it can be skipped if desired.

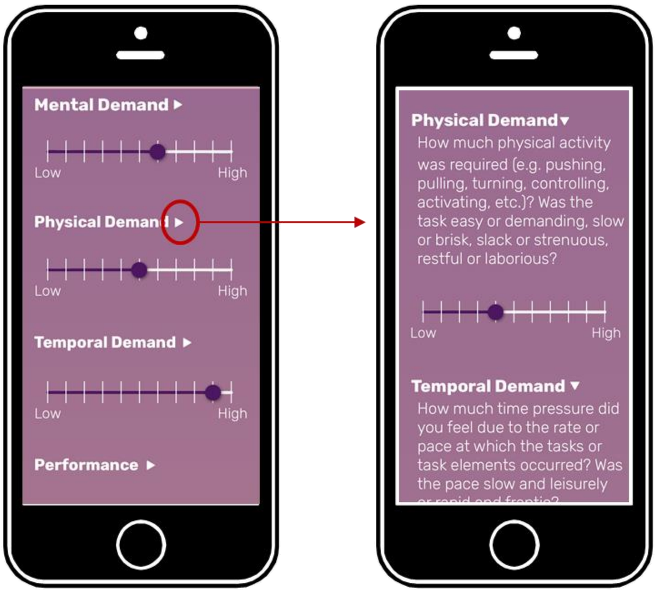


Figure 224: Implementation of NASA TLX in the WorkingAge App

2.15.4.2 Results

Users were asked to rate their working day based on six subsections from the NASA TLX. Five of the six subsections were scored on a scale from Low to High, whilst the subsection Performance was scored on a scale from Good to Poor. Figure 225 shows the average user scores for the six subsections of the NASA TLX. Grupo Antolín (GA) rated themselves highest in all six subsections stating that they had high Mental-, Physical- and Temporal-demands, as well as high Effort and Frustration, but good Performances. EXUS and MUTUA seemed to rate themselves pretty similarly, except for the sections Frustration, Effort and Temporal Demands, all in which EXUS rated themselves higher.

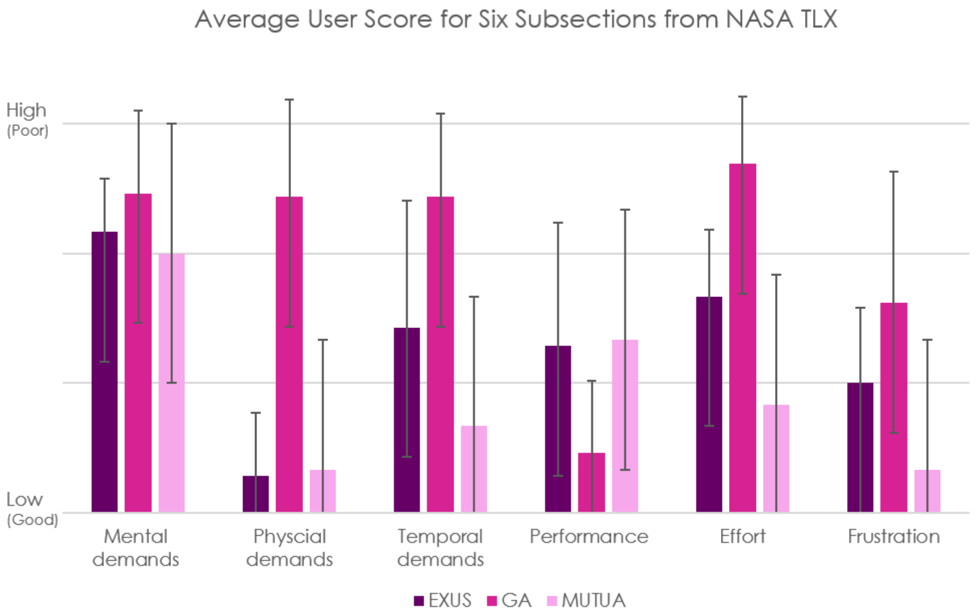


Figure 225: Average user score of the six subsections from the NASA TLX

2.15.5 (Physical) Health

2.15.5.1 Introduction

As described in D3.1, to monitor the self-perceived general health status, assess its evolution throughout the phase of use of the WA technology and assess the correlation of such health state with the compliance and adherence to the intervention plan provided by the system, once a week the user will be asked to rate his current health through the following item of the SF-36 questionnaire:

Figure 226: SF-36 item (1992)

2.15.5.2 Results

Due to the lack of answers from the volunteers of the three use cases, we didn't get enough data to show measurable results.

2.15.6 Social Relations

2.15.6.1 Introduction

As described in D3.3, the volunteers were asked weekly about their social relationships, using two items from the SF-36 questionnaire. The below Figures show the screen that the users had to reply.

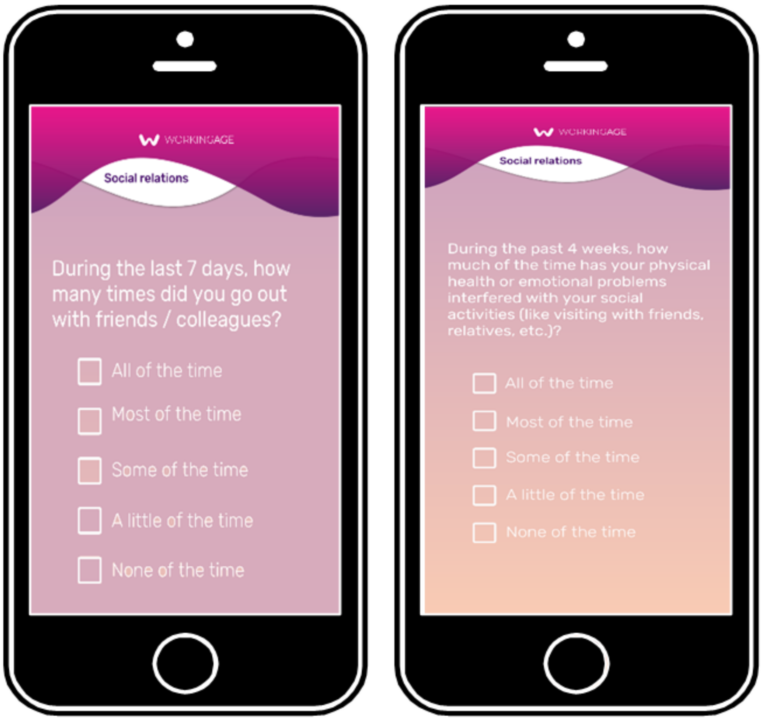


Figure 227: SF-36 items (1992)

2.15.6.2 Results

As in the previous questionnaires we didn't receive enough data to show conclusive results.

2.15.7 Nutritional Habits

2.15.7.1 Introduction

The nutritional habits were assessed in the pre and post questionnaires using the questions from "How healthy is your diet?" questionnaire. This is explained with detail in D9.1.

The only question included in the WAOW tool, was about water intake, sent to the users every day.

Figure 228: Water intake screen

2.15.7.1 Results

Unfortunately, the volunteers didn't give enough answers to this question to present valid results.

2.16 Wi-Fi Network

This section reports the performances of the WAOW Tool edge network deployed on the three pilot sites. The performances KPIs chosen to assess the network are the network size, traffic, and distance crossed by the information in the network. Thus, taking in account each pilot's specific settings. We also investigate the issues we met during the deployment and the mitigation procedures, and we propose improvements for next version of the WAOW Tool or any similar product.

2.16.1 Use Case: Teleworking (EXUS)

EXUS was hosting the Teleworking Pilot in which the WAOW Tool users were located at home. The challenge was to provide the WAOW Tool monitoring services at multiple locations while ensuring same level of security and privacy at both home and office.

We choose to replicate the office's IT environment at each home and connect both site via secure VPN. Then the office's servers could collect teleworker's data directly, without any intermediary cloud and in a secure manner.

To do so, we deployed 11 WAOW Tool routers at each home and at the office (Figure 229). The routers connect together and form a mesh network. Routers that are located nearby, at the office, connect together wirelessly via Wi-Fi (red link between YOI-RELAY-1 and YOI-VPN). Remote routers, the teleworkers one, connect to the office via Internet and secure VPNs (black link between YOI-TELEWORKER-X and YOI-VPN). Thus, the servers located at the office (e. g. the zeromqproxy, waexus-server and voice servers) could connect and collect data generated by Teleworkers' IoT devices (e. g. camera1-2 of YOI-TELEWORKER-2) directly, without going through any third party's cloud.

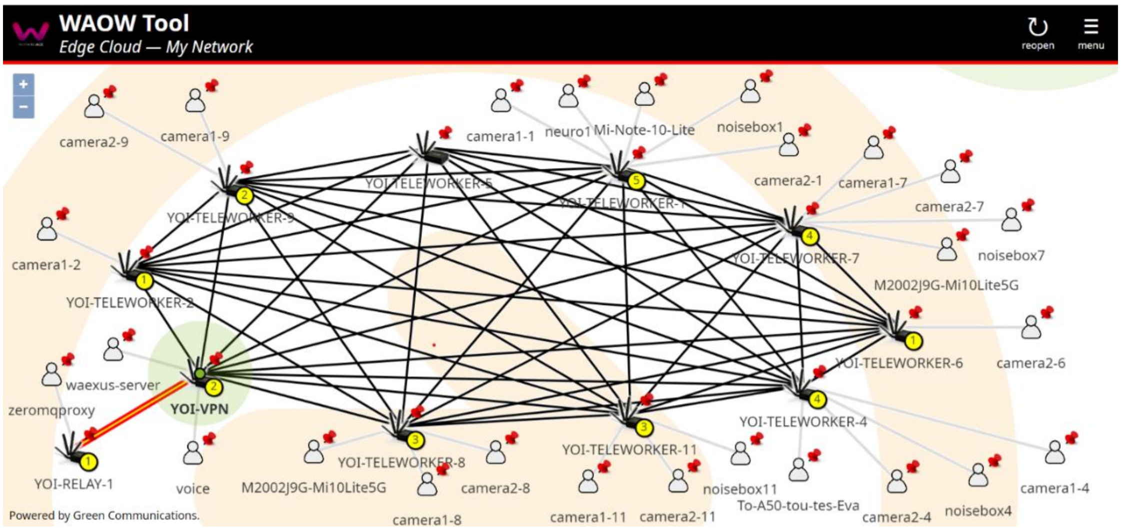


Figure 229: EXUS network view

2.16.1.1 Network Size

The network size reached 12 network nodes at peak for connecting 10 teleworkers and the office. There were 26 IoT devices connected when fetching Figure 229 out of which 3 edge servers, 14 cameras, 4 noiseboxes, and 5 other devices (e.g. smart phones, Mi-Band, etc.).

2.16.1.2 Traffic analysis

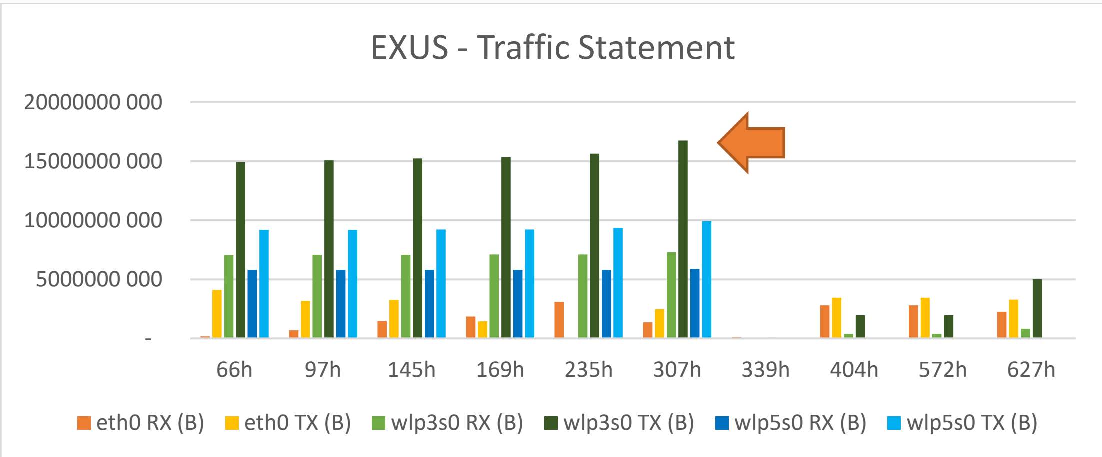


Figure 230: EXUS traffic analysis

We registered 16GB Output traffic at peak (307h in Figure 230) on the wlp3s0 interface of the YOI-VPN network node.

wlp3s0 is the wireless backhaul interface of the YOI-VPN that is involved in the transfer of data between teleworkers and the ZeroMQ server (zeromqproxy in Figure 161) and the transfer of data between Voice and EXUS's servers with the ZeroMQ server. Hence, the traffic over wlp3s0 is representative of all the traffic generated in the network.

Wlp5s0 is the wireless interface involved in the transfer of data between teleworkers and the waexus and voice servers and Eth0 is the ethernet interface of the YOI-VPN involved in the transfer of data from teleworkers to the office.

Due to a flooding incident at EXUS office that damaged most servers, little data could be recorded after 307h.

2.16.1.3 Distance crossed by Bytes of information

Data was traveling from IoT devices deployed at home to the servers deployed at the office via a mesh network of 3 hops maximum. For instance, Camera1-2 reaches the zeromqproxy server passing through YOI-TELEWORKER-2, YOI-VPN and YOI-RELAY-1 in Figure 229.

The maximum distance crossed by Bytes of information was the physical distance between the home and the office valued at several tenths of kilometres. Which is way under the average distance crossed by any Byte of information sent over the Internet and clouds which is 15 000 Km. WAOW tool data travelled, was stored, and processed in Greece and the company had full control of its data.

2.16.2 Use Case: Office (Mutua Universal)

MUTUA was hosting the Office Pilot in which the WAOW Tool users were all located in the same building. The challenge was to provide the WAOW Tool monitoring services to a large number of participants spread over multiple floors without any access to the building IT infrastructure (e.g. no ethernet LAN).

We choose to deploy a dedicated network infrastructure (Figure 231) composed of wireless nodes deployed on 3 different floors. The nodes were connecting together wirelessly via Wi-Fi (red and grey links) to form a mesh network. Each node also provided a Wi-Fi access point to connect participant's IoT devices (represented as buddy icons). WAOW Tool servers were located on a server room, next to the YOI-VPN and connected to the WAOW Tool network via Wi-Fi as well. The YOI-VPN provided a gateway to the Internet and a remote access via VPN to partners for technical support (internet cloud in Figure 232).

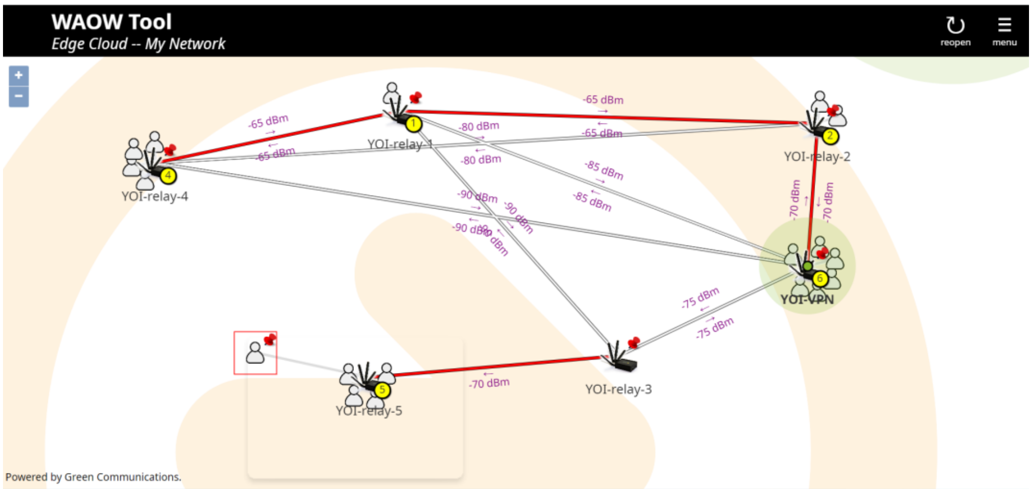


Figure 231: MUTUA network view

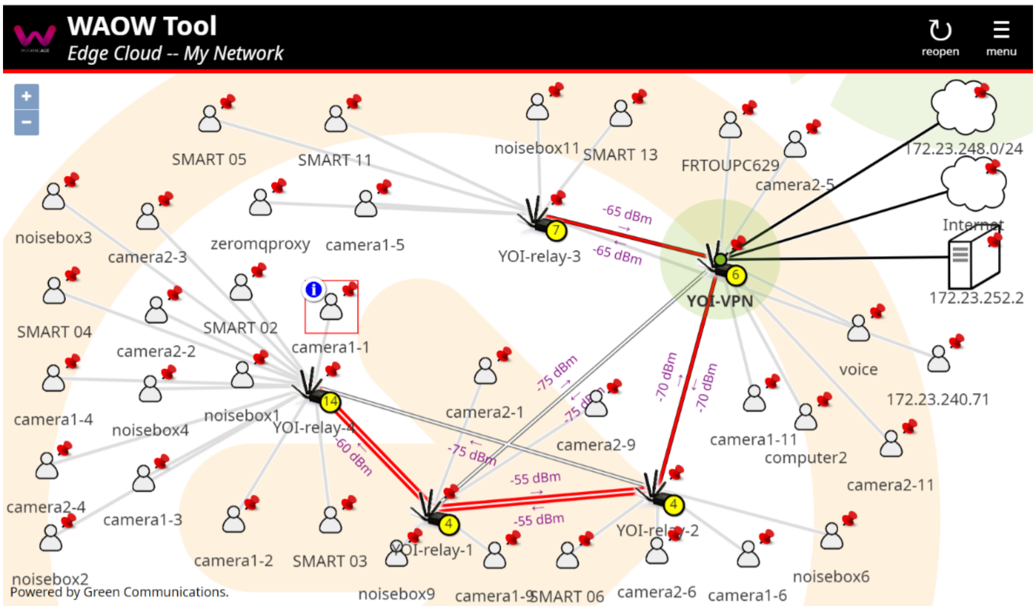


Figure 232: MUTUA network view with IoT devices

2.16.2.1 Network size

The network was composed of 6 network nodes - one YOI-VPN and 5 YOI-RELAYS. It connected 40 devices when fetching Figure 232 out of which 20 cameras, 11 smartphones, 6 noiseboxes and 3 servers.

The quality of the wireless signal between the network node was average (red links of -70 dbm) to poor (grey links of < -75 dbm) in Figure 231 due to the large area to be covered, the bad signal propagation between floors, and the interferences with other Wi-Fi in the area.

2.16.2.2 Traffic analysis

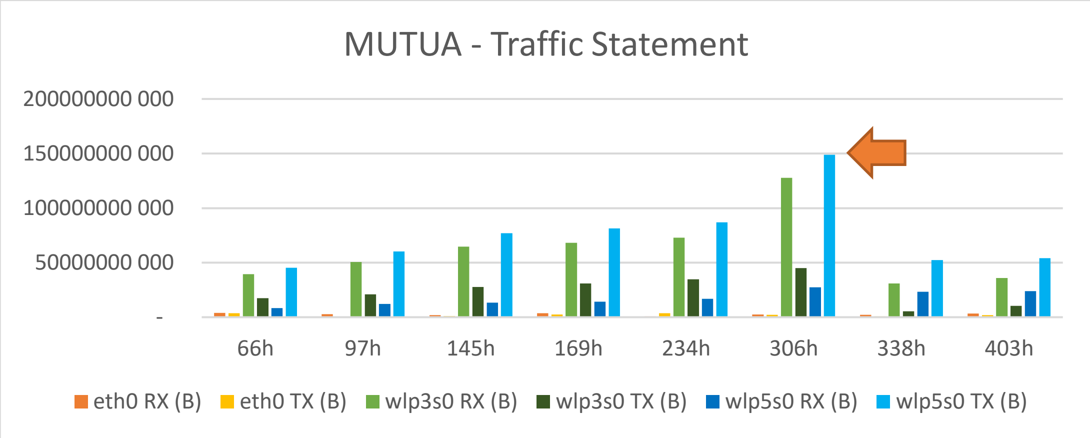


Figure 233: Mutua Traffic Statement

We registered 148GB Output traffic at peak (307h in Figure 230) on the wlp5s0 interface of the YOI-VPN network node.

Wlp5s0 is the wireless access interface of the YOI-VPN that is involved in the transfer of data between IoT devices and the server (e.g. between Camera1-1 and computer2 in Figure 232) and the transfer of data between server to another server (e.g. from voice server to zeroMQ server). Hence, the traffic over wlp5s0 is representative of all the traffic generated in the network except the traffic from and to the Internet.

Wlp3s0 is the wireless backhaul interface involved in the transfer of data between servers, IoT devices and Internet but the server-to-server communications.

Eth0 is the ethernet interface of the YOI-VPN involved in the transfer of data from or to the Internet.

It is to be noticed that servers could have been momentarily connected to another node of the network than the YOI-VPN. It is the case in Figure 232 zeromqserver is connected to YOI-RELAY-3. In that case the traffic from and to this server is not entirely recorded.

There was a flooding issue after 306h in the building basement that induced the network to restart. We have lost part of the data for the next recording.

2.16.2.3 Distance crossed by Bytes of information

Data was traveling from IoT devices deployed at the office to the servers also deployed at the office via a mesh network of 4 hops maximum. For instance, Camera1-1 reaches the computer2 server passing through YOI-relay-4, YOI-relay-1, YOI-relay-2, and YPOI-VPN in Figure 232.

The maximum distance crossed by Bytes of information was the physical distance between camera1-1 and computer2 valued at several tenths of meters. Which is way under the average distance crossed by any Byte of information sent over the Internet and clouds which is 15 000 Km. WAOW tool data was collected, stored and processed at MUTUA's office and the company had full control of its data.

2.16.3 Use Case: Production (Grupo Antolin)

Grupo Antolin (GA) was hosting the Production Pilot in which the WAOW Tool users were all located in the same production site. The challenge was to provide the WAOW Tool monitoring services to a large number of participants in a metal framework building without any access to the building IT infrastructure (e.g. no ethernet LAN).

We choose to deploy a dedicated network infrastructure (Figure 234) composed of wireless nodes deployed in the area so that they are in Wi-Fi range of each other. The nodes were connecting together wirelessly via Wi-Fi (characterized by red and yellow links) to form a mesh network. Each node also provided a Wi-Fi access point to connect participant's IoT devices (symbolized as buddy icons). WAOW Tool servers were located on a server room, next to the YOI-VPN and connected to the WAOW Tool network via Wi-Fi as well. The YOI-VPN provided a gateway to the Internet and a remote access via VPN to partners for technical support (represented by a cloud icon).

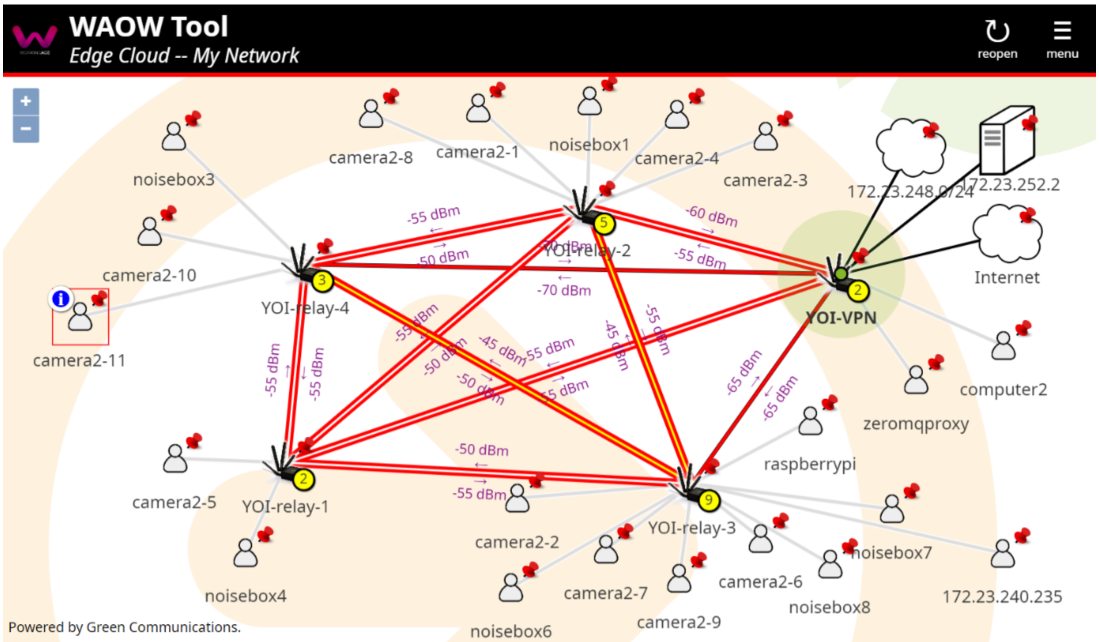


Figure 234: GA network view

2.16.3.1 Network size

The network was composed of 5 network nodes - one YOI-VPN and four YOI-RELAYS. It connected 21 devices when fetching Figure 234 out of which 11 cameras, 6 noiseboxes, 2 servers, and 2 other devices (smartphone and RPi).

The quality of the wireless signal between the network node was good to very good (red and yellow or red and white links of > -60 dbm) or average (red links of about < -60 dbm) in Figure 234 thanks to the good signal propagation in the area.

2.16.3.2 Traffic analysis

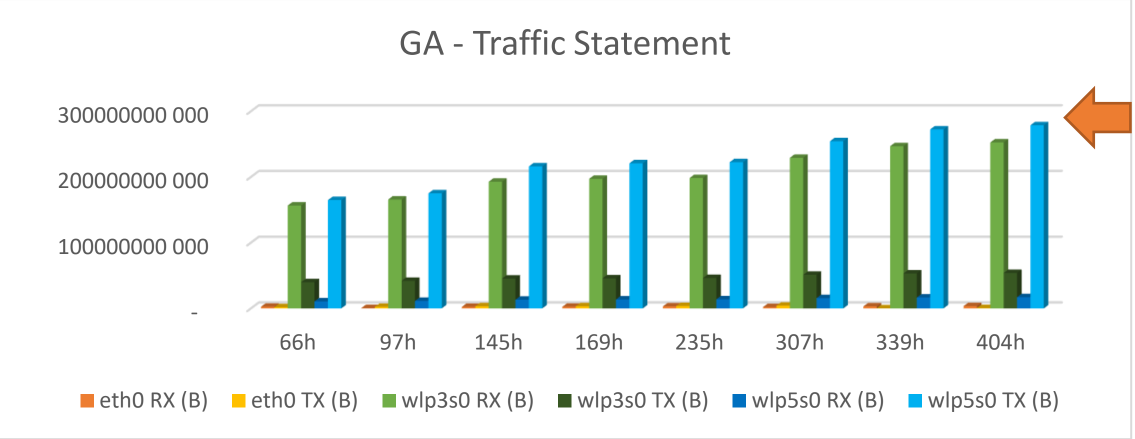


Figure 235: GA traffic statement

We registered 278GB Output traffic at peak (404h in Figure 235) on the wlp5s0 interface of the YOI-VPN network node.

Wlp5s0 is the wireless access interface of the YOI-VPN that is involved in the transfer of data between IoT devices and the server (e.g. between Camera2-11 and zeromqproxy in Figure 235) and the transfer of data between server to another server (e.g. from computer2 server to zeroMQ server). Hence, the traffic over wlp5s0 is representative of all the traffic generated in the network except the traffic from and to the Internet.

Wlp3s0 is the wireless backhaul interface involved in the transfer of data between servers, Io T devices and Internet but the server-to-server communications.

Eth0 is the ethernet interface of the YOI-VPN involved in the transfer of data from or to the Internet.

It is to be noticed that servers could have been momentarily connected to another node of the network than the YOI-VPN. In that case the traffic from and to this server is not entirely recorded.

2.16.3.3 Distance crossed by Bytes of information

Data was traveling from IoT devices deployed on premise to the servers also deployed on premise. For instance, Camera2-11 could reach the zeroMQ server using multiple path. The pass offering best link quality would route the information passing through YOI-relay-4, YOI-relay-3, YOI-relay-2, and YPOI-VPN in Figure 235.

The maximum distance crossed by Bytes of information was something in the order of the physical distance between camera2-11 and the zeroMQ server valued at several tenths of meters. Which is way under the average distance crossed by any Byte of information sent over the Internet and clouds which is 15 000 Km. WAOW tool data was collected, stored and processed at GA's office and the company had full control of its data.

2.16.4 Issues and Mitigations

Table 25: Wi-Fi - Issues and Mitigations

Issue	Description	Mitigation
Multicast	We discovered a bug in the multicast service provided by WAOW tool network leading multicast traffic to loop in the network forever. This overload of the network lead routers to disconnect.	The issue is fixed in more recent version of the software. To bypass the issue for WorkingAge Project we disabled multicast on WorkingAge netwoks as multicast is not used in the project.
Ejabberd	We discovered a bug in our ejabberd server overloading network node’s CPU.	We fixed the bug in our latest software version and disabled ejabberd on WorkingAge network as no ejabberd server is used in the project
Latency	We experienced latency of up to a second on MUTUA’s network. Our investigations led to think it was due to the interferences on 2.4GHz Wi-Wi frequencies at MUTUAS’s office.	We switched the network backhaul network frequency to 5Ghz. Although we have lost in link quality, the change improved significantly the latency. Thus, enabling the services to work properly.
No access to company’s IT infrastructure	Hosting companies restricted the access to their IT infrastructure due to IT security reasons. We could not use their Internet access nor their ethernet backhaul.	We created a fully autonomous solution using wireless network infrastructure and dedicated Internet lines.

2.16.5 Conclusion and Lesson Learned

Through this project, we created a fully autonomous WAOW Tool that can be deployed anywhere without interfering with IT system of the host organization. Once deployed, the WAOW Tool fetches, collects, stores, and processes data on premises or in the vicinity without any data traveling the entire Internet (15 000Km in average) to reach any remote cloud sometime located abroad. Thus, leading to grater data security, sovereignty, privacy, etc. Although the infrastructure was submitted to wireless interferences, signal obstructions, and unpredicted events (power shut down, flooding), the solution was resilient and provided sufficient capacity for performing the WA services.

2.17 ZeroMQ

2.17.1 Introduction

High-performance messaging middleware is an important infrastructure component of any distributed application. A well-designed messaging infrastructure can minimize dependencies between application components, enabling evolution of the software as requirements change and as new components are integrated. Here, we evaluate the performance of ZeroMQ, the communication middleware mechanism used to achieve continuous communication and exchange of information.

2.17.2 Method

A list of key performance indicators (KPIs) was developed in order to evaluate the system's functionality, e.g., the ability of ZeroMQ to handle simultaneously messages from all users. These 4 KPIs are:

4. KPI1: Add User functionality
    a. Mobile sent the user registration details to Edge Cloud, as prompted by the WAOW app
    b. Mobile sent the user registration details (including photo) to Facial/Gesture Recognition component
5. KPI2: Edge Cloud sent an ack to the Mobile App
6. KPI3: Mobile sends start-stop messages to the Edge Cloud
7. KPI4: Edge cloud sends high level data to mobile app
    a. Messages regarding performed gesture are sent back to the app
    b. Edge Cloud sent the high-level data to the Mobile App

The KPIs were achieved with all components operating simultaneously. The above messages were selected as they vary in function, complexity and payload size, parameters that could impact system performance. Example of the message for each KPI follow (Table 26):

Table 26: Example payloads for each KPI used to test the ZeroMQ functionality

KPI	Payload
1a	{ "userpseudoid": "U550e8400-e29b-41d4-a716-446655440000", "sensorgroupid": "adduser/S76c91ee3-323b-47f3-b595-79a3d533d9a6", "rsa4096publickey": "ssh-rsa AAAAB3NzaC1yc2EAAAABIwAAAQEAklOUpkDHrfHY17SbrmTIpNLTGK9Tjom/BWDSUGPl+nafzlHDTYW7hdI4yZ5ew18JH4JW9jbhUFrviQzM7xlELEVf4h9lFX5QVkbPppSwg0cda3Pbv7kOdJ/MTyBlWXFCR+HAo3FXRitBqxiX1nKhXpHAZsMciLq8V6RjsNAQwdsdMFvSlVK/7XAt3FaoJoAsncM1Q9x5+3V0Ww68/eIFmb1zuUFljQJKprrX88XypNDvjYNby6vw/Pb0rwert/EnmZ+AW4OZPnTPI89ZPmVMLuayrD2cE86Z/il8b+gw3r3+1nKatmIkjn2so1d01QraTlMqVSsbxNrRFi9wrf+M7Q=="} }
1b	{ "userpseudoid": "U550e8400-e29b-41d4-a716-446655440000", "sensorgroupid": "S4188c841-e28a-4865-b233-d39a465358ff", "rsa4096publickey": "ssh-rsa AAAAB3NzaC1yc2…GgtShbs9649r/Loufhl…"

	"photo": "/9j/4AAQSkZJRgABAgAAAQABAAD…pHje8vrprm9Hk2lv5sD\\n7llVI" }
2	{ "sender": "voice.workingage.eu" }
3	{ "action": "start" }
4a	{ "gesture": 1 }
4b	{ "probability": 0.68, "timeStamp": 20200412202123, "sensorType": "Neurophysiological", "values": [{ "sensor":"HeartRate", "value": "High" }] }

The communication between Mobile App and the Edge Cloud components is a two-way communication process, meaning that the sender and receiver listen to each other in order to be able to respond at any time. For this reason, messages are collected at 2 points: messages sent by the Mobile App to the Edge Cloud and received back to the Mobile App by the Edge Cloud components. ZeroMQ is a commercial off-the-shelf (COTS) tool, therefore its performance is already reported in the literature. Instead, we focus on the evaluation of the aforementioned KPIs. Results are presented below for short and long-term tests.

2.17.3 Results – Short Term Tests

The goal of the single-day tests for ZeroMQ was to ensure the system functionality (e.g. the publish and subscribe services) and the correct operation of the proxy server. As this system affects all sensors communicating with the edge servers, the individual sensor tests automatically become tests for the ZeroMQ as well.

The ZeroMQ messaging service functionality was verified during the short-term tests by monitoring the app subscription process for each of the participating volunteers, which, in essence, reproduced the specified KPIs. In this process, ZeroMQ plays a critical role, as it is responsible for receiving a unique registration message from the mobile phone and sending it to each edge server, and, subsequently, receiving the unique acknowledgement from each edge server of a successful registration and sending it to the app on the mobile phone. Upon registration, the users were asked to also "start" and "stop" the sensors, demonstrating the process. This process also works by a message exchange through the ZeroMQ middleware, thus confirming successful message exchange and the correct functionality and effectiveness of the ZeroMQ topics.

There were no errors or issues associated with ZeroMQ identified in the single-day test of the short-term testing phase.

2.17.4 Results – Long Term Tests

To verify correct operation of the ZeroMQ during the week test, the ZeroMQ log files generated by the mobile app were obtained from select volunteers. The log files were checked to verify:

- Receipt of high-level messages from edge servers with the established 1-minute interval.
- That only the messages for the specific SensorGroupID and UserPseudoID were received; thus no message interference was occurring.

Further, some sensors (e.g. the noisebox) where manually shut down and restarted with the mobile app remaining in the "start" recording mode. When the sensors restart, they are by default in the "stop" recording mode. The check performed was to verify that the sensors resumed operation (e.g. "start" recording) within 10 seconds, the frequency at which the system status messages are sent. This process was additionally monitored through the ZeroMQ logs.

Regarding the KPIs and system scalability, a number of messages (according to KPI payload indicated in Table 26) were generated at different frequencies, in order to measure latency from the generated log files as well as potentially lost messages. The results are shown in Figure 236. In all cases tested, a latency of 1ms was found. A notable exception is KPI 1b, where the payload size (containing the user photo) causes a larger latency in processing, particularly when messages are sent very frequently (e.g. at 1ms intervals). For this case, latencies of nearly 600 ms were measured in the most extreme cases. Although this is a rather large latency, we note that this specific message is exchanged only once, during user registration. In all cases tested, no messages were dropped.

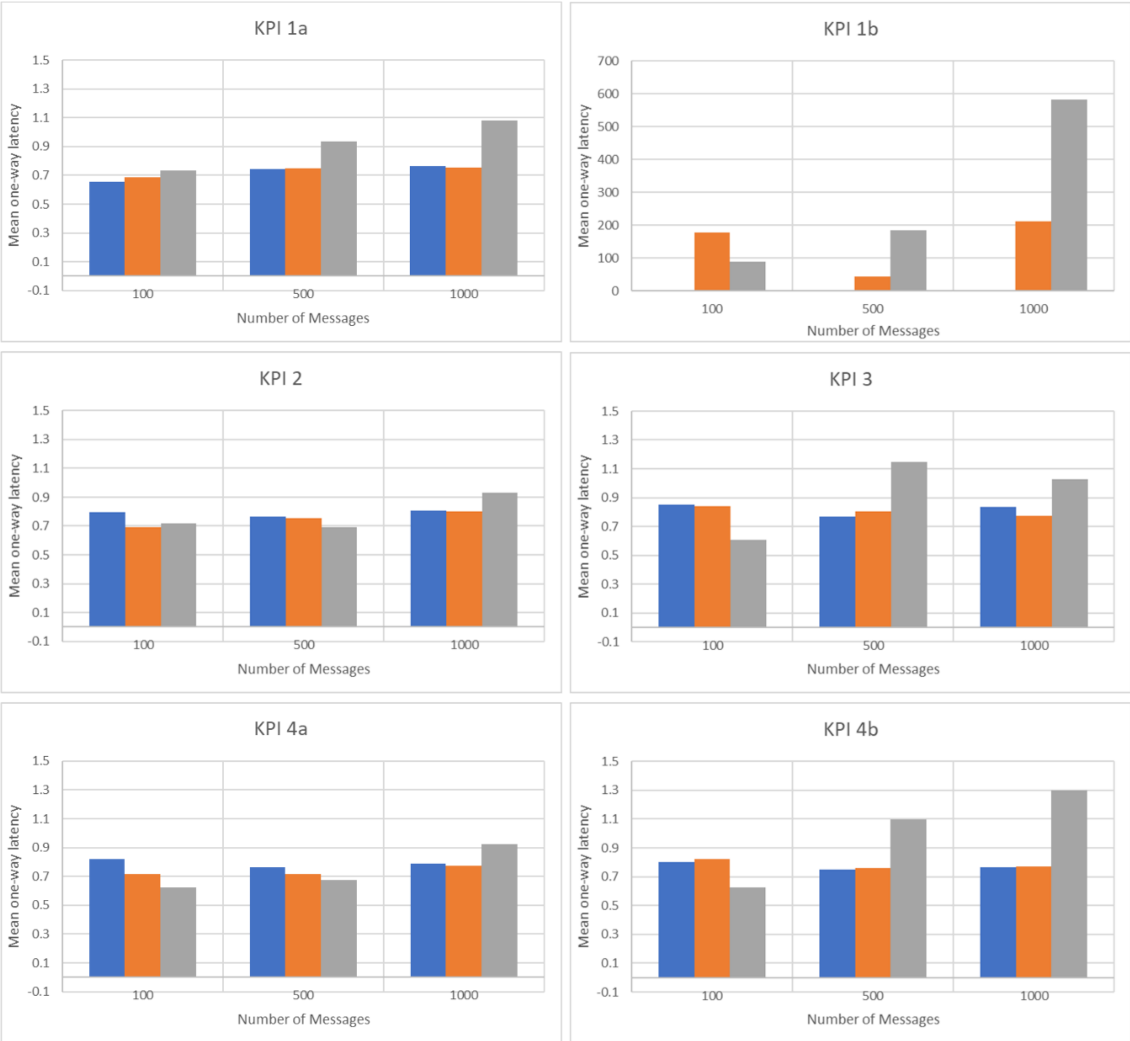


Figure 236: One-way latency for the defined KPIs, as a function of number of messages and message frequency (blue = 100ms, orange = 10ms, grey = 1ms). All latency results reported in ms.

Detailed results for each case tested are further reported in Table 27 through Table 29.

Table 27: Latency results for 100 ms message frequency

#messages	KPI1a [ms]	KPI1b [ms]	KPI2 [ms]	KPI3 [ms]	KPI4a [ms]	KP4b [ms]	Lost
100	Min:0.532000 Max:0.797000 Mean:0.654930 StdDev:0.040296	Min:0.798000 Max:3.868000 Mean:0.985000 StdDev:0.295924	Min:0.662000 Max:0.864000 Mean:0.796440 StdDev:0.035075	Min0.711000 Max:0.955000 Mean:0.850550 StdDev:0.056519	Min:0.600000 Max:0.917000 Mean:0.818770 StdDev:0.065397	Min:0.684000 Max:0.901000 Mean:0.799580 StdDev:0.039703	0
500	Min:0.590000	Min:0.696000	Min:0.731000	Min:0.650000	Min:0.597000	Min:0.592000	0

	Max:0.874000 Mean:0.744322 StdDev:0.062398	Max:9.452000 Mean:0.980702 StdDev:0.384261	Max:0.870000 Mean:0.764844 StdDev:0.033439	Max:0.907000 Mean:0.765397 StdDev:0.054743	Max:0.927000 Mean:0.761081 StdDev:0.069108	Max:0.866000 Mean:0.750312 StdDev:0.051373	
1000	Min:0.632000 Max:0.919000 Mean:0.762915 StdDev:0.050572	Min:0.765000 Max:20.047000 Mean:0.932204 StdDev:0.606868	Min:0.597000 Max:0.991000 Mean:0.807967 StdDev:0.063643	Min:0.651000 Max:0.970000 Mean:0.836545 StdDev:0.047572	Min:0.640000 Max:0.955000 Mean:0.788962 StdDev:0.07007	Min:0.551000 Max:0.974000 Mean:0.766519 StdDev:0.053214	0

Table 28: Latency results for 10 ms message frequency

#messages	**KPI1a [ms]**	**KPI1b [ms]**	**KPI2 [ms]**	**KPI3 [ms]**	**KPI4a [ms]**	**KP4b [ms]**	**Lost**
100	Min:0.521000 Max:0.808000 Mean:0.68840 StdDev:0.057611	Min:0.606000 Max:438.169000 Mean:178.678230 StdDev:150.667919	Min:0.476000 Max:0.897000 Mean:0.689998 StdDev:0.045579	Min:0.534000 Max:1.019000 Mean:0.844210 StdDev:0.092340	Min:0.548000 Max:0.776000 Mean:0.718810 StdDev:0.048561	Min:0.464000 Max:0.904000 Mean:0.824312 StdDev:0.079293	0
500	Min:0.512000 Max:0.866000 Mean:0.75054 StdDev:0.049666	Min:0.635000 Max:1245.859 Mean:44.1423 StdDev:175.2689	Min:0.532 Max:0.960000 Mean:0.756705 StdDev:0.062387	Min:0.491000 Max:0.982000 Mean:0.801860 StdDev:0.069687	Min:0.406000 Max:0.938000 Mean:0.714906 StdDev:0.077473	Min:0.548000 Max:0.930000 Mean:0.760002 StdDev:0.065665	0
1000	Min:0.566000 Max:1.031000 Mean:0.751716	Min:0.708000 Max:861.059 Mean:212.486	Min:0.499000 Max:0.914000 Mean:0.803020	Min:0.561000 Max:0.969000 Mean:0.771751	Min:0.378000 Max:0.929000 Mean:0.773369	Min:0.406000 Max:0.995000 Mean:0.772258	0

	StdDev:0.058136	StdDev:147.699	StdDev:0.09327	StdDev:0.050271	StdDev:0.043601	StdDev:0.063655	

Table 29: Latency results for 1 ms message frequency

#messages	KPI1a [ms]	KPI1b [ms]	KPI2 [ms]	KPI3 [ms]	KPI4a [ms]	KP4b [ms]	Lost
100	Min:0.409000 Max:0.619250 Mean:0.735040 StdDev:0.902670	Min:0.619000 Max:291.87000 Mean:98.26343 StdDev:88.9386	Min:0.309000 Max:0.520000 Mean:0.716880 StdDev:0.885565	Min:0.374000 Max:0.732000 Mean:0.607925 StdDev:0.061315	Min:0.401000 Max:0.827000 Mean:0.624200 StdDev:0.070625	Min:0.412000 Max:0.874000 Mean:0.625620 StdDev:0.088570	0
500	Min:0.373000 Max:13.115000 Mean:0.934901 StdDev1.823469	Min: 0.514000 Max: 913.737 Mean:78.7337 StdDev:185.450	Min:0.248000 Max:7.335000 Mean:0.694334 StdDev:0.446123	Min:0.423000 Max:20.528000 Mean:1.149050 StdDev:2.622211	Min:0.349000 Max:13.889400 Mean:0.673523 StdDev:1.72129	Min:0.428000 Max:20.214000 Mean:1.096454 StdDev:0.076804	0
1000	Min:0.391000 Max:19.993000 Mean:1.081554 StdDev:2.326847	Min:6.821000 Max:2349.228 Mean:1071.758 StdDev:581.524	Min:0.43600 Max:19.089000 Mean:0.929432 StdDev:1.620759	Min:0.479000 Max:21.588000 Mean:1.027350 StdDev:2.065248	Min:0.373000 Max:20.386000 Mean:0.921638 StdDev:1.827355	Min:0.404000 Max:21.273000 Mean:1.298354 StdDev:2.970244	0

2.17.5 Summary and Outlook

Brokerless systems, like ZeroMQ, offer considerably higher throughput and less latency than classical message-oriented middleware, but require greater orchestration of network topologies. ZeroMQ is especially suited for high performance and low latency scenarios. They offer high scalability but can lead to tighter coupling between components.

End-to-end latency for messages of up a few bytes remains fairly consistent, with a range from 0.5ms to 1ms. The actual values will depend, of course, on the physical location of the machines where the instances are allocated. As one would expect, latency increases with larger messages or larger distances of physical location.

The ZeroMQ system proved to be a reliable tool for communication between edge-servers, sensors and the mobile app. In fact, the ZeroMQ could also be used as a debugging tool, e.g., if an unexpected behaviour occurred it was a strong indication that some aspect of the system was not behaving correctly (e.g., sensor was offline, network communication issue, etc.) rather than an error in the ZeroMQ proxy itself.

2.18 WAOW App

WAOW app, It is an android app that communicates with the different sensors through ZeroMQ or Bluetooth, receives all the data from these sensors, processes that data and sends them to the DSS contained in the app, it also shows the user graphs, advices, questionnaires, and some of the processed data.

2.18.1 Method

Following the specifications given in the D9.1 "Evaluation protocol", we divided the evaluation in two parts: Short Term tests and Long Term tests, where the first phase was carried out at the premises of ITCL, while the second phase was conducted at the three In-Company tests sites: EXUS (Greece), Mutua Universal (Spain) and Grupo Antolin (Spain).

Test procedure:

- Install the app on the mobile phone
- Open the app.
- Log in with user and password
- Check if the main service is working (there should be a notification in the notification bar with the WorkingAge logo)
- Set all sensors
- Open profile settings and click on Logs button

You can access the settings screen from the button located at the top right of all the screens of the application (Figure 237 and Figure 238).

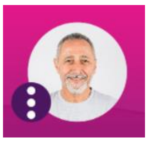


Figure 237: WAOW App settings button

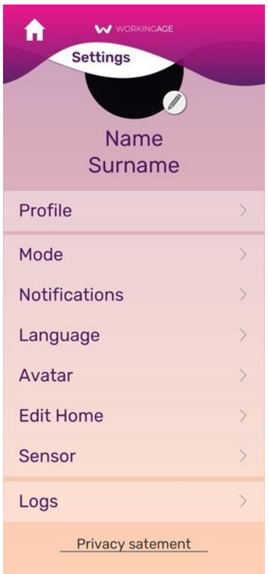


Figure 238: WAOW App settings screen

- Check the different logs (one per sensor/module). A generic view of WA personnel at these logs is sufficient to verify correct communication with each sensor/module (Figure 239).

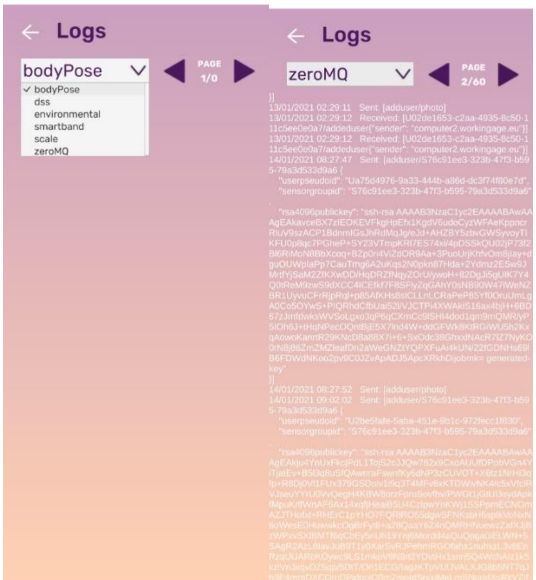


Figure 239: WAOW App settings screen

2.18.2 Results – Short Term Tests

During the short term tests, the app was tested on all the phones that were used in the project one by one, the method explained above was followed, due to the problems that occurred with the driver crisis and the pandemic, not all the sensors were available to be tested at the same time with the app but with those that were available it was a success, in all the phones it was verified that there was communication with the edge cloud, with the Bluetooth sensors, with the DSS and that the service in the background of the app that allows it to always be running.

2.18.3 Results – Long Term Tests

During the long term tests, monitoring was more difficult since the logs could only be accessed at specific times or at the end of the tests, during which time the users had to configure and register the app and the different sensors, some of the sensors in their homes, such as the scale, were provided with manuals and instructions and were given support whenever possible, but in spite of everything, some of the volunteers did not know how to handle these technologies well and made mistakes that affected the working of the app or some sensors.

When working with all the sensors at the same time in a network with several other users, the app slowed down due to having to process too many messages, and there were several complaints from users for this reason, in the middle of the long term test the app was updated with a new version of the app that improved several of these defects.

Despite these problems, the app worked as it should, and in most cases it collected the necessary data for the different studies of the project (Table 30).

Table 30: DSS logs recovered

Pilot	Type of pilot	Users Registered	DSS logs	Percentage
Grupo Antolín	Production	13	11	84
Mutua Universal	Office	14	12	100
EXUS	Telework	9	7	77

2.18.4 Discussion and Outlook

Building the app was hard work since it was necessary to unite many technologies that changed or were adapted throughout the project, in addition to being the centre of the entire system, any failure in the network, communication, sensors or any other of the system elements is shown as a malfunction of the app, as demonstrated with the improvement after the update during the long term test, perhaps with a longer period and some more updates to debug the system, it would have been possible to improve much more.

However, although it was not perfect, the goals we set were achieved and the app worked correctly, so we are satisfied with the result.

2.19 WAOW Decision Support System performance

2.19.1 Introduction

The Decision Support System (DSS) is the designed reasoning core of the WAOW Tool. Such a system supports the worker while dealing with working activities. The DSS is based on a model-driven approach, in particular, it is a rule-based engine, but it is able to work with the results of the classifiers, which generate a probabilistic output. A typical rule system is typically expressed as a set of facts, and a set of rules that describe how to derive new facts, as it happens in well-known declarative programming languages like Prolog (Muggleton et al., 2012). The DSS integrate in the WAOW tool was based on a data-driven model and it was technically developed to be able to handle probabilistic rules (De Raedt et al., 2007). Therefore, the DSS's reasoning engine is ProbLog, an extension of the popular Prolog. As the original Prolog, ProbLog uses procedural interpretation of c, in this case labelled with probabilities.

2.19.2 Method

The reasoning engine of the WAOW tool's DSS is Problog. ProbLog is implemented as a Python library. To respect the very stringent GDPR and privacy requirements of WorkingAge project, we chose to implement the DSS directly into the WAOW app running on each worker's mobile phone. In this way, each DSS works on information belonging to the owner of the smartphone. In addition, by using this solution each worker will be totally isolated form the others, and there will not be any single component that need to know all the workers' information. Unfortunately, this choice, albeit very secure, forced us to run the Python-based DSS into the Android-based WAOW app. The same Python library is compatible for the implementation in iOS environment. For doing that, we leveraged an Android-friendly Python distribution called Chaquopy (Chaquo Ltd), which permitted us to run the ProbLog Python code into the Java-based Android app. Then, we defined a Java API that hid the Python machinery and provided to the Android app a standard Java library. The API permitted to:

- Manage the DSS instance
- Manage persistence of the DSS state when the app is killed by Android
- Insert sensor measures
- Ask for new advice
- Insert worker's feedback
- Run the rule adaptation algorithm
- Manage backup/restore of the DSS stare

The DSS was technically implemented in the WAOW Android mobile app. Such implementation was placed in a larger context in which different entities, implementing the Decision-Making Tool (DMT) pipeline, are interfaced. Figure 240 shows how the DMT components communicated:

- Most of the sensors integrated in the WAOW tool were connected to the Edge Cloud through Wi-Fi (1); the Edge Cloud then sent the high-level information to the DSS (2).
- Some sensors sent measurements directly to the app, by means of Bluetooth, and did not require any intervention of the Edge Cloud (see Sensor "Z").
- Other sensors sent measurements directly to the app by means of Bluetooth, and the app forwarded data to the Edge Cloud (1' and 1") to get high-level information (2).

The Edge Cloud included data processing servers, i.e., miniaturized portable PC, and the middleware infrastructure.

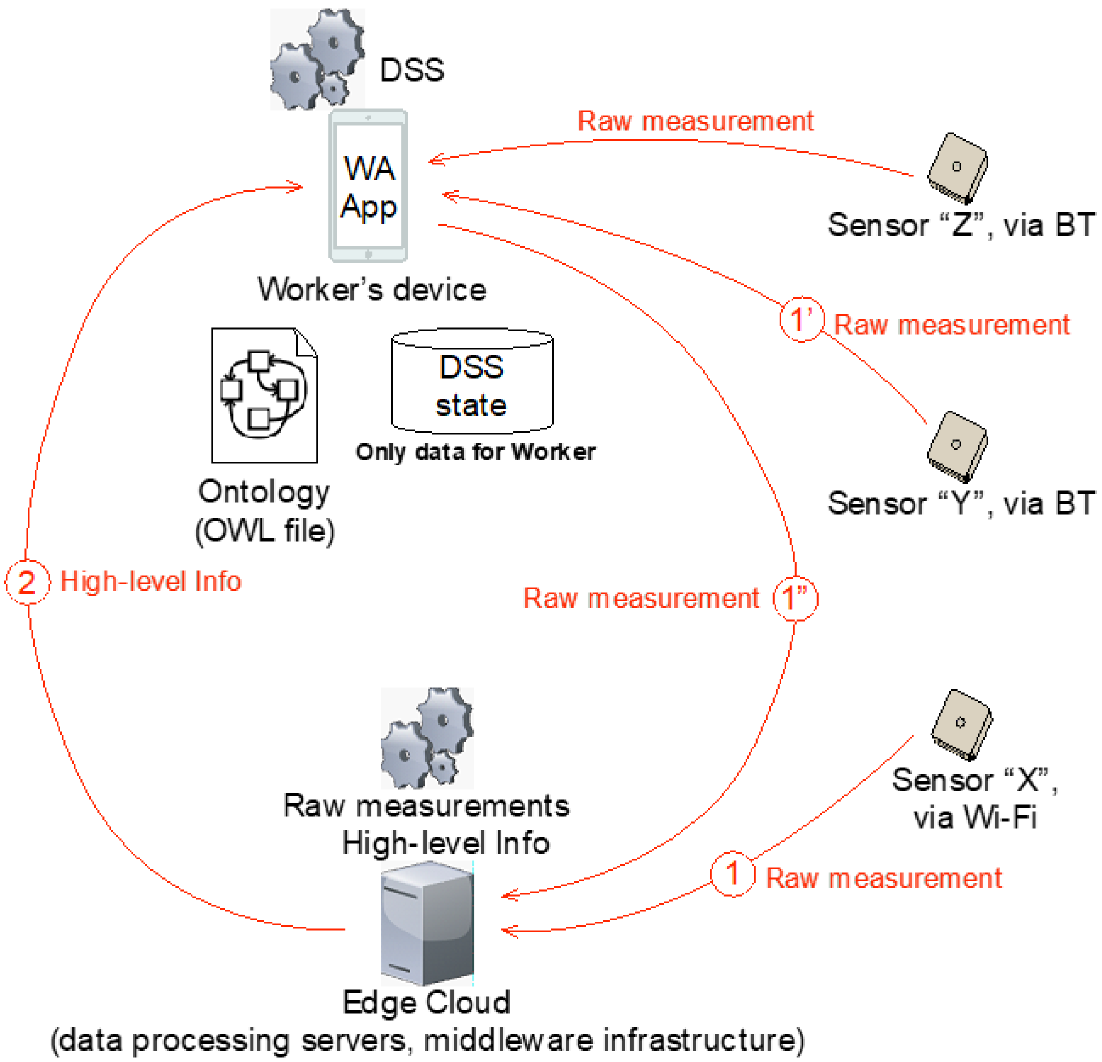


Figure 240: Communication schema of the Decision-Making Tool (DMT) pipeline including the DSS and the different Subsystems.

2.19.3 Results

Due to the shortness of the In-Company tests, on which different factors impacted, such as the pandemic and the global semiconductor crisis, the technical partners of the Consortium (BS, UCAM, AUD, RWTH) decided to perform an additional analysis to quantify the DSS performance. In this regard, it was chosen to feed the final version of the DSS with the mental, emotional, and physical states classifications of the participants involved in the In-Lab experimental phase of the WorkingAge project. In fact, in the context of the In-Lab tests, the above-mentioned technical partners agreed to perform additional data collection using a common set of sensors to enlarge the sample size. For further details, please refer to the Deliverable 9.3.

The analysis of the DSS performance through the simulation applied by using the In-Lab shared dataset corresponded to feeding the DSS final version with the classifications related to the mental workload, stress, arousal, and emotional state of the participants while they were performing activities to simulate the three different working environments involved in the In-Company tests. In particular, the DSS generated a detailed report for each involved participant, which included the sensors' inputs and the advice that the DSS generated accordingly. The DSS performance analysis consisted in:

- Estimation of the number of coherent suggestions corresponding to the cases in which the DSS provided advice and / or suggestions coherent with the sensors' inputs. As an example, the case in which the DSS suggested the participant to take a break because the Neurometrics Subsystem classified the stress level as High.
- Estimation of the number of missing suggestions corresponding to the case in which the DSS did not provide any advice and / or suggestions, even if the sensors inputs indicated a negative alteration of the evaluated mental and physical states.
- Estimation of the number of non-coherent suggestions corresponding to the cases in which the DSS provided advice and / or suggestions non-coherent with the sensors' inputs. As an example, the case in which the DSS suggested the participant to parallelize the working tasks when the Neurometrics Subsystem classified the mental workload as High.

The DSS performance analysis results are reported in Table 31 below:

Table 31: DSS performance analysis performed in the In-Lab shared dataset.

Subject ID	Mental workload	Stress	Arousal	Emotional state
Subj1	coherent suggestions: 89,8% missing suggestions: 10,2% non-coherent suggestions: 0%	coherent suggestions: 90% missing suggestions: 10% non-coherent suggestions: 0%	coherent suggestions: 89,8% missing suggestions: 10,2%	coherent suggestions: 80% missing suggestions: 10% non-coherent suggestions: 10%

			non-coherent suggestions: 0%	
Subj2	coherent suggestions: 90% missing suggestions: 10% non-coherent suggestions: 0%	coherent suggestions: 80% missing suggestions: 10% non-coherent suggestions: 10%	coherent suggestions: 90% missing suggestions: 10% non-coherent suggestions: 0%	coherent suggestions: 90% missing suggestions: % non-coherent suggestions: 10%
Subj3	coherent suggestions: 81,8% missing suggestions: 18,2% non-coherent suggestions: 0%	coherent suggestions: 80% missing suggestions: 20% non-coherent suggestions: 0%	coherent suggestions: 81,8% missing suggestions: 18,2% non-coherent suggestions: 0%	coherent suggestions: 80% missing suggestions: 10% non-coherent suggestions: 10%
Subj4	coherent suggestions: 89,8% missing suggestions: 10,2% non-coherent suggestions: 0%	coherent suggestions: 89,8% missing suggestions: 10,2% non-coherent suggestions: 0%	coherent suggestions: 90% missing suggestions: 10% non-coherent suggestions: 0%	coherent suggestions: 89,8% missing suggestions: 10,2% non-coherent suggestions: 0%
Subj5	coherent suggestions: 72,7% missing suggestions: 27,3% non-coherent suggestions: 0%	coherent suggestions: 81,8% missing suggestions: 18,2% non-coherent suggestions: 0%	coherent suggestions: 100% missing suggestions: 0% non-coherent suggestions: 0%	coherent suggestions: 89,8% missing suggestions: 10,2% non-coherent suggestions: 0%
Subj6	coherent suggestions: 90% missing suggestions: 10% non-coherent suggestions: 0%	coherent suggestions: 80% missing suggestions: 10% non-coherent suggestions: 10%	coherent suggestions: 89,8% missing suggestions: 10,2% non-coherent suggestions: 0%	coherent suggestions: 80% missing suggestions: 10% non-coherent suggestions: 10%
Subj7	coherent suggestions: 81,8% missing suggestions: 18,2%	coherent suggestions: 80% missing suggestions: 10%	coherent suggestions: 90% missing suggestions: 10%	coherent suggestions: 81,8%

	non-coherent suggestions: 0%	non-coherent suggestions: 10%	non-coherent suggestions: 0%	missing suggestions: 18,2% non-coherent suggestions: 0%
Mean	**coherent suggestions: 85,1%** **missing suggestions: 14,9%** **non-coherent suggestions: 0%**	**coherent suggestions: 83,1%** **missing suggestions: 12,6%** **non-coherent suggestions: 4,3%**	**coherent suggestions: 90,2%** **missing suggestions: 9,8%** **non-coherent suggestions: 0%**	**coherent suggestions: 84,5%** **missing suggestions: 9,8%** **non-coherent suggestions: 5,7%**

The averaged percentage associated to the DSS performance for each mental and emotional state are reported in Figure 241:

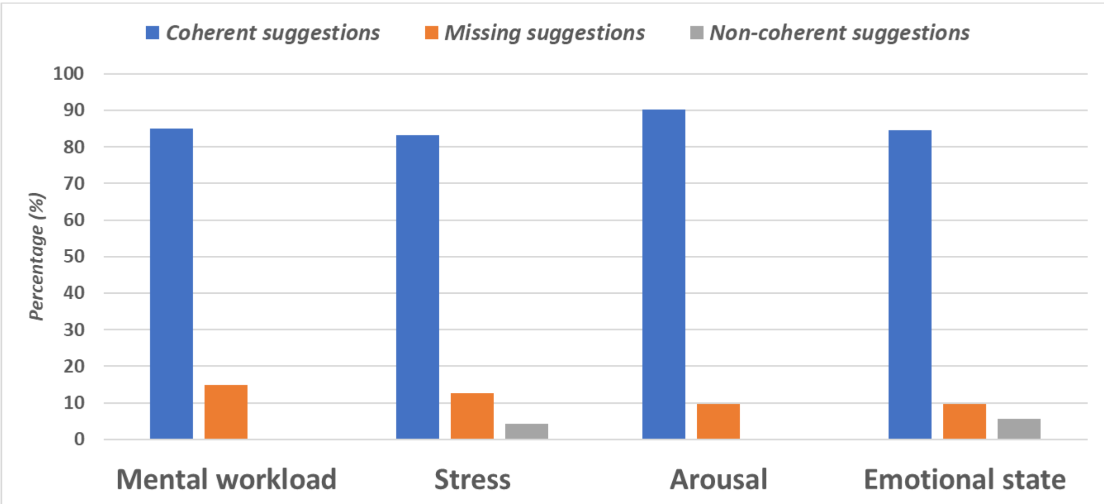


Figure 241: The averaged percentages associated to the DSS performance for each mental and emotional state.

2.19.4 Discussion and Outlook

The results obtained through the presented additional analysis conducted on the DSS performance are overall satisfying and relevantly promising. It can be observed that the DSS reacted at least in 85% of the cases by providing advice and / or suggestions to the participants according to the sensors' inputs. Furthermore, the results revealed that only for two mental states and for a minimum percentage (5.7%), the DSS generated a non-coherent advice and / or suggestions. In other words, the DSS simulation functioning revealed that it's reasoning engine positively impacted or did not negatively impact for at least the 94% of the cases. Additionally, the coherent advice and / or suggestions generated by the DSS occurred in at least the 83.1% of the cases, while for the arousal classifications such a percentage corresponded to over 90%.

The presented DSS performance are very promising and even more valuable if we consider the fact that the DSS reasoning engine was not properly trained before the simulation, due to the shortness of the In-Company tests. In other words, we obtained good results despite the short time for fine tuning of the DSS rules associated to the mental, emotional, and physical states classifications.

3. Secondary & Tertiary Users

3.1 Introduction

Advancements in psychophysiological and physical sensor-based measurements, as well as the development of the Internet of Things (IoT), have led to the feasibility of sensor-based measurements in real-world workplaces, with the goal of developing a workplace or technological system that, while considering employees' stress and strain, prevents human performance from being impaired by over- or under-strain. Given that major research efforts have been made to develop methods for monitoring employees´ strain, the lack of research on whether employers and employees support the usage of such measures is surprising (see van Acker et al. 2020). Even though general literature on technology acceptance exists, such as Venkatesh et al. (2016) and Venkatesh and Davis (2000), and may provide initial guidance, it lacks examination of the specific acceptance and connected advantages and risks of health risk monitoring technology at the workplace. User acceptance is defined as a decision made by an individual at a particular point in time to use technology intentionally. Related to health risk monitoring at the workplace, previous studies have predominantly focused on predictors related to the individual (the working person), such as organizational commitment, organizational identification, monitoring attitudes, fairness, and invasion of privacy or surveillance and perceived control and although the TAM-Model is one of the most widely used models there is a lack of research on what influences acceptance in terms of measuring health-related parameters in the work context (Abraham et al., 2019; Or & Karsh, 2009; Venkatesh & Bala, 2008). Furthermore, the transferability of existing research results from the private usage of systems to monitor health-related parameters are only possible to a limited extent due to the special relationship between employee and employer. It can be assumed that the motivators of use are different for these two stakeholders, as managers decide for employees but may not be the actual end users, while employees have less influence on the decision to implement such a system but are the end users of such technologies.

To gain a more comprehensive insight, additional secondary users, namely human resource managers, occupational health and safety professionals, and tertiary users, namely healthcare professionals, and platform developers, were asked about their thoughts, wishes, and ideas regarding the WAOW tool. With the help of two online questionnaires, advantages, and disadvantages of the WorkingAge system from the point of view of managers (in the roll of the acquiring of such a system) and employees (in the roll of the using such a system) are worked out and possible suggestions for improvement is discussed. The research questions to be answered are the following:

1. **Research Question 1:** Managers Insights: Assessment of expectations, concerns, and willingness to support the WAOW tool in their organizations.

2. **Research Question 2:** Employees Insights: Assessment of expectations, concerns, and willingness to use the WAOW tool.
3. **Research Question 3:** Work Design: Relationship between work design and user acceptance of the WAOW tool.

3.2 Research Question 1 – Manger Insights

3.2.1 Introduction

This first study aims to contribute to the first research question, in which the expectations and concerns of managers are queried to analyse factors which supports the implementation of sensor-based health risk monitoring system like the WAOW tool. The aim of this study is therefore to examine the managers' view of the extent to which and under what conditions they would support the use of the tool in their companies.

Research Objectives

This study aims at identifying the expectations and concerns of managers regarding employees' strain monitoring, specifically the WOAW-Tool, as this group of users is rarely considered yet will be key forces in the implementation of the tool. Examining expectations and concerns of this user group, allows us to assess if there is an overlap in concerns for managers and WAOW tool primary users, what aspects should be added and how these influence managers' willingness in implementing the tool. Therefore, the main research questions are:

1. Which expectations and concerns do managers have about sensor-based health risk monitoring?
2. How do these expectations and concerns relate to managers' willingness to support the use of sensor-based health risk monitoring in their company?
3. How does this differ in terms of cultural differences within Europe?

3.2.2 Method

3.2.2.1 Participants

The overall sample size of the study included N = 887 participants. Participants had to be at least 18 years old to be included. Even though most participants that answered the questionnaire had some sort of managing job, there was an abundance of different occupations amongst the 887 participants. Overall, 292 females (33%), 591 males (66.7%) and two participants identifying as other (0.2%) answered the questionnaire. Participants were recruited from three countries: Germany, UK, and Spain. Table 32 shows an overview of the sample size for each country. Care was taken to ensure that specifically human resource managers, occupational health, and safety professionals, as well as healthcare professionals were included. In addition, the analysis of workplaces with predominantly physical or mental as well as combined work tasks was considered in order to be able to include as many viewpoints as possible.

Table 32: Overview of sample size per-country

	N	Sex	Mean Age (SD)	Work Tasks
Germany	604	Male: 67% Female: 33%	43.3 (14.9)	Physical: 25% Mental: 48% Both: 27%
UK	154	Male: 68% Female: 32%	43.3 (13.7)	Physical: 20% Mental: 51% Both: 29%
Spain	128	Male: 66% Female: 34%	41.1 (14.3)	Physical: 38% Mental: 24% Both: 38%

3.2.2.2 Questionnaire

After an introductory section which collected participants' demographics, their form of employment, and information about their employer or company, participants were introduced to the concept of sensor-based health risk monitoring, as it is done with the WAOW tool. The subjects were presented with a summary of the different sensors and their measurement methods that are used in the WAOW tool. In the next step, participants were asked to rate every single sensor with regard to possible concerns that may arise through its implementation within their companies. These concerns were predetermined and selected based on Schall et al. (2018). The three predetermined concerns were privacy concerns, work distractions and lack of employee cooperation. Afterwards, participants were asked to rate the whole WAOW tool using four expectations and four concerns based on the literature review of Mettler and Wulf (2019). Finally, participants were asked to rate whether they would support the implementation of the WAOW tools´ sensors within their company on a 4-point scale (No, rather no, rather yes, yes).

3.2.3 Data Collection and evaluation procedure

Data was collected via an online survey. Participants were recruited in May 2021 and the recruitment took place in three European countries: Germany, the United Kingdom (UK), and Spain. Participants were contacted via a survey panel provider that accessed a random sample of panel members who met the inclusion criteria for the survey. Inclusion criteria for respondents was as follows: employed full-time, in a managing position, and at least 18 years of age.

In order to make sure that the data quality was acceptable, the collected data was examined using a multi-level review of the responses. Using Leiner's (2019) relative speed index with a limit of 2.0, participants who finished the questionnaire too quickly were eliminated. This was reinforced with an attention check adapted from Shamon and Berning (2020) which participants had to pass. Furthermore, the open response questions were reviewed for automated answers. Participants that did not pass these filters were excluded from the data evaluation.

3.2.4 Results

3.2.4.1 Willingness to Implement the WAOW tool

Figure 242 shows the percentage of managers who would like to implement the respective sensors of the WAOW tool within their company. 75% of all managers said they would implement the environmental sensor and 57% would implement the Wearables. The Camera (48%), Microphone (48%) and Eye Tracker (46%), were the least mentioned sensors to be implemented, however, all three sensors were still mentioned by almost half the managers.

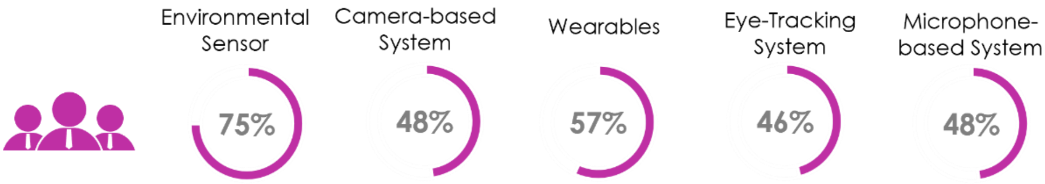


Figure 242: Overall number of managers who would implement the respective sensor in their company.

The percentage of managers who would like to implement the respective sensors of the WAOW tool, divided into manager´s respective countries are displayed in Figure 243. It can be seen that there is a slight descriptive tendency in that managers in the UK and Spain are more inclined to implement such sensors than managers in Germany. However, statistically significant differences are not measurable.

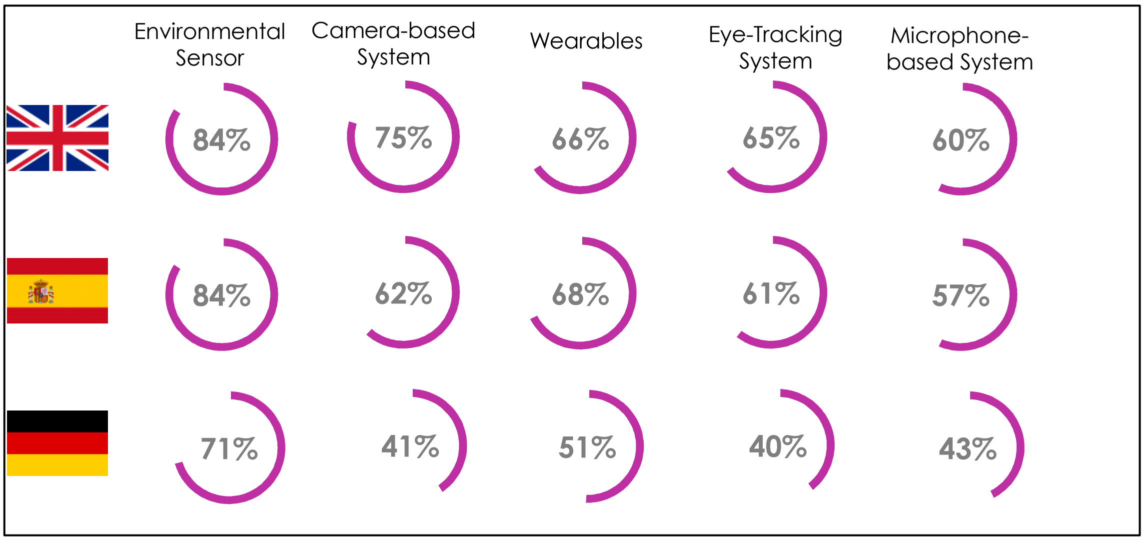


Figure 243: Number of managers per country who would implement the respective sensor in their company.

3.2.4.2 Expectations regarding the WAOW tool

Overall, the advantages of the WAOW tool are clear. All managers agreed that the implementation of a system like the WAOW tool would lead to reduced occupational safety risks (Germany: 3.5, Spain: 3.8, UK: 3.9), improved working conditions, due to a better identification and correction of a poor adaptation of working conditions and tasks to employees (Germany: 3.3, Spain: 3.8, UK:

3.9), as well as improved awareness of the negative effects of certain work practices on health and well-being (Germany: 3.3, Spain: 3.7, UK: 3.9). Most managers seemed to agree that the WAOW tool could aid in creating positive aspects, such as creating an incentive to change unhealthy working habits. As can be seen in Figure 244, both the UK and the Spain subsample show higher levels of support than the German subsample, however, the differences between the countries studied are minor.

Figure 244: Results of overall advantages of WAOW tool per county

3.2.4.3 *Concerns regarding the WAOW tool*

In line with the study by Schall et al. (2018), participants could indicate their single biggest concern about the use of the individual sensors in their company through an open response answer. Subsequently, managers rated the five scenarios according to three predetermined concerns, namely invasion of employees' privacy, distraction of employees from their work and a lack of employee compliance, on a 5-point ordinal scales (strongly disagree to strongly agree).

3.1.4.3.1 Open answers

In line with the study by Schall et al. (2018), participants could indicate their single biggest concern for each sensor that may arise from the implementation of the systems. Figure 245 shows users greatest concerns for all systems.

Figure 245: Results of open response for all sensor regarding concerns about the implementation of the WAOW tool ("*Please indicate your main concerns regarding the use of the systems listed below in your company*").

In Figure 245 we can see that amongst all three nationalities most managers indicated that they see "No concerns" with the implementation of the environmental sensor within their company (Germany: 88.2%, Spain: 94.6%, UK: 96.5%). German and Spanish managers indicated that their biggest concern was Data Privacy and having the system be used as a means of surveillance (Germany: 9.1%, Spain: 2.3%). English participants answered that their biggest concern was the accuracy of the system (1.8%). A few managers also mentioned concerns such as the functionality and costs of the system (Germany: 2.7%, Spain: 3.1%, UK: 1.8%), which are referred to as "Other".

Similar to the environmental sensor, most managers indicated that they see "No concern" with the implementation of the camera (Germany: 72.1%, Spain: 88.3%, UK: 91.5%). The next biggest concern was Data Privacy and having the system be used as a means of surveillance (Germany: 22.8%, Spain: 4.7%, UK: 5.1%). A few managers also mentioned concerns such as the invasion of privacy and security of the system (Germany: 5.1%, Spain: 7%, UK: 3.4%), which are referred to as "Other".

Most managers answered that they had "No concerns" with the implementation of wearables within their company (Germany: 82.4%, Spain: 95.5%, UK: 94.8%). Data Privacy and fear of surveillance were the most commonly mentioned concerns for this system (Germany: 13.5%, Spain: 2.2%, UK: 1.7%). Some managers also mentioned concerns such as the invasion of privacy and feeling as if they are controlled by the system (Germany: 4.1%, Spain: 2.2%, UK: 3.4%), these concerns are referred to as “Other”.

The implementation of the eye tracker was also met with most managers answering that they see “No concern” about the implementation of this system within their company (Germany: 79.6%, Spain: 91.9%, UK: 93.2%). The most mentioned concerns amongst the three nationalities are Data Privacy and fear of surveillance (Germany: 15.1%, Spain: 3.7%, UK: 2.5%). Concerns such as the system being an invasion of privacy as well as the implementation of the system being unnecessary are grouped together and referred to as “Other” (Germany: 5.3%, Spain: 4.4%, UK: 4.2%).

Similarly, to all other systems, most managers indicated that they had “No concerns” in having the microphone implemented into their company (Germany: 80.1%, Spain: 90.45, UK: 91.4%). Data Privacy and fear of surveillance were the second most mentioned concerns for the system (Germany: 16.6%, Spain: 6.6%, UK: 5.2%). Manager concerns such as invasion of privacy and infringement of personal rights were grouped together and referred to as “Other” (Germany: 3.3%, Spain: 2.9%, UK: 3.4%).

3.1.4.3.2 Invasion of my employees´ privacy

Invasion of employees´ privacy is particularly rated as relevant for the cameras and the microphone. In all countries, the greatest concerns are expressed in relation to these two sensors. The environmental sensor is rated best in all countries in terms of privacy concerns. The full results are shown in the figure below (Figure 246).

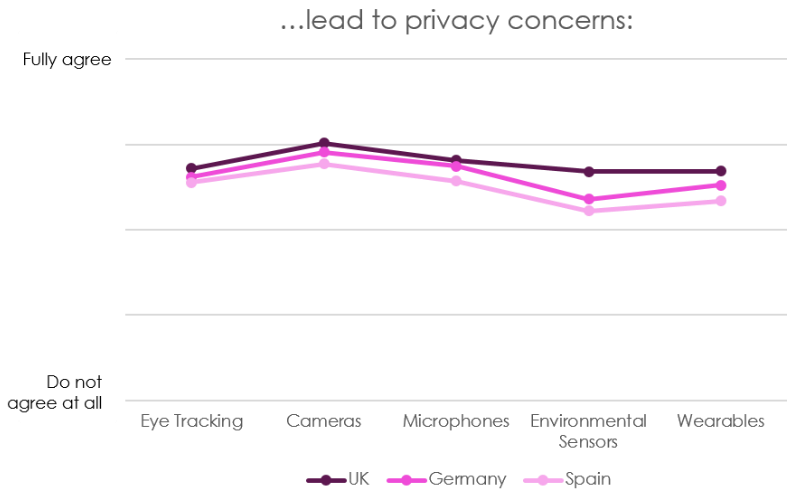


Figure 246. Results of managers privacy concerns for their employees about the use of the measurement system by country.

The results illustrate that cultural affiliation has an influence on privacy concerns regarding the usage of measurement systems for monitoring. The evaluation of the environmental sensor as well as the wearable differ significantly, however only a very small effect size is measurable (Table 33).

Table 33: Effect size of the different sensors that could lead to privacy concerns

Sensor type	F statistics
Environmental Sensor	$F(2, 884) = 9{,}965$, $p < .05$, $\eta_p^2 = .013$
Camera	$F(2, 884) = 1.758$, $p > .05$
Wearables	$F(2, 884) = 3{,}382$, $p < .05$, $\eta_p^2 = .008$
Eye-Tracking	$F(2, 884) = 0.777$, $p > .05$
Microphone	$F(2, 884) = 1.779$, $p > .05$

Post-hoc tests with Bonferroni correction show that regarding the usage of an environmental sensor for monitoring, participants from the UK differ significantly from those from Germany ($p < .05$) and from Spain ($p < .01$). Similarly, Germany differs significantly from Spain ($p < .05$), this shows that the UK has the greatest privacy concerns related to the environmental sensor ($M = 3.68$; $SD = 1.14$), followed by Germany ($M = 3.36$; $SD = 1.20$) and lastly Spain ($M = 3.22$; $SD = 1.15$) expressing the least privacy concerns.

Almost the same results were evaluated for the usage of wearables for monitoring. Post-hoc tests with Bonferroni correction show that regarding the usage of a wearable, the UK differs significantly from Spain ($p < .05$), whereas the UK and Germany, as well as Germany and Spain, show no statistically significant difference ($p > .05$). Participants from the UK have the greatest privacy concerns ($M = 3.69$; $SD = 1.08$) related to wearables, followed by Germany ($M = 3.52$; $SD = 1.15$) and Spain expressing the least privacy concerns ($M = 3.33$; $SD = 1.08$).

For all other sensors, related to the country, no statistically significant differences can be found in regard to the invasion of employees´ privacy. The results illustrate that the statistical differences occur for systems that are basically expressed as less intrusive with regard to privacy concerns. The evaluation of camera and microphone, which are rated as particularly intrusive, do not differ significantly.

3.1.4.3.3 Distraction of my employees from their work

Distraction from work by any of the systems studied is evaluated differently within each country. This concern is generally rated highest for all sensors by participants from the UK. Separated per sensor, in the United Kingdom, the eye tracker and the environmental sensor are rated as particularly distracting, while in Germany the camera followed by the eye tracker and in Spain wearables followed by the microphone are rated as distracting. In all cases, however, the differences are marginal. The full results are shown below (Figure 247).

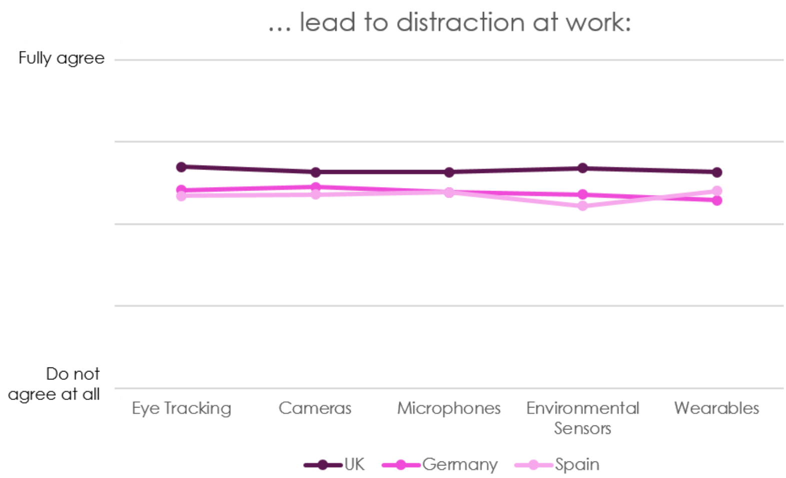


Figure 247. Results of the distraction of employees from work when using the measurement system of the UK, Spain, and Germany.

As already recognized for the privacy concerns, distraction from work due to a sensor is evaluated differently in the analysed countries. Statistically significant differences are measured for the environmental sensors as well as for the wearable (Table 34).

Table 34: Effect size of different sensors that could lead to work distractions

Sensor type	F statistics
Environmental Sensor	$F(2, 884) = 9{,}965$, $p < .01$, $\eta_p^2 = .022$
Camera	F(2, 884) = 2,343 , p > .05
Wearables	F(2, 884) = 5,471, p < .01 , $\eta_p^2 = .012$
Eye-Tracking	F(2, 884) = 0.777, p > .05
Microphone	F(2, 884) = 2,583, p > .05

Post-hoc tests with Bonferroni correction show that in regard to the usage of the environmental sensor for monitoring, participants from the UK differ significantly from Germany ($p < .01$) and Spain ($p < .01$). Germany also differs significantly from Spain ($p > .05$). Respondents from the UK see the greatest impact in the distraction from work by the environmental sensor ($M = 3.68$; $SD = 1.14$), followed by Germany ($M = 3.35$; $SD = 1.20$) and Spain ($M = 3.22$; $SD = 1.15$) expressing the least concerns regarding distraction from work.

Regarding the Wearables, post-hoc tests with Bonferroni correction show that only the UK differs significantly from Germany and Spain ($p < .05$), whereas participants from Germany and Spain show not statistically significant difference ($p > .05$). The UK has the greatest concerns regarding distraction from work ($M = 3.63$; $SD = 1.07$) related to the wearables, followed by Spain ($M = 3.40$; $SD = 1.15$) and Germany expressing the least concerns regarding distractions ($M = 3.36$; $SD = 1.20$).

For all other sensors, related to the country, no statistically significant differences can be found with regard to distraction from work.

3.1.4.3.4 Lack of my employee compliance

Lack of employee compliance refers to the concern, that employees will not use the system or will use it incorrectly. This concern is generally rated highest for all sensors by participants from the UK. If the individual sensors are considered, it initially becomes apparent that all sensors are rated similarly. Descriptively, however, the concerns are named in particularly for the eye tracker, even if the differences are only slight (Figure 248).

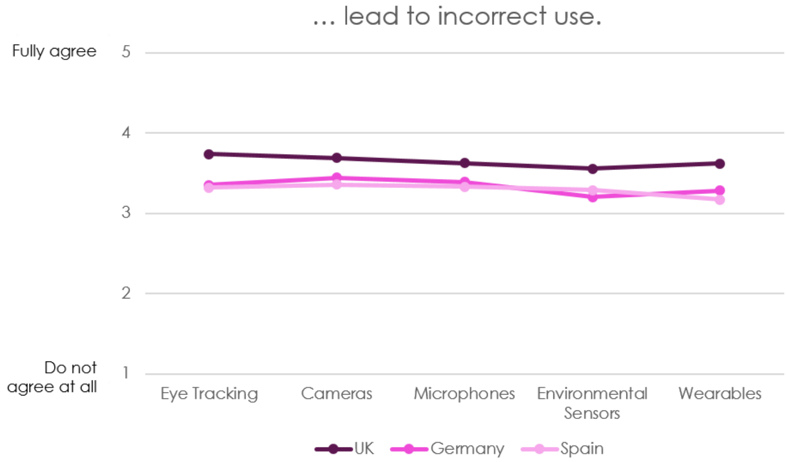


Figure 248. Results of the concerns of the employees´ about the lack of interaction with the measurement systems from the UK, Germany, and Spain from "strongly disagree", to "disagree", to "neither", to "agree" to "strongly agree".

As for the other concerns, the results illustrate that cultural affiliation has an influence on usage behaviour. In terms of usage, there are significant differences for the environmental sensor, the camera, wearables, and the eye tracker between the countries analysed, however only very small effect sizes are measurable (Table 35):

Table 35: Effect size of different sensors that could lead to lack of employee compliance

Sensor type	F statistics
Environmental Sensor	$F(2, 884) = 5{,}521$, $p < .005$, $\eta_p^2 = .012$
Camera	$F(2, 884) = 3{,}710$, $p < .005$, $\eta_p^2 = .008$
Wearables	$F(2, 884) = 6{,}242$, $p < .001$, $\eta p^2 = .014$
Eye-Tracking	$F(2, 884) = 4{,}387$, $p < .005$, $\eta_p^2 = .010$
Microphone	$F(2, 884) = 2{,}803$, $p > .005$

In summary, it can be said that the UK differs significantly from the other two countries, in that it considers this concern to be more serious than the other two countries. This was also reported before in the descriptive evaluation.

3.2.4.4 *Associations with system support*

The results described above were further used to compare them with the question regarding the support of the WAOW tool. The following Figure 249 gives an overview of participants' ratings of the investigated system specific concerns towards sensor-based health risk monitoring. The colour gradient depicts the relative frequency of binned system support across all five systems for the participants who gave the respective rating for the particular concern, with darker values corresponding to higher frequencies of support.

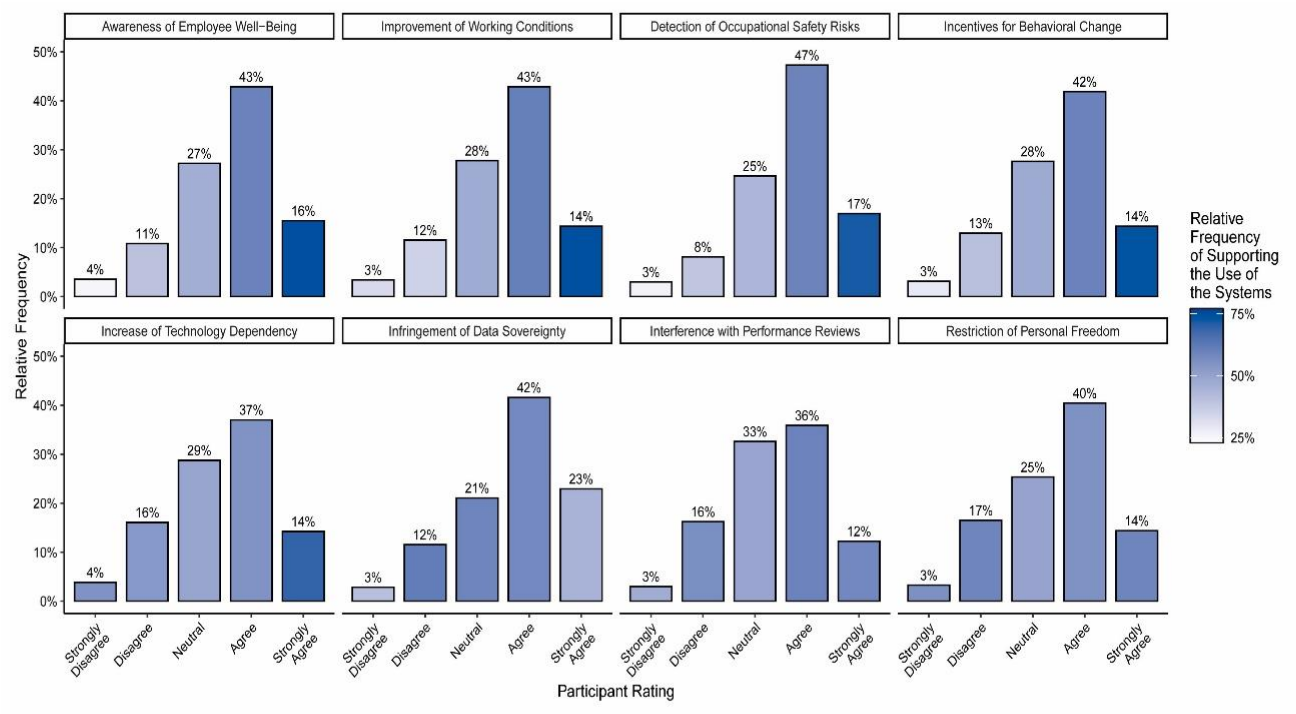


Figure 249: Overview of participants rating of system specific concerns

For system-specific concerns, the colour gradient shows a close monotonic decrease of support among participants with increasing intensity of the respective concern. The most pronounced decrease in support exists for those who are highly concerned about privacy invasions. A similar pattern can be seen for all four general expectations which yield strict monotonic increases in support with increasing expectation intensity. Conversely, the relationship between the general concerns and health risks monitoring support seems less clear.

3.2.5 Conclusion

Monitoring employee health risks has been a focus of human factors research in recent decades. Yet, our understanding of the factors that influence employers' and employees' decisions to support its implementation is very limited. Therefore, this study examined the overall willingness to implement sensor-based health risks monitoring systems as well as prevalent expectations and concerns among managers. We are pleased to note that managers have a very positive attitude towards the use of sensor-based health risks monitoring systems and that almost every second respondent would support the use of such a system in their own company. We also found that only minor cultural differences were discernible, and these were only descriptive; no statistically significant differences were measurable.

The general expectations for system support were shared by the majority of the sample. These results are promising as they suggest that a relevant proportion of managers recognize the benefits that health risks monitoring can provide. More importantly, all four expectations are significantly positively related to managers' system. This underlines that the expectations queried are appropriate for increasing support among organizational decision makers. By emphasizing the importance of expected benefits to system adoption, the results are consistent with existing literature regarding both technology adoption in general and IoT device adoption in particular.

For concerns, a negative correlation was demonstrated between privacy and distraction concerns and system support. One of the main findings of this study is the high importance of privacy concerns, as these concerns were the most prevalent in the quantitative assessments, by far the most frequently mentioned in the open-ended responses and showed the strongest relationship with system support. Although the role of privacy in IoT device adoption has been discussed previously, the estimated magnitude of this relationship exceeds general reports in the literature regarding the potential sensitivity of workload data as well as the intrusiveness of sensor systems such as cameras and microphones. The problem of individual privacy is exacerbated in the workplace because there is an inherent power imbalance between employers and employees, with employers deciding whether to deploy new technological solutions. Therefore, the implementation of workload monitoring systems will require privacy measures such as anonymizing data and limiting data access to balance the intended benefits with preventing opportunities for abuse. It is important that employees do not perceive workload assessments as a loss of control over personal data and that sufficient trust is established in the employer's data processing practices.

As a summary, the study presented is one of the first to examine the expectations and concerns of relevant stakeholders regarding the implementation of sensor-based workload monitoring systems. In principle, the opinion of the sample studied is very positive. The four specific expectations of increased awareness of employee well-being, improved working conditions, identification of occupational safety risks, and the ability to incentivize behaviour change were shown to be strongly associated with managers' willingness to support workload monitoring in their organizations. In contrast, the concerns examined differed in their association with system support, with the risk of invasion of employee privacy emerging as the strongest barrier to managers' support for implementation.

3.3 Research Question 2 – User Insights

3.3.1 Introduction

This second study aims at contributing to the second research questions, in which the expectations and concerns of potential users are queried, to analyse factors which support the usage of sensor-based health risk monitoring systems.

The aim of this study is, therefore, to examine the users' view of the extent to which and under what conditions they would use the WAOW tool.

Research Objectives

This study aims at identifying the expectations and concerns of potential users regarding the WOAW-Tool, as user acceptance of technology is considered an essential factor for the effective utilization of any technology, product, and process (Venkatesh & Bala, 2008). Examining expectations and concerns of this user group, allows us to assess if there is an overlap in concerns for managers and WAOW tool primary users, what aspects should be added and how these influence the willingness in using the tool. Therefore, the main research questions are:

1. Which expectations and concerns do potential users have about sensor-based health risk monitoring?

3.3.2 Method

3.3.2.1 Questionnaire

Users were given an overview of the sensors that are used in the WAOW tool. Here, all sensors were depicted and described regarding how they work, as well as the way in which health-related aspects of the working person are measured to support him or her in working healthily and safely in the long term. To best address privacy concerns, participants received a specific explanation on how the data would be processed, that the data collected would only be available to the person using it and could not be shared with or viewed by others. First, the respondents were asked which of the presented sensors they would like to use at their workplace. Afterwards, an overall rating of the WAOW tool was asked as well as the behavioural intention (BI) to use the WAOW tool. Both were recorded with the aid of one item: "I think the system is useful", to assess the overall rating of the WAOW tool, and "If I had the system available, I would use it" to assess the behavioural intention. In addition, we collected participants' demographics (age, gender, technical affinity, form of employment, professional experience, and work hours per week).

3.3.2.2 Procedure

Data was collected via an online survey. Participants were recruited in January 2022 which took place in Germany. Participants were contacted via a survey panel provider that accessed a random sample of panel members who met the inclusion criteria for the survey. Inclusion criteria required respondents to be at least 18 years of age and not older than 67 years (retirement age in Germany) and to be employed full time. The sample was selected to be age and gender representative for the German working population. Respondents received monetary compensation for their participation.

3.3.2.3 Data analysis

To ensure sufficient data quality for the analysis, a multistage screening of the user responses was performed. First, participants with implausible completion times were excluded, using the relative speed index with a lenient cut-off of 2.0 as criterion (Leiner, 2019). Second, an attention check item was included which participants had to pass (Shamon & Berning, 2020). As a result, 104 participants were excluded, and 493 participants remained for the further analysis.

3.3.2.4 Participants

The overall sample size of the study included N = 493 participants. Participants had to be at least 18 years old, hold an office job, as well as work full-time to be included in the evaluation. Even though all participants worked in an office setting, participants were asked to categorize themselves into 5 predetermined job descriptions. Participants were able to choose and identify with the following job descriptions: Upper Management (5.8%), Manager (12.2%), Team Leader (15.4%), Employee (65.1%) and other (1.6%). Users were also asked to answer if they predominantly worked from home (Home Office) or in an office.

Overall, 224 females (45.5%), 265 males (53.8%) and three participants identifying as other (0.6%) answered the questionnaire. The average age of participants was 43.8 years with a standard deviation of 12. Table 36 shows an overview of the sample characteristics.

Table 36: Overview of sample characteristics

		N	*%*
Gender	Female	224	45.4
	Male	266	54.0
	Other	3	0.60
Age	18 - 29	87	17.6
	30 - 39	97	19.7
	40 - 49	133	27,0
	50 - 59	128	26.0
	60 - 67	48	9.7
Position	Upper Management	28	5.7
	Management	60	12.2
	Team Leader	76	15.4
	Employee	321	65.1
	Other	8	1.6
Main workplace	Home-Office	224	45.4
	Office	269	54.6
Main work tasks	Knowledge work	110	22.4
	Routine work	144	29.3
	Conferences and meetings	65	13.2
	Mixed activities	174	35.4

Professional experience [1 – very low to 7 – very high] (*M/SD*)	5.884 (1.127)
Working hours / week [hours] (*M/SD*)	40.078 (4.805)
Technical affinity [1 - no affinity to 5 - complete affinity] (*M/SD*)	2.781 (1.338)

3.3.3 Results

3.3.3.1 Willingness to use the WAOW tool

Respondents were asked to indicate whether or not they found the WAOW tool useful and if they would implement the WAOW tool in their work environment if it were available to them. The first descriptive finding is that more than half of all participants find the tool to be useful and are in favour of implementing the system in their work environments. These results can be seen in Figure 250.

Figure 250: Evaluation of user responses that find the WAOW tool useful and would implement it into their workplace

Users were also asked to choose which sensors they would like to have in their WAOW tool. As described above, the subjects were informed about the functioning of each sensor and the data collected.

Among the most selected sensors are the environmental sensor (44%), the activity tracker (41%), the neurometrics wristband (38%) and the body posture analysis (35%). The results for all sensors can be seen in Figure 251.

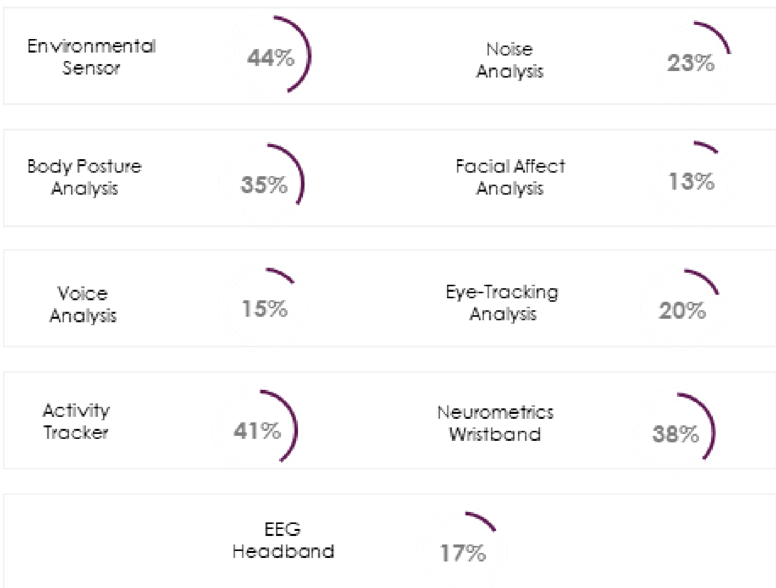


Figure 251: Evaluation of user responses as to which sensors they would like to use in their WAOW tool

3.3.3.2 Expectations regarding the WAOW tool

Users were asked to answer via an open-response question what they find to be the best aspect of the WAOW tool. These results were analysed and can be seen in Figure 252.

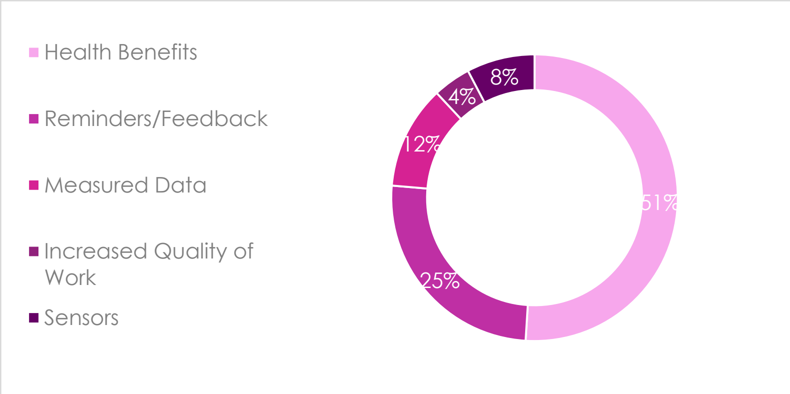


Figure 252: Open response question – "What is the best aspect of the WAOW tool?"

Figure 252 shows user responses to what they believe is the best aspect of the WAOW tool. A little bit more than half of the users (51%) indicated that the health benefits that the system brings is the best aspect of the tool. The second most mentioned aspect were the Reminders/Feedback (25%) the tool would give users throughout the day, and the third best aspect of the tool was the data that the tool measures (12%).

3.3.3.3 Concerns regarding the WAOW tool

Users were also asked to answer via an open-response question what would prevent them from implementing of the WAOW tool within their workplace. These results were analysed and can be seen in Figure 253.

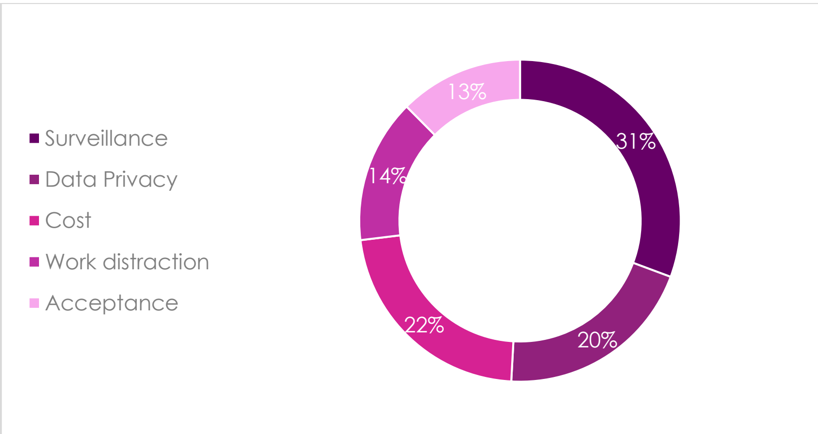


Figure 253: Open response question – "What would prevent you from using the system in your company?"

Figure 253 shows the open response answers of users regarding what would prevent them from implementing the tool within their workplace. The most mentioned user prevention was surveillance (31%), users indicated that they believed the tool would be used to control and monitor their work. Users also indicated that the cost of the tool (22%) and data privacy concerns (20%) would hinder them from implementing the tool within their company.

3.4 Research Question 3 – Work Design

3.3.4 Conclusion

The aim of this study was, after the view of managers, now also to capture the view of actual users in order to be able to compare them with each other.

Taken together, the results are very positive. Almost 2/3 of the respondents recognize the benefits of the tool and consider it useful for their own workplace. A few fewer users would actually implement the system in the workplace. The concerns of the benefits provide information about the concerns related to the implementation, which are especially about data protection and privacy aspects. In contrast, however, there is a strong conviction that the tool provides health support, with one in two respondents stating this. With regard to the individual sensors, it seems that contactless sensors in particular are preferred, sensors, which do not have to be worn directly on the body; the activity tracker is an exception here, which can probably be explained by the fact that users are familiar with wearing an activity tracker or a watch. In this context, it should be noted that the participants in this study were probably confronted with sensor-based health risk monitoring systems for the first time. The fact that the individual sensors could not actually be tried out made it more difficult for participants to evaluate them, as they thus had to evaluate potential rather than real systems. This may have led to lower acceptance, especially for those sensors that are less familiar to the general population (e.g., EEG headband and eye tracking), which is a well-known phenomenon in the acceptance research literature.

3.4.1 Introduction

This third study aims at contributing to the third research questions, in which the work design, namely job demands and resources, are involved in the willingness to use the WAOW tool at work.

User acceptance can be described as the relevance for one's own work and the possibility that using a particular system will improve the user's work or job performance. This concept of perceived usefulness (PU) is considered a key factor in explaining the behavioural intention to use technology. Since PU depends on the relevance to one's own work, it can be assumed that job demands, and resources contribute to how the system is evaluated. Previous research shows that PU corresponds with a willingness to use advanced mobile devices (Y.-S. Wang et al., 2006) and mobile health systems (Wu et al., 2007). However, studies that have been conducted in a work context are very limited.

Related to health risk monitoring at the workplace, previous studies have predominantly focused on predictors related to the individual (the working person), such as organizational commitment, organizational identification, monitoring attitudes, fairness, invasion of privacy or surveillance and perceived control (Abraham et al., 2019).

Research Objectives

Based on the previously described findings, this analysis was conducted in order to provide initial insight into how work demands, and resources affect the behavioural intention to use the WAOW tool while working. The theoretical approach states that external variables will first affect perceived usefulness and thus indirectly influence behavioural intention. This direct and indirect effect will be examined for job demands and resources as external variables. Therefore, the main research question is:

1. How does the acceptance differ depending on the workplace (Office vs. Home-Office workplace)?

3.4.2 Method

3.3.2.1 Questionnaire

The questionnaire explained above (section 3.3.2.1) was composed of two parts. First, the WAOW tool was presented as it is described in section 3.3.2.1. All sensors as well as the whole WAOW tool were depicted, including a description how they work as well as the way in which health-related aspects of the working person are measured to support him or her in working healthily and safely in the long term. In addition to the WAOW tool's general assessment and behavioural intention, four items were used to measure the perceived usefulness (PU) of the WAOW tool. The items were selected based on the corresponding scale of the Technology Acceptance Model and adapted to the context of health risk monitoring systems (e.g. "Such a system can help improve health habits") (Venkatesh et al., 2016).

3.3.2.2 Procedure, Analysis & Participants

Procedure and sample characteristics are described in section 3.3.2.

3.4.3 Results

In addition, a comparison was made to determine the extent to which the workplace affects the acceptance of the entire WAOW tool as well as the individual components. For this purpose, the respondents were asked to indicate whether they work mostly from home or predominantly in an office.

Respondents were asked on a 6-point scale to indicate whether they would implement the WAOW tool in their work environment if it were available to them. The first descriptive finding is that more than half of all participants (54.5%) are in favour of implementing the system in their work environments.

Implementation of the WAOW tool is slightly more likely to be sought in a normal office environment than in a home office (56% vs. 53%) (Figure 254).

Differences between office and home-office workplaces were analysed using an unpaired t-test. There was no statistically significant difference between participants working at home and participants working at normal offices ($t(491) = -.207$, $p = .836$).

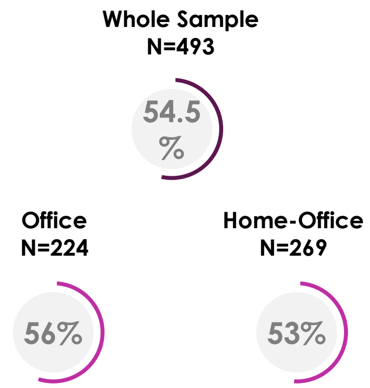


Figure 254: Percentage of respondents who would like to implement the WAOW tool in their work environments, depending on their workplace and work tasks.

Respondents were asked to choose which sensors they would like to have in their WAOW tool. As described above, the subjects were informed about the functioning of each sensor and the data collected.

Among the most selected sensors is the environmental sensor (44%), the activity tracker (41%), the neurometrics wristband (38%) and the body posture analysis (35%). The descriptive evaluation shows that there are differences between the two workplaces studied with regard to the use of the individual sensors (Figure 255). Therefore, all sensors were tested for significant correlations regarding the workplace. A chi-square test was used to compare workplace and acceptance. No expected cell frequencies were below 5. Results show that the activity tracker is significantly correlated with the workplace, $\chi^2(1) = 4.783$, $p = .029$). However, the effect size is negligible (CC = .098; Cramer's V = .098).

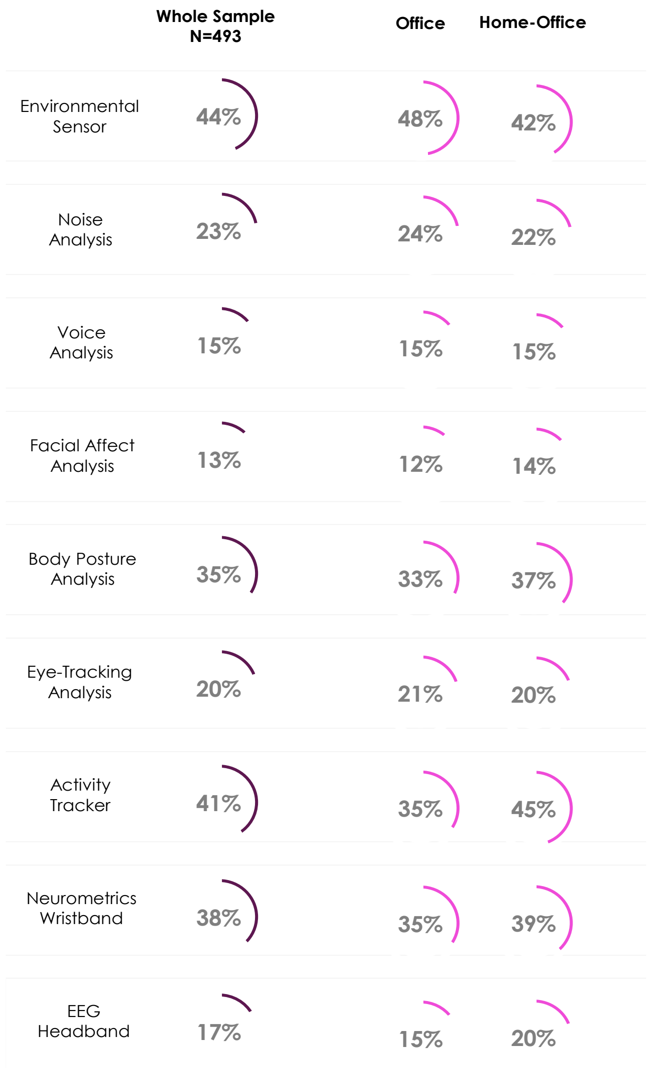


Figure 255: Percentage of respondents who would like to use sensor during their working day, divided according to their prior workplace.

3.4.4 Conclusion

In summary, it can be seen that the acceptance regarding the WAOW tool is very high, every second person interviewed would like to implement and use the WAOW tool at their workplace. The physical working environment has no significant influence on the willingness of the implementation of the WAOW tool, not even on the selection of the tool's components. Furthermore, the primary work task is not related to the desire to implement the WAOW tool or to the selection of components. Despite this, a descriptive tendency can be seen that working persons with many meetings and conferences want to implement the WAOW tool more often and would like to use sensors more often than others. Overall, 77% of respondents who would like to implement the tool, but of those 77% only 52% of knowledge and routine workers would want to use the tool. One explanation for this may be that the study was conducted in times of the Corona pandemic, meetings therefore take place predominantly virtually, and people who have to attend many meetings during their working day are therefore already frequently using cameras and microphones in their normal working environment. This could lead to fewer concerns regarding the use of such systems.

3.5 Summary: Secondary & Tertiary Users

As can be seen in Figure 256, Managers tended to rate the implementation of sensors within their workplace higher compared to possible users of the tool. However, both Managers and Users rated the Environmental sensor (Managers: 75%, Users: 44%) and Wearables (Managers: 57%, Users: 32%) the highest, in terms of implementing them into their workplace. Overall, we can see that Managers tend to be more willing to implement specific sensors of the WAOW tool, whilst the actual users of the tool are a bit more hesitant towards specific sensors, such as microphones (Managers: 48%, Users: 19%).

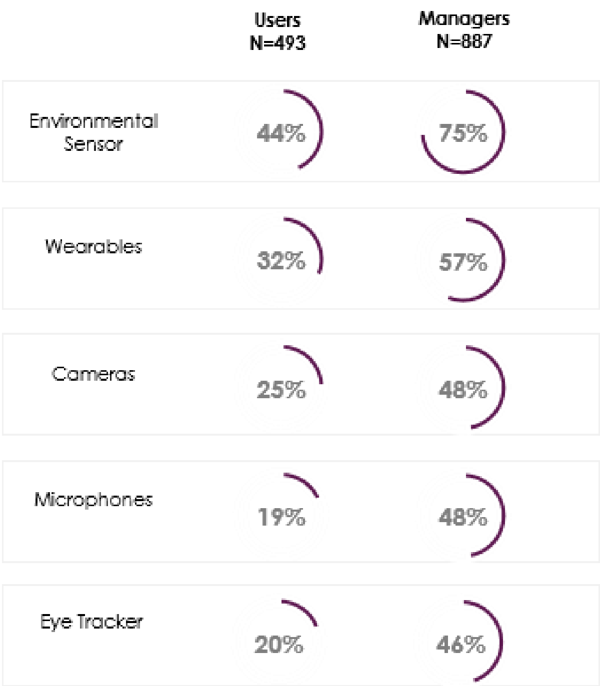


Figure 256: Implementation of sensors from the WAOW tool, divided into Users and Managers

Specific user concerns also show differences between Managers and Users of the tool. In Figure 245 and Figure 253, we can see that both Managers and Users share concerns in regards to the tool being used. However, most Managers indicated that they had “No concerns” about the implementation of the tool. Users, on the other Hand, mentioned concerns such as costs, distractions as well as acceptance. The differences in Managers and User concerns shows that Managers are more open to trying to implement the WAOW tool within their workplace, whilst Users show more hesitancy towards it. It is also interesting to note that very few managers mentioned “Cost” as a legitimate concern, however, it seemed to be a big concern amongst Users.

On the basis of the above-mentioned surveys, we were able to help identify the wishes and concerns of both sides, the users and the supervisors, in order to derive meaningful conclusions for the successful implementation of such technologies at work. The results suggest that users and managers generally see the benefits of such technologies, especially when they are aware of relevant health issues. Both sides are particularly aware of data protection and data security. This aspect was recognized and taken into account at an early stage in the WorkingAge project. As the comparison to the data of the primary users shows, the processing in the WorkingAge project could be implemented successfully, as the primary users showed only few concerns about data security. At this point, therefore, a limitation of this survey study with tertiary users should also be mentioned. Most respondents were likely encountering the nature of a workplace health risk monitoring system for the first time, as this is not a practice in common workplaces. This situation makes it more difficult for participants as they evaluate a potential system rather than a real one. This may have led to a different level of acceptance, as participants may have lacked expertise and specific interest (McClendon, 1991) and as surveyed in an online survey format the likelihood of approval can be lower (Weijters et al., 2008). In addition to the limitations specific to this study, there is a need for further research on monitoring health in real workplaces. This would not only allow an investigation of the real-world applicability of methods developed in controlled environments (Alberdi et al., 2016), but it would also improve the basis for analysing stakeholder attitudes toward such systems. Finally, it could provide the basis for needed long-term studies examining how opinions about occupational health risk monitoring change over time of system use (Gorm & Shklovski, 2016).

As a summary, the present studies conducted with secondary and tertiary users of the WAOW tool focused on the factors that support the use and implementation of the WAOW tool among managers as well as employees. Individual expectations and concerns regarding the overall system as well as the individual sensors were considered, and additionally, working conditions were included to provide a more nuanced picture regarding the use of health risk monitoring technologies. The results presented may contribute towards a better understanding of the potential of modern technologies in supporting everyday work to help manage physical and psychosocial demands and improve mental and physical health of workers around the globe.

4. Conclusions

Our aim of this deliverable is to assess the sustainability, accessibility, usability, validity and usefulness of the WorkingAge system according to the standards and indicators of the criteria set out in the evaluation protocol (D9.1). the following points were addressed:

1. The evaluation of the benefits for occupational health and quality of life from the primary users' perspective (section 2.3), including a pre-questionnaire (questionnaire to assess the situation before using the WAOW tool) and a post-questionnaire (questionnaire to assess the situation after using the WAOW tool). The results illustrate a positive change in the area of physical working conditions, especially for office and teleworking workplaces. Positive psychosocial changes, on the other hand, are measurable for production workplaces; the same applies to general well-being and the assessment of quality of life. Overall, the picture is positive despite the shortened time of use; the users recognise the advantages of the tool and the positive effects on their own health and well-being.
2. The evaluation of the entire WAOW tool as well as individual components from the perspective of primary users in short and long-term testing (section 2.4), taking into account factors of usability and acceptance of the tool, as well as an evaluation of all individual sub-modules, including interventions. The results obtained are very promising. As described, the usability of the tool has been significantly improved since the in-lab tests. Furthermore, the usability shows no significant differences in terms of user groups or use cases, which is a very good sign for a broad applicability and illustrates that the WAOW tool has successfully mastered the difficult task of an adaptive solution for a wide range of use cases and also with regard to the older target user group. Particularly noteworthy are also the positive evaluations with regard to ethical, legal and social implications (ELSI). Neither social or ethical implications nor concerns about data protection and data security are perceived as seriously as they appeared to be in the in-lab tests. In fact, the ELSI-aspects could even be strengthened in the course of use, so that the users considered social, ethical and legal implications considerably less after the longer period of use. A very good sign that speaks for transparency and security towards the WAOW tool.
3. The evaluation of the validity and usefulness of the individual components of the WAOW tool based on the data collected with the primary users during short and long-term field tests (section 2.5 - 2.19). With regard to the sensors, it must of course first be mentioned that the in-company tests had to be carried out in a more limited way than was initially planned. Despite the difficult circumstances and numerous challenges, however, almost all sensors were deployed in real workplaces and used by the users. Unfortunately, due to the prolonged period where wearing facemasks was obligatory, and additionally some

technical difficulties, the facial affect analysis submodule could not be used; and some problems arose with regard to data storage, which is why some sensors were used but the subsequent evaluation is limited due to missing data, although they were used by the users and the data were fed into the WAOW tool. The sensors used were perceived positively by the users, as also shown by the assessment in the evaluation. While it is clear that some submodules still need to be revised, many of the problems could have been improved with a longer period of use and can be attributed to pandemic-related difficulties. It is worth highlighting that with this project we have created a fully autonomous WAOW tool that can be deployed anywhere without interfering with the host organisation's IT system. Once deployed, the WAOW tool retrieves, collects, stores and processes data on-site or nearby. This allows to answer important aspects of data security, sovereignty and privacy. A user-friendly interface was created that had to unite many technologies, which was also the difficulty, as the app thus represents the centre of the WAOW tool and has to regulate through any failure of the network, communication, sensors or other system elements. Important findings in this regard could be collected and also remedied in the short-term tests; the long-term tests helped to identify further potential for improvement, which will led to significant improvements.

4. Finally, the deliverable provides an evaluation of the WAOW tool from the perspective of secondary and tertiary users (section 3). Worker health risk surveillance has been a focus of human factors research in recent decades. However, we know very little about the factors that influence employer and employee decisions to support implementation of a system like the WAOW tool. Therefore, as part of the WorkingAge project, we wanted to explore the overall willingness to implement the WAOW tool. On the one hand, we surveyed decision makers (managers, health professionals, human resource professionals, etc.) and, as a second step, potential users. Encouragingly, we find that managers are very positive about the use of WAOW-like systems and that almost every second respondent would support the use of such a system in their company. On the basis of the surveys conducted with secondary and tertiary users, we were able to help identify the wishes and concerns of both sides, the users and the supervisors, in order to derive meaningful conclusions for the successful implementation of WAOW-like systems at work.

Despite the difficulties and challenges that the last few years have brought us, as a consortium we have been able to continue the development of the WAOW tool and apply it in real working environments. In three different use cases, the WAOW tool has been evaluated in terms of functionality and reliability, but also in terms of usability, acceptability and impact on health, well-being and effectiveness. Many emerging problems could be overcome, some questions are still open and need further research which can be done on the basis of the collected data. Nevertheless, the evaluation by the users was very positive and encouraging and thus offers in summary a positive picture of the WAOW tool.

5. References

Abraham, M., Niessen, C., Schnabel, C., Lorek, K., Grimm, V., Möslein, K., & Wrede, M. (2019). Electronic monitoring at work: The role of attitudes, functions, and perceived control for the acceptance of tracking technologies. *Human Resource Management Journal*, *29*(4), 657–675. https://doi.org/10.1111/1748-8583.12250

Alberdi, A., Aztiria, A., & Basarab, A. (2016). Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review. *Journal of Biomedical Informatics*, *59*, 49–75.

FER-2013: Facial Expression Recognition 2013 Dataset, Available at: https://www.kaggle.com/datasets/msambare/fer2013

Borghini, G., Bandini, A., Orlandi, S., Di Flumeri, G., Arico, P., Sciaraffa, N., Ronca, V., Bonelli, S., Ragosta, M., Tomasello, P., Turhan, U., Acikel, B., Ozan, A., Imbert, J. P., Granger, G., Benhacene, R., Drogoul, F., & Babiloni, F. (2020). Stress Assessment by Combining Neurophysiological Signals and Radio Communications of Air Traffic Controllers. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, *2020*, 851–854. https://doi.org/10.1109/EMBC44109.2020.9175958

Chaquo Ltd (Ed.). *Chaquopy: Python SDK for Android*. https://chaquo.com/chaquopy/

Charles, R. L., & Nixon, J. (2019). Measuring mental workload using physiological measures: A systematic review. *Applied Ergonomics*, *74*, 221–232. https://doi.org/10.1016/j.apergo.2018.08.028

Cheong, J., Kalkan, S. & Gunes, H. (2021) The Hitchhiker's Guide to Bias and Fairness in Facial Affective Signal Processing: Overview and techniques, IEEE Signal Processing Magazine 38 (6), pp. 39-49, 2021. https://ieeexplore.ieee.org/abstract/document/9591544

Churamani, N., Kara, O. & Gunes, H. (2022) Domain-Incremental Continual Learning for Mitigating Bias in Facial Expression and Action Unit Recognition, IEEE Transactions on Affective Computing, 2022. https://ieeexplore.ieee.org/document/9792455

Churamani, N., Kalkan, S. & Gunes, H. (2021) AULA-Caps: Lifecycle-Aware Capsule Networks for Spatio-Temporal Analysis of Facial Actions, Proc. of 16th IEEE International Conference on Automatic Face and Gesture Recognition, 2021. https://ieeexplore.ieee.org/document/9666978

De Raedt, L., Kimmig, A., & Toivonen, H. (2007). *ProbLog: A Probabilistic Prolog and Its Application in Link Discovery*. https://www.aaai.org/papers/ijcai/2007/ijcai07-397.pdf

Gorm, N., & Shklovski, I. (2016). *Sharing steps in the workplace: Changing privacy concerns over time*.

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(9), 904–908. https://doi.org/10.1177/154193120605000909

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology. Human Mental Workload* (Vol. 52, pp. 139–183). Elsevier. https://doi.org/10.1016/S0166-4115(08)62386-9

Hupont, I., Tolan, S., Gunes, H. & Gómez, E. (2022) The landscape of facial processing applications in the context of the European AI Act and the development of trustworthy systems, Nature Scientific Reports, volume 12, Article number: 10688 (2022).

https://www.nature.com/articles/s41598-022-14981-6

Jäger, R. (2004). Konstruktion einer Ratingskala mit Smilies als symbolische Marken. *Diagnostica*, *50*(1), 31–38. https://doi.org/10.1026/0012-1924.50.1.31

Leiner, D. J. (2019). Too Fast, too Straight, too Weird: Non-Reactive Indicators for Meaningless Data in Internet Surveys. *Survey Research Methods*, *13*(3), 229–248. https://doi.org/10.18148/srm/2019.v13i3.7403 (229-248 Pages / Survey Research Methods, Vol 13 No 3 (2019)).

Luo, C., Song, S., Xie, W., Shen, L., & Gunes, H. (2022). Learning Multi-dimensional Edge Feature-based AU Relation Graph for Facial Action Unit Recognition. arXiv preprint arXiv:2205.01782.
https://doi.org/10.48550/arXiv.2205.01782

McClendon, M. J. (1991). Acquiescence and recency response-order effects in interview surveys. *Sociological Methods & Research*, *20*(1), 60–103.

Mettler, T., & Wulf, J. (2019). Physiolytics at the workplace: Affordances and constraints of wearables use from an employee's perspective. *Information Systems Journal*, *29*(1), 245–273. https://doi.org/10.1111/isj.12205

Muggleton, S., Raedt, L. de, Poole, D., Bratko, I., Flach, P., Inoue, K., & Srinivasan, A. (2012). Ilp turns 20. *Machine Learning*, *86*(1), 3–23. https://doi.org/10.1007/s10994-011-5259-2

Nelles, J., Kohns, S., Spies, J., Bröhl, C., Brandl, C., Mertens, A., & Schlick, C. M. (2017). Best-Practice Approach for a Solution-Oriented Technology Assessment: Ethical, Legal, and Social Issues in the Context of Human-Robot Collaboration. In *Advances in Ergonomic Design of Systems, Products and Processes* (pp. 1–14). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-53305-5_1

Or, C. K. L., & Karsh, B.-T. (2009). A systematic review of patient acceptance of consumer health information technology. *Journal of the American Medical Informatics Association*, *16*(4), 550–560.

Sauro, J. (2011). *A practical guide to the system usability scale: Background, benchmarks & best practices.*

Schall, M. C., Sesek, R. F., & Cavuoto, L. A. (2018). Barriers to the Adoption of Wearable Sensors in the Workplace: A Survey of Occupational Safety and Health Professionals. *Human Factors*, *60*(3), 351–362. https://doi.org/10.1177/0018720817753907

Shahidi, B., Haight, A., & Maluf, K. (2013). Differential effects of mental concentration and acute psychosocial stress on cervical muscle activity and posture. *Journal of Electromyography and Kinesiology : Official Journal of the International Society of Electrophysiological Kinesiology*, *23*(5), 1082–1089. https://doi.org/10.1016/j.jelekin.2013.05.009

Shamon, H., & Berning, C. (2020). *Attention Check Items and Instructions in Online Surveys: Boon or Bane for Data Quality?* https://doi.org/10.2139/ssrn.3549789

Song, S., Jaiswal, S., Shen, L., & Valstar, M. (2020). Spectral representation of behaviour primitives for depression analysis. *IEEE Transactions on Affective Computing.* https://doi: 10.1109/TAFFC.2020.2970712

van Acker, B. B., Bombeke, K., Durnez, W., Parmentier, D. D., Mateus, J. C., Biondi, A., Saldien, J., & Vlerick, P. (2020). Mobile pupillometry in manual assembly: A pilot study exploring the wearability and external validity of a renowned mental workload lab measure. *International Journal of Industrial Ergonomics*, *75*, 102891. https://doi.org/10.1016/j.ergon.2019.102891

Vecchiato, G., Cherubino, P., Maglione, A. G., Ezquierro, M. T. H., Marinozzi, F., Bini, F., Trettel, A., & Babiloni, F. (2014). How to Measure Cerebral Correlates of Emotions in Marketing Relevant Tasks. *Cognitive Computation*, *6*(4), 856–871. https://doi.org/10.1007/s12559-014-9304-x

Venkatesh, V., & Bala, H. (2008). Technology Acceptance Model 3 and a Research Agenda on Interventions. *Decision Sciences*, *39*(2), 273–315. https://doi.org/10.1111/j.1540-5915.2008.00192.x

Venkatesh, V., & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science*, *46*(2), 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926

Venkatesh, V., Thong, J. Y., & Xu, X. (2016). *Unified Theory of Acceptance and Use of Technology: A Synthesis and the Road Ahead.*

Wang, X., Lavender, S. A., Sommerich, C. M., & Rayo, M. F. (2022). Exploring the relationships between computer task characteristics, mental workload, and computer users' biomechanical responses. *Ergonomics*, 1–10. https://doi.org/10.1080/00140139.2022.2026490

Wang, Y.-S., Lin, H.-H., & Luarn, P. (2006). Predicting consumer intention to use mobile service. *Information Systems Journal*, *16*(2), 157–179. https://doi.org/10.1111/j.1365-2575.2006.00213.x

Weijters, B., Schillewaert, N., & Geuens, M. (2008). Assessing response styles across modes of data collection. *Journal of the Academy of Marketing Science*, *36*(3), 409–422.

Wu, J.-H., Wang, S.-C., & Lin, L.-M. (2007). Mobile computing acceptance factors in the healthcare industry: A structural equation model.

International Journal of Medical Informatics, *76*(1), 66–77. https://doi.org/10.1016/j.ijmedinf.2006.06.006

Xu, T., White, J., Kalkan, S. & Gunes, H. (2020) Investigating bias and fairness in facial expression recognition, Proc. of the European Conference on Computer Vision Workshops, pp. 506-523, 2020. https://link.springer.com/chapter/10.1007/978-3-030-65414-6_35

Zhang, X., Yin, L., Cohn, J. F., Canavan, S., Reale, M., Horowitz, A., … & Girard, J. M. (2014). Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database. *Image and Vision Computing*, 32(10), 692-706.

https://doi.org/10.1016/j.imavis.2014.06.002

Zhao, J., Li, R., Liang, J., Chen, S. & Jin, Q., Adversarial Domain Adaption for Multi-Cultural Dimensional Emotion Recognition in Dyadic Interactions, Proc. of the 9th International on Audio/Visual Emotion Challenge and Workshop, Pages 37–45, 2019.