



Team usability testing: development and validation of a groupware usability evaluation method

Dalma Geszten¹ · Balázs Péter Hámornik¹ · Károly Hercegi¹

Received: 30 June 2023 / Accepted: 26 March 2024 / Published online: 17 April 2024
© The Author(s) 2024

Abstract

As virtual teamwork became widespread, the importance of the usability of groupware has increased, as groupware must have high usability in order to properly support teamwork. Several studies have focused on the usability evaluation of groupware, yet, both in terms of science and practice, there is a lack of a method that, like existing empirical single-user methods, can be routinely applied in software development practice. The new Team Usability Testing method differs from previous methods in that it is an empirical method for evaluating synchronous (real-time) collaborative software that explores the team usability problems with the help of real or potential users. The method consists of questionnaires, screen recording videos and group interviews, and the framework of data processing is based on the theory of the mechanics of collaboration. In the four steps of the development of the method, empirical studies under laboratory conditions and real working conditions, and then, for validation purposes, analytical, heuristic evaluation with the involvement of usability experts were performed. During the development and the application of the method, the communication patterns of different teams were examined in terms of their relationship to software usability. In this paper we present the results of the second laboratory study and the heuristic evaluation in details. Overall, the method is able to explore team usability problems of different types of teams, making it a valuable element in existing methods for evaluating the usability of groupware.

Keywords HCI design and evaluation methods—laboratory experiments · HCI design and evaluation methods—usability testing · Computer supported cooperative work · Empirical studies in collaborative and social computing

1 Introduction

The strengthening trend toward remote work cause the digital transformation of organisations. The technological support of this transformation is important, because providing the right groupware support for a team is one of the key factors of team effectiveness (Aldag and Kuzuhara 2015). In other words, technology is a crucial element of team performance, especially in virtual teams, so it is essential to be

supported by the right groupware (Kirkman and Mathieu 2005; Martins et al. 2004). Therefore, the need for highly usable groupware is increasing, which include the growing demand for highly usable groupware evaluation methods.

The results presented in this paper are part of a bigger research project, which is related to the usability evaluation of groupware. The main goal of the project was to create a groupware evaluation method, which is able to explore team usability problems.

During the development of the Team Usability Testing method, our research questions were the followings:

- What factors influencing groupware usability can the Team Usability Testing reveal?
- What types of team usability problems can the Team Usability Testing reveal?
- What types of team usability problems can the field study reveal compared to the laboratory study?
- What types of team usability problems can the laboratory study reveal compared to heuristic evaluation?

✉ Dalma Geszten
geszten.dalma@gtk.bme.hu

Balázs Péter Hámornik
hamornik.balazs@gtk.bme.hu

Károly Hercegi
hercegi.karoly@gtk.bme.hu

¹ Department of Ergonomics and Psychology, Faculty of Economic and Social Sciences, Budapest University of Technology and Economics, Műegyetem Rkp. 3, 1111 Budapest, Hungary

- What is the relationship between usability problems and a team's communication patterns?

The method was created in four research phases, from which the last two phases are presented in this paper (Citations deleted to maintain the anonymity of the review process). First, groupware usability is defined, along with the demonstration of groupware evaluation. Then the results of two research phases: a laboratory study and a heuristic evaluation is discussed. Finally, future research possibilities will be presented.

2 Related work

2.1 Groupware usability

Generally, software usability is a part of user experience (Hassenzahl 2007; Hassenzahl and Tractinsky 2006) and is related to the software's efficiency, effectiveness, learnability and satisfaction (Nielsen 1994).

According to the ISO 9241–11 standard usability is: "the extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use". In new definitions the meaning of effectiveness is also related to appropriateness, besides accuracy and completeness. In addition, "efficiency has been redefined in the revised standard as the resources (time, human effort, costs and material resources) that are expended when achieving a specific goal (e.g., the time to complete a specific task)" (Bevan et al. 2015; ISO 2015). Furthermore, the definition of satisfaction (positive attitudes and no discomfort) was extended to include emotional and physiological effects (either positive or negative) while using the product. (Bevan et al. 2015).

Groupware is a special type of computer technology, a multi-user computer system that helps users to collaborate (Gutwin and Greenberg 2000; Salomón et al. 2019, p. 11). Ellis et al. (1991) define groupware as follows: "computer-based systems that support groups of people engaged in a common task (or goal) and that provide an interface to a shared environment" (Ellis et al. 1991, p. 40). The authors state that groupware is a spectrum, there is no rigid line between groupware and non-groupware systems. It depends on the extent to which the system supports two important dimensions: common task and shared environment.

There are several types of groupware, which can be categorized in many ways: time (synchronous/asynchronous) and space (personal/remote), group size (small to large groups), type of group tasks (e.g. planning, decision-making), characteristics of the group (e.g. group composition), type of software or hardware and collaborative functionalities (e.g. screen sharing, file/document sharing, synchronous work

on documents) (Bafoutsou and Mentzas 2002). As stated in Ellis et al. (1991)'s work, groupware systems can be categorized based on the groupware's primary functionalities and goals. The authors highlight that these categories can overlap. Examples include, but are not limited to message systems, multiuser editors, group decision support systems and electronic meeting rooms, computer conferencing, intelligent agents and coordination systems.

Groupware usability, according to Pinelle et al. (2003) is defined as the extent to which groupware enables teamwork to be carried out—efficiently, effectively and satisfactorily—for a given team's particular collaborative activity (Pinelle et al. 2003). According to the authors the three important aspects of groupware usability is effectiveness, efficiency and satisfaction (Gutwin and Greenberg 2000).

Effectiveness refers to whether a collaborative action has been carried out successfully and to the number and severity of errors associated with the action. A groupware with high usability support the users in collaborative actions, thus the number of (solved) usability problems and breakdowns are in connection with this phenomena. Efficiency refers to the time or the users effort to carry out a collaborative action and to solve a common task in the shared workspace. A highly usable groupware support collaborative actions to happen quickly. Satisfaction refers to whether team members are relatively satisfied with the outcome of the collaboration and the collaborative process supported by the groupware (Gutwin and Greenberg 2000).

2.2 Groupware usability evaluation

The scientific discourse on the evaluation of groupware usability started already in 1988, when Grudin summarized the difficulties of evaluating groupware (Grudin 1988). Afterwards, intensive research on the topic started, with researchers experimenting with different methods. Pinelle's (2000) research summarizes these "early" studies on the evaluation of groupware, but at the end of his article he makes a sharp criticism: the studies were either not documented at all (only reporting results) or not well documented. This makes it difficult to reproduce the research in practice and difficult to verify their scientific reliability. Pinelle therefore also points to the need for new groupware evaluation methods that are time and cost effective (Pinelle 2000). In addition, although researchers have developed many different low-cost methods, they have not been widely disseminated. In the next section, we present some of these methods for illustration purposes.

The Evaluation Working Group (EWG) framework is a low-cost method for collaborative software evaluation. The framework distinguishes four levels of groupware: requirement (e.g.: "requirements generated from the tasks being performed by the group"), capability ("functionality that is

needed to support the different requirements"), service ("services ... that can be used to support the capabilities needed in CSCW systems"), and technology ("specific implementations of services") (Cugini et al. 1997, pp. 9–10). The first step of the method is to design collaborative scenarios based on the four levels. The method then evaluates the scenarios along various metrics and measures to determine the usability of the collaborative software. Some examples for metrics include: countables, task completion, time, user ratings, and conversational constructs, while for measures are: task outcome, cost, user satisfaction, awareness, collaboration management and breakdowns (Cugini et al. 1997; Damianos et al. 1999).

Fuks et al. (2005) have created the 3C model, which they recommend for use when designing and evaluating groupware. The 3C refers to communication, coordination and cooperation, with an emphasis on awareness. The 3C model represents collaboration in group work as an iterative cycle, where communication affects coordination, coordination affects cooperation and cooperation affects communication. The central element of the model is awareness, which affects all three elements simultaneously. At the same time, the correct functioning of the groupware functions associated with each C element increases awareness.

Along with the development of low-cost evaluation methods, a large body of research focuses on how certain features (mostly related to awareness) affect usability. Early research focused on which techniques are most likely to enhance awareness in the collaborative workspace (e.g. radar view, telepointer) (Gutwin and Greenberg 1996; Gutwin et al. 2004). Later, researchers have investigated how exactly awareness techniques enhance software usability (Gutwin and Greenberg 1998). Subsequent research has already looked at the impact on collaboration when the workspace is out of sync (Ignat et al. 2015); and exactly which awareness techniques should be used for different types of collaborative software (Lopez and Guerrero 2017). Furthermore, Collazos et al. (2019)'s study recommends a design framework for how to integrate different awareness support features into groupware. The framework consists of five phases and offers specific awareness support features to different aspects of the software (e.g.: for providing information about people's state emoticons, auditory icons, and avatars can be used) (Collazos et al. 2019).

Research has approached the evaluation of the usability of groupware from several directions. Analytical methods have evaluated groupware based on the knowledge of experts, but without involving real or potential users. These methods usually proposed a software design based on some kind of task model. A common feature is that their aim was to design the most usable groupware for a given organisation (Herskovic et al. 2009; Pinelle et al. 2003; Veer Van Der G and Welie Van M 2000). Analytical methods also

include research on expert analysis based on the mechanics of collaboration and on collaborative heuristics based on the mechanics (Baker et al. 2002; Pinelle and Gutwin 2008). The popularity of analytic methods in groupware research is underlined in Kutlu et al. (2021) review article. The article analyse groupware research papers from 2010 to 2020 and highlights that most studies use analytical methods (design science or conceptual modeling) related to groupware research (Kutlu et al. 2021). The greatest advantage of analytical methods that they can be used earlier in the groupware development process, than other methods.

Despite the advantages of analytical methods, many studies stress the importance of user involvement in the software development process (Heikkilä et al. 2021; Leso and Cortimiglia 2022; Parnell et al. 2021).

One early example of user involvement in groupware evaluation is the evaluation of the G^Roup Outline Viewing Editor (GROVE) collaborative text editor. Fifteen sessions were carried out with 3–6 participants in a variety of formats: face-to-face, distributed and mixed mode. The authors grouped design issues into four categories: problems related to group interfaces, group processes the groupware need to support, concurrency control and other system issues (Ellis et al. 1991).

Solano et al. (2016) suggest using another method, called CUEM (Collaborative Usability Evaluation Methods) for the usability analysis of groupware. According to CUEM, a combination of different methods should be used in a collaborative software usability evaluation situation, depending on the groupware analysis main goal and timeframe. The authors designed the CUEM method by examining three different types of groupware using seven usability evaluation methods and comparing the type and number of usability problems each method uncovered. The seven evaluation methods were: heuristic evaluation, cognitive walkthrough, formal experiment, constructive interaction, coaching method, interview, questionnaire. The authors suggest using a combination of inspection methods with test methods and differentiate between three method combinations: (1) Global Evaluation: heuristic evaluation+constructive interaction+interviews; (2) Specific Evaluation: Time Reduction: heuristic evaluation+coaching method+questionnaires; (3) Evaluation Focused on Specific Tasks: No Time Restrictions: cognitive walkthrough+formal experiments+questionnaires (Solano et al. 2016).

Another direction in the evaluation of the usability of groupware is represented by empirical methods that involve users in the evaluation of the usability of software in real, everyday working conditions. Field studies mostly investigate the impact of the groupware on collaboration (Gumienny et al. 2013; Tang et al. 1994) or what makes groupware successful within an organisation (Pipek and Wulf 1999; Vyas et al. 2013). Field studies use a variety

of methods, including observation, questionnaires, interviews, usability situations (scenarios) and log-file analysis (Christensen and Ellingsen 2016; Gumienny et al. 2013; Haynes et al. 2005; Marlow et al. 2016).

Besides, several attempts were made to automatize the data collection (Grigera et al. 2021) or the data analysis (Bringas et al. 2021) parts of groupware evaluation. The advantage of empirical methods is that by involving real or potential users, usability problems that are striking for them can be explored.

A different approach is represented by the mechanics of collaboration theory. While using groupware users must be able to perform individual and team-level tasks, because teams work processes consist of a mixture of these tasks. Therefore, groupware usability evaluation methods need to evaluate both individual and teamwork tasks (Pinelle and Gutwin 2002; Pinelle et al. 2003). In order to support the evaluation of individual and team-level tasks during a collaboration in a shared digital workspace, Gutwin and Greenberg (2000) created the mechanics of collaboration framework, which aims to offer an analytical framework for the evaluation of groupware. The novelty of the theory was that the authors argued that poor usability of groupware is caused by a lack of support for the basic collaborative actions, rather than by organisational or team factors. This point is still valid and crucial making the theory an important foundation of current research. These basic collaborative actions have been defined in several ways by the authors over the years:

"These activities, which we call the mechanics of collaboration, are the small-scale actions and interactions that group members must carry out in order to get a shared task done" (Gutwin and Greenberg 2000, p. 98).

"The mechanics of collaboration are the basic operations of teamwork—the small-scale actions and interactions that group members must carry out in order to get a task done in a collaborative fashion" (Pinelle et al. 2003, p. 287).

"A set of collaboration primitives that specify low-level actions that are needed to carry out a task in a shared manner, such as communicating with other members of the group, keeping track of what others are doing, negotiating access to shared tools or empty spaces in the workspace, and transferring objects and tools to others. T-CUA" (Pinelle and Gutwin 2008, p. 238).

To summarise the authors' definitions, collaboration mechanics are various basic actions that users must be able to perform on the shared collaborative workspace in order to solve a task together and collaborate successfully. They are like a system of mechanical parts working together as an efficient machine.

In 2003, the authors refined the original broadly defined mechanics to specific actions that can be observed and thus evaluated (Pinelle et al. 2003).

In summary, the theory of the mechanics of collaboration is useful because it allows us to decompose collaboration into different smaller actions, thus to analyse collaboration in terms of observable actions. Therefore, it can be applied to empirical and analytical studies, which will be demonstrated later (Pinelle and Gutwin 2002).

The main criticism of the theory is that the mechanics should be even more concrete, and should refer to specific actions to better use it in practice when designing and evaluating groupware (Pinelle and Gutwin 2008). To overcome this critique, the authors propose the combined use of several data collection methods, e.g., observation, interview, contextual interview (Pinelle et al. 2003). Another critique of the theory is that it works differently in synchronous and asynchronous collaborative situations, and in larger groups (Pinelle et al. 2003). This is not surprising, as the theory was originally developed for shared workspace groupware, which is characterised by users working together at the same time and in relatively small groups (3–7 people).

The framework was used in different evaluation situations: in a cognitive-walkthrough (Pinelle and Gutwin 2008), and in an analytical evaluation of early prototypes (Dew et al. 2015).

Moreover, Baker et al. (2001) developed usability heuristics for evaluating groupware, based on the Nielsen heuristics and the mechanics of collaboration theory as a synthesis (Baker et al. 2001). Rusu's (2016) work confirms that, in many cases, it is better to use specific heuristics that are better suited to the evaluated software rather than the general Nielsen heuristics. Although Nielsen's heuristics are easy to apply, in some cases alternative heuristics can detect more specific usability problems (Rusu et al. 2016).

Heuristics are rules of thumbs, user interface design guidelines. The eight heuristics developed for groupware evaluation are the following:

1. Provide the means for intentional and appropriate verbal communication
2. Provide the means for intentional and appropriate gestural communication
3. Provide consequential communication of an individual's embodiment
4. Provide consequential communication of shared artifacts
5. Provide protection
6. Manage the transitions between tightly and loosely-coupled collaboration
7. Support people with the coordination of their actions
8. Facilitate finding collaborators and establishing contact

Heuristic groupware evaluation method, which was developed specifically for evaluating the usability of shared visual workspace software is based on these heuristics. In 2002, the authors tested the applicability of the method and found

that novice/inexperienced evaluators could successfully use it (Baker et al. 2002).

As early as 2000, Gutwin and Greenberg propose to adapt the mechanics of collaboration theory to usability studies. Despite this, only one study (Dew et al. 2015) has applied the framework in this form, and that was only for individual testing. More research is therefore needed to confirm the usefulness of the framework in this form.

Due to the strong theoretical grounding of the framework and its practical applicability as discussed previously, we considered it appropriate to form one of the theoretical foundations of our research. Thus, the mechanics of collaboration theory is a key part of Team Usability Test, as it is one of the main theoretical frameworks for data analysis.

To summarize, although more than two decades have passed since the publication of Pinelle's (2000) article, there is still a lack of well-documented and rapid groupware evaluation methods despite the growing number of groupware. Therefore, the main goal of our research project was to develop a new method, which is able to explore the team level usability problems of groupware.

2.3 Earlier work related to the development of the team usability testing method

The development process of Team Usability Testing method consisted of four stages, in chronological order: first lab study, field study, second lab study and heuristic evaluation. Furthermore, based on the first and second lab study's data we conducted communication analysis to explore the relationship between usability problems and a team's communication patterns. The results of the first laboratory study, the field study and the communication analysis have already been published, therefore here we just summarize them shortly to facilitate the understanding of the whole method development process (Geszten 2021; Geszten and Hámornik 2023; Geszten et al. 2019, 2021, 2023).

Based on the results of the first lab study, the Team Usability Testing method can identify two types of factors affecting software usability: team usability problems and contextual factors. The team usability problems are related to awareness, explicit communication and the management of shared access. These are typical problems in real-time groupware (Gutwin and Greenberg 1998; Gutwin et al. 2004). Conversely, contextual factors refer to previous experience with the investigated groupware and team mood. The results show a relationship between contextual factors and team usability problems. The positive presence of contextual factors (more experience with the investigated groupware, positive mood during task solving) positively influenced collaboration and software usability. Although team usability problems appeared in all teams, they were handled differently by the teams. When problems arose, there were teams

that quickly solved them and overcame them. Conversely, other teams experienced the same problems causing serious communication and collaboration difficulties. As the problems were the same, in our future work we thought it important to investigate further exactly how each team differed. We did this by analysing team communication patterns using sequential analysis, which we will discuss later (Geszten and Hámornik 2023).

According to several literature sources, to support the validity of data analysis and interpretation, it is worthwhile to use multiple methods in the research process (Szokolszky 2020; Thurmond 2001). Therefore, we continued the development of the method with a field study. The field study (34 h in the field), which was based on participant observation and an interview, took place at a software development company. The participants were two user interface designers, who collaborated on the same project at the time of the observation using the same design groupware. The evaluation of the groupware took place under natural everyday working conditions, during which one of the researchers was in the role of an "observer as participant" (did not manipulate the events and did not interfere with the participants' work). She observed how the participants used the design groupware in their daily work. The data collected consist of written notes taken during the observation and an interview at the end of the observation. As in the first lab study, the field study also identified team usability problems and contextual factors that affect collaboration. The team usability problems observed in the Field Study were the lack of possibility to switch between different parts of the workspace and the visibility of notes. The contextual factors referred to the importance of the use of physical tools for collaboration, i.e., whiteboard and notebook, which play a significant role in understanding the process of collaboration. This is in line with the literature finding that collaborative software is characterised by usability problems due to inadequate support for the mechanics of collaboration and contextual problems (Steves et al. 2001). Therefore, the main finding of the field study was that team usability problems and contextual factors that affect collaboration, as revealed by the Team Usability Testing method, are valid aspects of software usability that exist in the field (Geszten et al. 2019, 2023).

We used sequence analysis to examine the communication patterns of the teams in the first and second lab studies. We considered this analysis necessary because the results of the lab studies showed that different types of team usability problems were typical for each team. The study of team communication is a significant topic in the psychological literature, but it is a less researched area in relation to software usability.

We chose the overwriting problem for analysis, which is justified by the fact that this is the most serious team usability problem, since in this case, one participant accidentally

(due to inadequate support for collaborative features) overwrites or deletes the work of another participant in the collaborative workspace. Overwriting problems were avoided by some teams and not by others, so we examined the nature of the difference in communication patterns. The study of team communication in the context of software usability is a poorly researched topic, so in our research we developed a code system for this purpose by merging team process theory and the mechanics of collaboration theory (Marks et al. 2001; Pinelle et al. 2003). Our results indicate that in those teams where there was discourse about awareness, no overwriting appeared. So if someone communicates or requests information about what is happening in the workspace, team members respond with this type of information. In addition, teams that have been effective in helping each other (if a team member asks for help, they get help) or have a tight organisation and planning of joint work (if a team member shares information about the organisation of joint work, they also receive this type of information in response), can avoid overwriting.

The novelty of the results is that we examined this phenomena in the context of software usability. As certain conflicts and problems occur in teams that communicate differently, this is also true for the usability testing situation, teams with certain communication patterns will have certain usability problems and different teams will encounter different types of problems in their communication. This can also have an impact on the interpretation of the usability test results, as teams' communication strategies can compensate for serious usability problems (Geszten and Hámornik 2023).

After a brief summary of the previous research phases, in this paper, the focus is on the results of the last two phases of the method development: the second laboratory study and the heuristic evaluation. Therefore, in the next sections the further development and the validation process of the new Team Usability Testing method will be discussed.

3 Method—the four stages of developing the team usability testing method

The process of developing the Team Usability Testing method consisted of four stages. As a first step, we conducted a laboratory study involving teams collaborating at the same time and at different locations. The main objective of the first laboratory study was to assess the types of usability problems the Team Usability Testing can reveal. In addition, it was also important to investigate the extent of useful data the different data collection methods can explore. As a next step, before the second laboratory study, we were interested in what kind of problems would occur under real, workplace conditions compared to controlled laboratory

conditions. Therefore, a field study was conducted before the second laboratory study. Summarizing the results and lessons learned from the field study and the first laboratory study, the second laboratory study followed, during which we analyzed a different groupware than in the first laboratory study. To investigate the relationship between usability problems and team communication patterns, sequence analysis was performed using a self-developed code system (based on the communication transcripts of the first and second laboratory studies). As a final step of the research project, we tested how results of expert analysis of given groupware compares to results of a laboratory study. Therefore, as the last part, a heuristic evaluation was performed.

4 Second laboratory study

As the third step of our research project, we conducted the second laboratory study, which was further developed based on the results and experiences of the first laboratory study and the field study. The subject of the second lab study was the Miro collaborative whiteboard software, which allows users to work simultaneously on the same workspace.

The research questions in the second lab study were:

- 1 What factors influencing groupware usability can the Team Usability Testing reveal?
- 2 What types of team usability problems can the Team Usability Testing reveal?

4.1 Participants

The study involved 11 teams, which are presented in detail in Table 1. As in the first lab study, participants were given two roles: collaborator or observer. In each case, a team consisted of three collaborators and in each case in the second lab study there were also observers. (In the second lab study, we invited four participants to a time slot, with the aim of ensuring that if someone did not come, the study could still be carried out. As in the first lab study, the roles were drawn from an envelope by the participants. Since all four participants came on each occasion, there were observer participants in each case).

The key characteristics of the participants are summarised in Table 1. The values in the table always refer to the team as a whole, whereas in the following description we present the total number of participants (observers included). In the second lab study, 10 males and 23 females participated in collaborator roles, aged 18–22 years (mean: 19.42, standard deviation: 1.27). In the before-task questionnaire, we asked participants "Do you prefer working alone or in a team?" (1-prefer individual work; 7-prefer teamwork). The combined mean for all teams was 4.00, standard deviation

Table 1 Characteristics of participants in the second laboratory study

	Male	Female	Age (mean)	Age (SD)	Team player (mean)	Team player (SD)
1	0	3	22	0	3	0
2	0	3	19	1	5.33	1.16
3	1	2	19.33	0.58	6	0
4	2	1	18.33	1.53	3.33	1.53
5	1	2	19	0	3.33	2.52
6	1	2	20	1.73	4.33	1.53
7	0	3	19	1	3	0
8	3	0	18.67	0.58	5.33	1.53
9	1	2	20	2	3.33	1.16
10	0	3	19.33	0.58	3.33	0.58
11	1	2	18.67	0.58	3.67	2.08

1.54. None of the participants in the second lab study had previous experience with the Miro collaborative whiteboard software before the study. As in the first lab study, there were observer participants in all cases in this study.

4.2 The subject of the test, the evaluated groupware

Miro is a collaborative online whiteboard software that provides a common visual workspace for collaborators, mainly to visualise different ideas and (work) processes (Fig. 1). In our experience, users of Miro are mostly students who use the software to visualise ideas, and UX researchers and

designers who use the software to visualise and brainstorm processes. In our experiment Miro was used as an online collaborative mind-mapping application. "A mind map is a multi-coloured and image-centred, radial diagram that represents semantic or other connections between portions of learned material hierarchically" (Eppler 2006, p. 203).

4.3 Procedure

The procedure for the study was the same as for the first laboratory study. The differences between the two studies are explained at the end of this section.

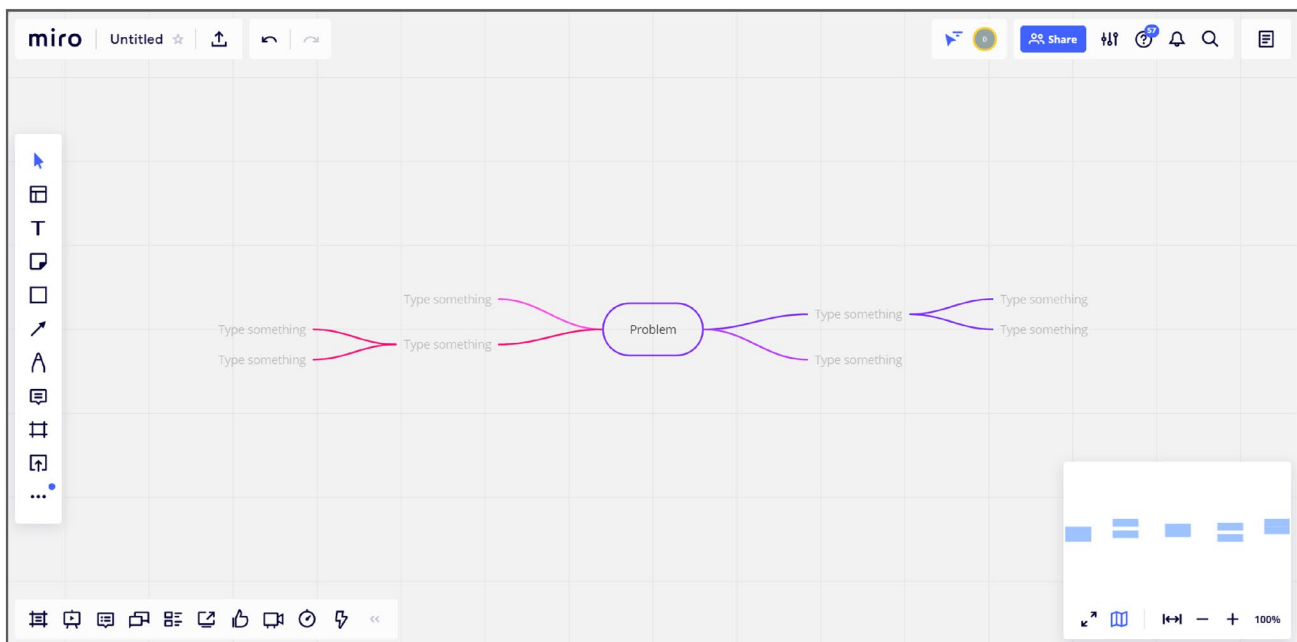


Fig. 1 Miro collaborative whiteboard software user interface, Miro (version 1.0.26)

During each session of the second lab study, the three collaborating participants were tasked with creating a shared visualisation in 30 min. Their task was to organise a university event and to create a Miro visualisation from their ideas. We chose this type of task because it is similar to the first lab study task, and it can be done in Miro in this time frame (30 min) and does not require any "special" skills to solve it. The exact questions for the before and post-task questionnaire used in the second lab study and the post-experiment group interview are available in Appendix 10.3, 10.4, 10.5.

Based on the preliminary experience of the first lab study and the field study, we developed the Team Usability Testing as follows. Observers were involved in the overall research process and received written instructions. In the first lab study, the observer participant did not complete the post-task questionnaire and did not participate in the group interview. Rethinking the research design, we considered it a major loss of information. The tutorial video on the collaborative software was removed from the research design based on feedback from the participants of the first lab study. According to the participants, this was not enough time for a comprehensive presentation, and they did not use it. During the first lab study (questionnaire and interviews), several participants mentioned that collaboration was influenced by how well they knew each other as teammates. Although all participants in the first lab study belonged to the same university group, there may be differences in the depth of their relationship, which can have an impact on cooperation. Therefore, in the second lab study, we asked team members in the before-task questionnaire how well they knew each other, so we have numerical data on this.

Besides, we also asked questions about team mood, since it was an important contextual factor, which played a role in successful collaboration in the first lab study. Two questions were asked related to team mood: "How activated was the team when working together?" (1–9) and "How pleasant was working together?" (1–9). Activation and pleasantness are the two dimensions of mood, according to Larsen and Diener's (1992) circumplex model, which was used successfully by Bartel and Saavedra (2000) to measure team mood.

4.4 Tools/instruments

In the second lab study, we used the same instruments as in the first lab study. We made video and audio recordings using the free screen recording software Open Broadcast Software (OBS). In the second lab study, the participants were in the same room, so we also recorded their discussion of task solution using a dictaphone. The before and post-task questionnaire was in Google Forms format. We also recorded the group interview using a dictaphone and a mobile phone voice recording application.

4.5 Steps of the analysis

The second lab study generated exactly the same type of data as the first lab study. So the data analysis is based on the following data:

- Questionnaire data (before and post-task questionnaire)
- Communication transcript (Communication transcript of the communication during collaboration: participants' discussion during the task)
- Post-experiment group interview transcript

The data from the second lab study was analysed using content analysis, as was the data from the first lab study. The teams' voice communication was transcribed from the on-screen behaviour videos of the collaboration. The post-experiment group interviews were also transcribed. The mechanics of collaboration framework was the base of our analysis: we used the different mechanics as separate codes for communication analysis and analysis of post-experiment interviews. We analysed the communication transcripts based on the mechanics of collaboration theory: first, we defined if an utterance is related to usability or not, then we categorized the utterances by the mechanics. After that we performed the same process with the interview transcripts and the questionnaire data. Observers' notes were not part of the analysis because of its various qualities. We would like to highlight that only a part of the mechanics of collaboration framework could be interpreted because of the characteristics of the experiment and the groupware. The mechanics of collaboration is a general framework for any type of groupware, and some mechanics can only be used when participants work in distributed or asynchronously (Geszten et al. 2021).

In this case, the development of the coding system was also carried out in a multi-step iterative process. The final code system is presented in Table 2. In the analysis of the data from the second lab study, we distinguished a total of nine factors influencing collaboration, of which five factors (awareness, avatar, zoom, overwriting, allocation and protection of shared workspace) are team usability problems and four factors (Miro expertise, acquaintanceship, task division and team mood) are contextual factors.

4.6 Results

In this section first, the descriptive statistics of the collaboration, then the results related to the usability evaluation of Miro and the results of the development of the Team Usability Testing method will be discussed.

Table 3 demonstrates the descriptive statistics of the teams' collaboration: collaboration time (until the participants reached a consensus that the work is done) and

number of ideas (effectiveness). The mean collaboration time is 33 min and 10 s ($M = 1990.09$, $S = 162.50$) and teams came up with an average of 11 ideas ($SD = 3.26$). There is no significant correlation between collaboration time and the number of ideas (Kendall-tau- $b = 0.17$;

$p = 0.477$). This may suggest that teams did not have more ideas because they spent more time collaborating.

In the content analysis, we identified two types of factors that influence collaboration: team usability problems and contextual factors. Table 4 provides an overview of the frequency of codes. In total, 131 factors affecting collaboration

Table 2 Second laboratory study–coding framework for factors affecting collaboration

Factors influencing collaboration			
Name of code group	Related mechanics according to (Pinelle et al. 2003)	Name of code	Description of code
Team usability problems	Information gathering—“basic awareness”	Awareness	The user’s knowledge of who, when and what they are doing in the common workspace during the collaboration
		Avatar	Representation of users in the shared workspace
		Zoom	The zoom in/ out function of the shared workspace
	Shared access to tools, objects, space and time–protect work	Overwriting	Inadvertently modifying or deleting each other’s work while working together
		Allocation and protection of shared workspace	Division of the shared workspace, who works where and efforts to protect the area
Contextual factors	–	Miro expertise	Previous experience with the software
		Acquaintanceship	Acquaintanceship of team members
		Task division	Task sharing among collaborators
		Team mood	The mood of the team during the collaboration

Table 3 Descriptive statistics of collaboration time and number of ideas

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
Collaboration time	36:19	36:14	31:09	30:27	31:02	37:41	34:31	33:23	32:49	31:49	29:27
Number of ideas	11	16	10	8	12	11	10	4	15	13	11

Table 4 Factors affecting collaboration during the second laboratory study

		Total occurrence	Frequency(%)	Average occurrence per team	SD	Min	Max
Team usability problems	Awareness	50	54.35	4.55	3.50	0	10
	Avatar	4	4.35	0.36	0.88	0	3
	Zoom	12	13.04	1.2	0.6	0	2
	Overwriting	15	16.30	1.36	1.61	0	5
	Allocation and protection of shared workspace	11	11.96	1	1.48	0	5
	Total	92					
Contextual factors	Miro expertise	25	64.10	2.27	1.66	0	5
	Task division	2	5.13	0.18	0.57	0	2
	Team mood	7	17.95	0.64	0.98	0	3
	Acquaintanceship	5	12.82	0.45	0.50	0	1
	Total	39					

were identified 92 (70%) team usability problems and 39 (30%) contextual factors.

For the team usability problems, awareness was the most common, occurring 50 times, or 54.35% of the time. This was followed by overwriting (16.3%), zoom (13.04%) and allocation and protection of shared workspace (11.96%). Of the contextual factors, Miro expertise (lack of) caused the greatest proportion of difficulties (64.10%). The occurrence of different factors affecting collaboration in each team is summarized in Table 5.

4.6.1 Team usability problems identified by the team usability testing

4.6.1.1 Team usability problems related to the basic awareness collaboration mechanic Support for basic awareness and the functionality to provide it is key for real-time groupware. This is supported by the results of the second lab study. Several teams highlighted that collaboration was greatly helped by being able to see who was where on the workspace and what they were doing. Teams 1, 7 and 11 also stressed that it was helpful to see every letter they typed when writing.

P2: I think it was also helpful that when someone was typing, it was visible as a letter, not just when they had finished typing, so that helped us to follow it. (Interview transcript–Team 7)

Awareness In the post-experiment group interviews and in the communication transcripts, a number of problems related to awareness were also raised. We will first present them in general terms, but in the next section we will discuss problems related to Miro's awareness features (avatar and zoom).

P2: It's in there. What? Where are you now?

P3: Above the points.

P2: Above the points, yeah, ahhaaa (Communication transcript–Team 6)

Avatars One of the features supporting awareness is the display of users on the workspace. This was done in Miro

with avatars. Several teams highlighted that avatars (5, 6, 11) greatly facilitated collaboration. Only one out of eleven teams noticed that the avatar disappeared from the workspace when the team member was searching for something on the web and not working on the workspace.

For some teams, the ability to name the avatar helped (Team 3, 4), but one team pointed out that this was not very helpful as they did not remember the names of their teammates (Team 3). Regarding the avatar's appearance, two suggestions were made: one participant would like to see avatars in a more contrasting colour, while another would prefer to see symbols instead of the current solution. In addition, participants would like the avatar to be visible not only in the editing workspace but also in other parts.

Zoom When solving the task, several teams had problems with the zoom function in relation to awareness (Team 1, 3, 4, 9, 11). This is unfortunate because it impeded the cooperation of the participants, as they cannot control what they see on the workspace.

P3: (...) But I can't see (...), why do you write it in such a small letter?

P1: It's not small, it's a 64 font size.

P3: Yeah, I'm just zoomed in. (Communication transcript–Team 3)

This problem was most severely evaluated by a participant from Team 4, who experienced it like "being thrown back and forth" by the software:

P1: What (s)he said first, that we don't see the same as the others: it happened to me a lot of times that I just saw one word mixed up or completely different, like I was thrown into these empty spaces (...) (Interview transcript–Team 4).

4.6.1.2 Team usability problems related to the management of shared access mechanic With this mechanic, all problems were basically related to the common editing of the shared workspace. Some of the teams identified the ability to work on a common workspace and edit it together at the same time as a factor that helped them to collaborate.

Table 5 Occurrence of factors affecting collaboration in different teams during the second laboratory study

		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
Team usability problems	Awareness	10	7	6	5	8	9	1	1	0	2	1
	Avatar	0	0	3	0	0	0	1	0	0	0	0
	Zoom	1	0	1	2	1	2	0	1	2	1	1
	Overwriting	5	3	0	2	3	1	1	0	0	0	0
	Allocation and protection of shared workspace	1	1	5	2	0	2	0	0	0	0	0
Contextual factors	Miro expertise	3	1	0	2	5	4	1	5	2	1	1
	Task division	0	0	0	0	0	0	0	2	0	0	0
	Team mood	0	0	0	1	0	0	1	2	3	0	0
	Acquaintanceship	0	0	0	1	0	0	1	1	1	0	1

However, for some teams (Teams 2 and 3) this was a source of uncertainty, fear of overwriting, accidental deletion or the fact that a teammate might delete something from their work (they have the possibility to do so). Interestingly, this type of problem did not occur in these teams. However, in other teams there were a number of other usability problems, even serious ones.

Overwriting Of the usability problems (subjectively, but also theoretically), one of the most serious problems that can occur in a shared workspace is overwriting. In Miro's case, this meant that participants edited the same object and eventually one of the works was deleted.

P3: Why has it disappeared now?

P1: Did you delete it?

P3: I didn't delete it.

P2: That wasn't on purpose (laughter). Isn't there an undo button?

P3: (laughter)

P2: There it is.

P3: Yes, there is! Yikes! I pressed it too, sorry! (Communication transcript–Team 1)

Allocation and protection of shared workspace Some teams made an effort to avoid the overwriting problems, in which case they explicitly divided up the workspace during collaboration or, if they noticed that someone was trying to edit into their work, they indicated this to their partner, thus avoiding overwriting.

Q3: (Name of participant), I don't understand anyway. Now, move it!

P2: Well, which one is which?

P1: I'll take the bigger one (Communication transcript–Team 3)

4.6.2 Contextual factors

4.6.2.1 Miro expertise Almost all teams mentioned that the novelty of Miro made collaboration difficult. Most teams overcame this difficulty, although for one Team 7 member, the learning process was an unexpected and surprising experience.

P1: The software was new to all three of us, so it took some getting used to and figuring out how it worked (Questionnaire–Team 10).

Several teams pointed out that it helped that they had used similar software before and that Miro was similar to other software such as PREZI, Paint or MS Word. Two teams (Team

5 and Team 6) had serious difficulties using Miro. None of the participants in these teams had used similar types of software before. Team 5 therefore felt Miro was unmanageable, a software that did not do what they wanted it to do.

P5: Okay, so I find the task to be ingenious, but I think the system within which it has to be implemented is unmanageable (...).

P1: That the screen was jumping around, or that I couldn't insert text boxes with it, or that if I did, it was either very small or it was spreading the whole screen and I had to look for where I could delete it. (Interview transcript–Team 5).

4.6.2.2 Task division Among the contextual factors, the division of tasks was mentioned by only one team as a complicating factor. Factors affecting the usability of the groupware, which appear in a single team, are also important in the analysis of software usability, and therefore, although rare, contribute to the completeness of the analysis.

P5: we should have defined it better at the beginning ... but I think it would have been more efficient if we had discussed that well, you do that and then you do this and then afterwards within that how he is going to implement it (...) just that it was declared that way, that he does this, then it would have been better (Interview transcript–Team 8).

4.6.2.3 Acquaintanceship Almost all teams highlighted that the fact that they had known each other before (even just by sight) and the common life situation also helped them to collaborate.

There were few teams where a low level of acquaintanceship made collaboration difficult. One such team was Team 9, whose participants were completely unknown to each other.

One of the questions in the before-task questionnaire was also used to get information about how well the team members knew each other. We placed a coloured sheet of paper in front of each team member's computer, and the question "How well do you know the participant sitting at the red/yellow/blue computer?" appeared in the before-task questionnaire referring to this colour. Participants could answer the question on a scale of 0 to 10, with 0 being "We just met for the first time" and 10 being "We are best friends". Table 6 summarises the participants' responses.

The results of the Kruskal–Wallis test show that there is no significant difference in terms of acquaintanceship

Table 6 Aggregated questionnaire responses to the question "How well do you know the participant sitting at the red/yellow/blue computer?" (scored from 0 to 10)

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
Team mean	11.67	10	10	6.67	7.33	6.33	5.33	4.66	0	14.67	0.67

between teams $\chi^2(2) = 18.19, p = 0.052$. The test is not significant, but the results support the idea of testing this question with a larger number of teams.

4.6.2.4 Team mood Based on the results of the first lab study, we assumed that team mood plays a role in the success of collaboration, so we also asked about this phenomenon in the second lab study in separate questionnaire questions. The qualitative results show that, overall, the teams describe the collaboration as a positive, fun, entertaining and creative experience.

Please describe the collaboration in a few words.

P2: Fun, help, support

P3: Purposeful, fun

P1: I think it was good, we talked a lot, we brainstormed.

(Questionnaire–Team 1)

In contrast, when examining the quantitative questionnaire data, there is a significant difference between teams in the extent to which they perceived the collaborative work as pleasant, based on the results of the Kruskal–Wallis test [$\chi^2(2) = 21.43, p = 0.018$] (Table 7).

In addition, there is a significant difference between teams in how they felt they were able to collaborate with their teammates [$\chi^2(2) = 19.29, p = 0.037$]. Teams 7, 9 and 4 differ most from the other teams in a negative direction. However, the Kruskal–Wallis test shows no significant difference in how activated participants feel the collaboration was [$\chi^2(2) = 17.06, p = 0.073$]. There is no significant correlation between the number of ideas (effectiveness) and the rating of collaboration (Kendall's tau- $b = 0.06; p = 0.809$). In addition, how activated and pleasant participants felt when working together also showed no significant correlation between the number of ideas (effectiveness): Kendall's tau- $b = 0.176, p = 0.471$; Kendall's tau- $b = 0.218, p = 0.377$.

4.7 Discussion

4.7.1 Summary of the second laboratory study

In this section, a short summary of the results of the second laboratory study is presented. The detailed discussion of the results is presented in the 6. Discussion section. Similar to the first lab study and the field study, two types

of factors emerged in the second lab study: team usability problems and contextual factors. As in the first lab study, team usability problems are related to the basic awareness and the management of shared access collaboration mechanics. However, the explicit communication problem did not appear here, due to the fact that in the second lab study the participants were working in the same space and could talk to each other without technical barriers. As for contextual factors, team mood and task division also appeared in the second lab study as factors influencing collaboration. However, acquaintanceship also appeared as a new important supporting factor. To summarize, the results of the first and second lab studies are very similar to each other, therefore for real-time groupware evaluation, this is the kind of result that the team-level usability testing method can reveal.

4.7.2 Results on the development of the team usability testing

Based on the results of the first lab study and the field study, we carried out an improved version of the Team Usability Testing method. The changes relate to the role and instruction of the observers and the omission of the tutorial video, as well as the added acquaintanceship and team mood questions of the questionnaire.

The presence of independent observers added the most value during the post-experiment group interviews. They often brought up aspects of how they saw a phenomenon from the outside and this was well related to by the collaborating participants. The exact written instructions of the observers did not influence the quality of their written notes, it seems that how well someone can relate to this task depends on some other factor (probably the personal motivation of the participant). Leaving the tutorial video out from the process sped up the whole test situation. Overall, all teams were able to manage the task without it, although some teams mentioned that they would have liked some kind of comprehensive presentation. The question of acquaintanceship in the questionnaire provides important additional numerical information on how well the participants know each other, which can play an important role in the interpretation of the data, just like the questions related to team mood.

Table 7 Aggregated questionnaire answers to the questions "How do you think you could collaborate with your teammates?"(1–7); "How activated was the team when working together?"(1–9) and "How pleasant was working together?"(1–9)

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
Collaboration	6.00	6.67	9.50	5.00	6.00	7.00	3.67	6.00	3.67	7.00	6.67
Activated	7.67	8.33	8.00	5.33	7.67	8.67	3.00	8.33	7.33	8.33	6.33
Pleasant	9.00	8.67	8.67	6.67	8.33	9.00	4.33	8.33	7.33	9.00	7.67

In the steps of the method development so far, we have tested the method in laboratory conditions and in field conditions. As a validation of the method, we were interested to see what types of problems experts identify in a heuristic evaluation compared to the problems identified in the laboratory studies.

5 Validation of the team usability testing method

In the previous research phases, we developed a Team Usability Testing method, which is suitable for exploring team usability problems of groupware.

In the final phase of the research project, we validated it by comparing the results of the method we developed with those of the Nielsen heuristic evaluation method. We chose Nielsen's method as the basis for comparison because its effectiveness is scientifically supported (Lazar et al. 2017; Nielsen 1994; Rubin and Chisnell 2008). Instead of the original heuristics, the evaluators used heuristics suitable for groupware evaluation (Baker et al. 2001). The analysis was conducted by four experts with different professional backgrounds, who analysed the Miro software, which was also tested in the second lab study. (Although Miro has been upgraded from version 1.0.26 to 2.0.1 by the time the heuristic evaluation was performed, this only includes minor improvements that did not affect the collaborative features). This gave us the opportunity to observe the types of usability problems identified by the experts and to compare how the Team Usability Testing method is similar and different.

Our main research question in the heuristic evaluation was:

What types of team usability problems can the laboratory study reveal compared to heuristic evaluation?

5.1 Participants

Four evaluators with different academic and industry experience participated in the analysis. In order to ensure that as wide a range of professional perspectives as possible is

represented, the evaluators' industrial and academic experience is diverse, and this was deliberately sought in the selection process. Some evaluators have more academic (Evaluator 1), while others have more industrial experience (Evaluator 2). Two evaluators have equal experience in both academia and industry, but to different degrees (Evaluators 3 and 4). The evaluators also have varying degrees of experience in UX research, ranging from less than 5 years to more than 15 years of experience. All evaluators use a variety of different groupware (e.g. Google products, Slack, Trello, Asana, Miro, Prezi, MS Teams, Github, Jira, Figma) and all are also familiar with Miro. Evaluator 2 is the most experienced Miro user, (s)he uses the software on a daily basis. The evaluators tend to work in teams and also have previous experience of Nielsen heuristic evaluation (Nielsen 1994). They all emphasised that they do not use the strict Nielsen method, but use the Nielsen heuristics as a framework for their work. Evaluator 2 and Evaluator 3 also noted that they had used the heuristics in their work on the day of the research. The participants' characteristics are summarised in Table 8.

5.2 Procedure

Our study, based on Nielsen's heuristic evaluation methodology, consists of two parts: (1) Individual evaluation and (2) Problem severity scoring (collaborative), which took place online (Nielsen 1994). In part (1) participants rated the usability of the Miro collaborative software according to collaborative heuristics (Sect. 2.2). The eight heuristics created by Baker (2001) are intended to provide evaluators with a common set of criteria for assessing the usability of real-time groupware. These heuristics are specifically designed to evaluate visual shared workspaces used by up to five users at a time. During the evaluation, participants verbally analysed the software using the think-aloud method, while the researcher took notes. The exact task of the evaluators is given in Appendix 10.6. Prior to the (1) Individual evaluation, an online questionnaire was completed, which included questions on demographics and work experience. During part (2) Usability problems were summarised beforehand, here the participants were asked to discuss and jointly

Table 8 Heuristic evaluation—characteristics of participants

	Academic experience (years)	Industrial experience (years)	UI design and UX research experience (years)	Collaborative whiteboard software experience	Typically working alone or in a team	Experience in the Nielsen heuristic evaluation (nr. of times)
1	5–10	<5	<5	Miro, Axure, Prezi	Alone	1–5
2	<5	11–15	5–10	Miro, Mural, MindNode	In a team	>5
3	5–10	5–10	5–10	Miro, Prezi, LucidCharts	In a team	>5
4	>15	>15	>15	Miro	In a team	1–5

determine the severity of each problem. In both parts of the study, a video and audio recording of the participants' computer screens was made using screen recording software to evaluate Miro.

1. Individual evaluation - online, approx. 90 min
 - Verbal information, information and consent form for the participant
 - Online questionnaire (approx. 5 min)
 - Usability analysis and evaluation of groupware (approx. 60 min)
2. Problem severity scoring (collaborative) - online, approx. 30 min

Summarising the usability problems found during the individual evaluation and scoring the severity of the problems with the other evaluators. The severity of the problems was rated by the evaluators on a five-point scale using the Nielsen's heuristic evaluation methodology, where the values were as follows (Nielsen 1994).

- 0 - Not a usability problem, no problem in use, or technical problem and not usability.
- 1 - Only a cosmetic problem, not important to fix (only if there is time).
- 2 - Minor problem, should be fixed, but not urgent.
- 3 - Serious problem, important to fix quickly.
- 4 - Usability "disaster", must be fixed immediately.

5.3 Tools/instruments

The study was conducted with Microsoft Teams. Microsoft Teams is a complex software for video calling, which features (screen sharing, camera view) enabled the study to be carried out online. During both parts of the study, video and audio recordings were made of the participants' computer screens using screen recording software. This was done using the recording feature of Microsoft Teams. The backup recording was made using the OBS screen recording software.

5.4 Steps of the analysis

We have summarised the usability problems mentioned by each evaluator, based on the notes from the individual evaluations. In the summary, we focused only on the problems with the collaborative features, not on the problems with individual use. The researchers have presented these in an online presentation in the (1) Problem severity scoring part, during which the evaluators collectively decided on the severity of each problem using Nielsen's five-point scale

(Sect. 5.2). The detailed results are presented in a table in Appendix 10.1, grouped by the severity of problems and heuristics.

5.5 Results

During the evaluation, participants identified 21 problems, of which 16 were eventually rated as usability problems (problems with a severity greater than 0). Out of the 16 usability problems, 3 were serious problems, 4 were minor problems and 9 were cosmetic problems. Most of the problems (5 problems) were related to heuristic 1, which emphasises the importance of supporting verbal communication or some alternative to it between participants (Provide the means for intentional and appropriate verbal communication). Of the 3 serious problems, 2 were related to heuristic 1, the covering of some elements of the workspace with the video chat window and the possibility of the communication menu bar disappearing were rated by participants as serious usability problems. In terms of severity of problems, a lot of them were related to heuristic 5, Provide protection, which is related to accidentally modifying each other's work and preventing overwriting and deleting someone's work on the workspace. According to the evaluators, a serious usability problem related to heuristic 5 is that it is not possible to see in the edit history in details. The other usability problems that were rated as minor or cosmetic problems are related to heuristics 2, 3, 4, 6. Heuristic 2 focuses on the visibility of actions related to the common task, in this regard the visibility of the colour of comments and the possibility of hiding the participants' arrows were found to be problematic by the analysts. With heuristic 3, which focuses on the importance of keeping users' movements visible on the workspace, evaluators highlighted that they found it easier to navigate through the content rather than between users, which can make collaboration more difficult. Heuristic 4, similar to 3, only focuses on the continuous display of the movement of objects in the collaborative workspace. According to the evaluators, a problem may be that the tracking of changes is not clear (e.g., during a real-time collaboration, when team members edit in a shared workspace together, if someone goes out for a moment without paying attention, it is difficult to assess exactly what changes have occurred in the time missed). In addition, the edit history was not found where it would have been expected. With regard to heuristic 6 (Manage the transitions between tightly and loosely-coupled collaboration), the evaluators were unsure whether the map showed users working together at the same time. In addition, it took them a long time to realise that the map could be hidden if one did not want to follow what others were working on. They found it problematic that it was not clear what would happen when the "bring everyone to me" button was clicked, whether the invited collaborators

would receive a warning or whether they would "just find themselves there". Based on the results, there are no problems with heuristics 7 (Support people with the coordination of their actions) and 8 (Facilitate finding collaborators and establishing contact), and Miro was considered adequate by the evaluators in these aspects. In addition, the evaluators noted that although the heuristics helped them to evaluate Miro, they were very similar and therefore fewer heuristics would have been appropriate for analysis.

5.5.1 Summary of the heuristic evaluation

During the heuristic evaluation, experts evaluated the usability of the groupware based on collaborative heuristics built on the mechanics of collaboration. The heuristic evaluation revealed different types of problems than lab studies. The results address both synchronous and asynchronous team usability problems. Most of the problems and most of the severe problems are related to communication (heuristic 1). Ensuring proper communication is a key factor for effective collaboration in real-time groupware (Gutwin and Greenberg 2002). The severity scores for communication-related problems also support the view of the evaluators.

The evaluators also indicated a serious problem related to the Provide protection heuristic, suggesting that one of the features supporting overwriting avoidance was not considered fully adequate. In this respect, the results are similar to the results of the laboratory studies, where explicit communication and protection were also identified as team usability problems.

It is important to underline that the evaluators overall rated the usability of Miro high and several noted that they liked to use it in their daily work. Some respondents also added to the evaluation that Miro can present difficulties for first or novice users, but that these can be easily overcome.

5.5.2 Results on the development of the team usability testing

The results of the heuristic evaluation confirmed that the empirical Team Usability Testing method developed in our research project reveals different types of problems involving real or potential users than the expert analysis.

For example, although the heuristic evaluation also raises the problem of protection, it is related to the editing history. In contrast, with the Team Usability Testing, a very serious usability problem, the overwriting problem (accidental modification or deletion of each other's work) was explored in both lab studies. This type of problem was revealed only by the empirical Team Usability Testing method based on user involvement.

The results of the heuristic evaluation confirmed that involving real or potential users reveals different types of

problems than the expert analysis. The advantage of expert analysis is that it provides a quick and comprehensive picture of the groupware under evaluation, thus complementing the laboratory study.

Overall, we believe that expert analysis and the Team Usability Testing are complementary methods. Expert analysis is a method that can be used in the very early stages of software development, and should be applied. However, at later stages, by involving potential users, Team Usability Testing is better suited to identifying problems and understanding groupware usability.

6 Discussion

The goal of our research project was to develop a new usability evaluation method, which is able to explore the usability problems of real-time groupware. The method was developed in four phases: first lab study, field study, second lab study and heuristic evaluation. Here we discuss the results of the second lab study and the heuristic evaluation. (The result of the first lab study is discussed in our earlier work (Geszten et al. 2021).

In both the first and second laboratory studies we conducted a content analysis of the communication and post-experiment group interview transcripts as well as the textual responses to the before and post-task questionnaire based on the theory of the mechanics of collaboration (Pinelle et al. 2003). The team usability problems identified in the research project are usability problems that occur during the team's collaborative work and affect the team's collaboration while using the software. The Team Usability Test was able to identify team usability problems in a laboratory context. Team usability problems were related to explicit communication, basic awareness and the management of shared access in the first lab study, and basic awareness and the management of shared access in the second lab study.

In the first lab study, the problems related to basic awareness were avatar, synchronization and saving, while in the second lab study, the problems were awareness, avatar and zoom. These are all factors related to the features of the software and problems with them negatively affected collaboration. In groupware, supporting awareness is key, as evidenced by the large body of research and design frameworks on this topic (Collazos et al. 2019; Gutwin and Greenberg 1996, 1998; Gutwin et al. 2004; Ignat et al. 2015; Lopez and Guerrero 2017). In addition, the results are consistent with the 3C model, which shows that awareness is central to collaboration and thus particularly salient in the use of collaborative software (Fuks et al. 2005). The results also confirm previous literature, that real-time groupware often has awareness problems (Baker et al. 2002; Dew et al. 2015). According to the EWG framework the problems with

awareness functions indicate that the usability problems of the groupware are on the technology level, indicating that specific implementations of awareness functions do not have a high level of usability (Cugini et al. 1997).

The overwriting problem related to the management of shared access mechanic also appeared in the first and second lab studies. Although the research design was similar, this result contradicts Ellis et al. (1991)'s results, in which the overwriting problem rarely arose. According to the management of shared access mechanic, a groupware should prevent team members from accidentally deleting or overwriting each other's work (Pinelle et al. 2003). Overwriting is the most severe usability problem, with serious negative effects on collaboration. Overwriting was repeatedly avoided by the participants by precisely dividing the workspace. Problems with the management of shared access have also been reported in previous research (Dew et al. 2015; Pinelle and Gutwin 2008).

The difference between the problems identified in the first and second lab study is that in the first lab study, team usability problems related to explicit communication also appeared. This is not surprising, since while in the first lab study participants worked together in different locations to simulate virtual teamwork, in the second lab study they were able to collaborate in person. The results confirm Ellis et al. (1991)'s findings that face-to-face collaboration is smoother than distributed or hybrid. The occurrence of problems with explicit communication negatively affected collaboration, in line with the 3C model and previous research about groupware (Fuks et al. 2005; Pinelle et al. 2003).

The results confirm that Team Usability Testing is suitable for identifying team usability problems, and thus may have a valuable role among existing groupware evaluation methods.

As a final step in the development of the methodology, we carried out a heuristic evaluation (involving experts). The subject of the analysis was the Miro collaborative whiteboard software, also evaluated in the second lab study. By comparing the lab studies and the heuristic evaluation, we examined the types of problems that the method could detect compared to the heuristic evaluation. According to the literature, heuristic evaluation can reveal different types of problems than a usability study with real users, which is supported by our results (Solano et al. 2016). While in heuristic evaluation, evaluators tend to evaluate the software more comprehensively, users are best able to identify usability problems related to their daily tasks (Nielsen 1994; Steves et al. 2001). It mostly depends on the purpose of the research which method to choose, but when there is a possibility, it is recommended to use both methods together (Lazar et al. 2017; Rubin and Chisnell 2008).

In the heuristic evaluation, experts evaluated the usability of the groupware based on collaborative heuristics built on

the mechanics of collaboration (Baker et al. 2001). For the analysis of special software, the literature recommends the use of specific heuristics instead of the general Nielsen heuristics (Rusu et al. 2016).

The key difference is that while both the laboratory studies and the heuristic evaluation revealed problems with awareness, management of shared access and verbal communication, the heuristic evaluation did not reveal the most serious problem in the laboratory study: overwriting. In the lab study, there were several instances of overwriting in the shared workspace, where one participant inadvertently altered or deleted the work of another. This type of problem was only identified in the lab study.

It is important to stress that the evaluators overall rated the usability of Miro high, with several noting that Miro can present difficulties for first-time or novice users, but that these can be easily overcome. In contrast, the results of the second lab study suggest that Miro is a groupware that can cause serious difficulties for novice users. Although, in the second lab study, we observed that users were able to overcome usability problems with the help of each other, which is in line with Ellis et al. (1991)'s results: since users collaborate in the same software, friendly help is available from each other, many found the software difficult to use overall.

Thus, the Team Usability Testing method can reveal different types of results than a heuristic evaluation based on the experiences of real or potential users.

However, heuristic evaluation also has several important added values. The most prominent is that it does not focus on a single task, but examines all the collaborative features of the groupware in its entirety, thus providing an overview of the usability of the features.

The results of the two methods complement each other well, and in practice, the combination of comprehensive expert opinions and user experience provides a complete picture of the usability of the groupware. This result confirms previous research findings that both analytical and empirical methods are important when evaluating the usability of groupware, but that neither method can be substituted for the other (Steves et al. 2001). Moreover, Solano et al. (2016)'s work suggests that inspection and test methods should be used together, when evaluating the usability of groupware.

Overall, Team Usability Testing was able to identify additional findings compared to heuristic evaluation, and we therefore recommend its use in groupware evaluation situations.

7 Conclusion and summary of the team usability testing method

In our research project, we developed a method for evaluating the usability of real-time groupware in the four steps described above (first lab study, field study, second lab study,

heuristic evaluation). This method is called Team Usability Testing.

Team Usability Testing is an empirical method based on user involvement that is suitable for evaluating working prototypes or released software. Compared to the analytical and empirical methods mentioned above, which are relatively time-consuming (days to months), this method is less time-consuming (90 min plus analysis).

Team Usability Testing is suitable for exploring team usability problems in real-time groupware and contextual factors that influence collaboration while working together in a shared workspace. Team usability problems are usability problems that occur during collaboration while using the same groupware. They only occur in collaborative situations and cannot be investigated by single-user usability tests. These problems cause difficulties during collaboration. Team Usability Testing can be used to investigate teams collaborating in the same (face-to-face) or in different locations (remote).

Team Usability Testing consists of questionnaires, screen recording videos and post-experiment interviews. In terms of methods, the Team Usability Testing is similar to Solano et al. (2016)'s "Evaluation Focused on Specific Tasks: No Time Restrictions" method, which consists of cognitive walkthrough, formal usability experiments and questionnaires. The biggest difference is that the method we developed uses more empirical methods based on user involvement and bigger teams (three collaborators instead of two). The data analysis is also more comprehensive, which we will discuss in the next paragraph. Besides, Solano et al. (2016) highlights that "the documentation (guidelines) about how to conduct collaborative usability evaluations of interactive systems is scarce" (Solano et al. 2016, p. 14). Therefore, one of the added values of Team Usability Testing is that it is a well-documented method.

The data analysis is based on Pinelle's (2003) mechanics of collaboration and Marks' (2001) team processes theory. The data analysis consists of analysis of communication transcripts, interview and questionnaire data.

The method is able to reveal the communication patterns of each team based on the code system developed by the sequence analysis of the communication transcripts (Geszten and Hámornik 2023). Sequence analysis of communication transcripts and thus the identification of communication patterns is an important part of team-level usability analysis in scientific, academic research. For practical, industrial applications of the method, due to the generally shorter timeframe of the research, sequence analysis can be considered an optional part of the method. If sequence analysis is omitted, it is important to interpret the results with caution. Without identifying the communication patterns, it is not possible to know for sure whether the team usability problems identified relate to the points of the software to

be developed, the team's communication difficulties or their interaction. The recommended use of this method is summarised in Appendix 10.2.

8 Limitations and future work

One of the limitations of our research is that we examined three groupware that were used together in real time by users. For this reason, it is important to expand the number of groupware tested, and it is also worth extending the research and investigating the applicability of the method to asynchronous groupware (e.g., project management software). By examining multiple types of groupware, it will be possible to better distinguish between what is a usability problem specific to the groupware evaluated and what is a general problem specific to similar types of groupware.

In addition, now only Computer Support Collaborative Work (CSCW) groupware and scenarios were examined, the focus was on the workplace settings. In the future it would be exciting to investigate Computer Supported Collaborative Learning (CSCL) groupware and scenarios with a special focus on learning environments. In this way it would be also possible to make a distinction between CSCW and CSCL groupware usability problems.

Another limitation of the research project is the number of teams: a total of eighteen teams were examined (seven in the first and eleven in the second lab study). This number is common for individual usability studies, where the examination of 5–6 users reveal 80% of the potential problems (Nielsen and Landauer 1993). For usability studies on groupware, there is no precise limit to the number of teams that should be examined, and a more extensive study with more teams would help.

Furthermore, the questionnaires used in our studies are not standard questionnaires, therefore they should be tested with more teams to determine their reliability and validity. At the same time, in the future the current questionnaires can be supplemented or replaced by standard questionnaires, e.g., the Shared Workspace Usability Scale (SWUS) questionnaire (Berkman et al. 2018). Some scales of the SWUS questionnaire could offer a reliable and valid alternative to the before-and post-task questionnaire questions: the Grounding scale (items related to how well the users understand each other) to the acquaintance question ("How well do you know the participant sitting at the red/yellow/blue computer?") and the Team Integration scale (items related to understanding other users, satisfaction and reaching group agreement) to the collaboration and team mood questions ("How do you think you could collaborate with your teammates?"; "How activated was the team when working together?"; "How pleasant was working together?"). It is worth investigating what standard

questionnaires can add to the results and, in return, how much time they add to the length of data collection.

Another limitation of the research is that it focuses on the narrow concept of software usability as defined in Sect. 2.1. We plan to continue the research in several directions in the future. On the one hand, we would like to broaden the focus of the research to the user experience of groupware, which includes not only usability but also aesthetic and emotional experiences. On the other hand, we find it both academically and professionally challenging and exciting to study groupware at a systemic level. In most cases, teams do not use groupware on their own, but the process of collaboration takes place in a kind of "digital ecosystem" using multiple collaborative and single-user software and physical tools. Therefore, in the future, we would like to look at the whole collaboration process and the groupware and tools that support it.

Although the aim of the research project was to investigate the scientific value of the Team Usability Testing, it would be worthwhile to investigate the practical, business value of the method. Therefore, in the future, we would like to investigate the applicability of the Team Usability Testing in a real groupware development process.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10111-024-00759-5>.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Dalma Geszten. The first draft of the manuscript was written by Dalma Geszten and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open access funding provided by Budapest University of Technology and Economics. Phd scholarship by the Hungarian Government (only).

Data availability Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval Approved by the (Hungarian) Joint Psychological Research Ethics Committee (Reference No.: 2018/03; 2020/120).

Consent to participate On request.

Consent for publication On request.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are

included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Atlassian (2018) Miro (Version 1.0.26). <https://miro.com/>
- Atlassian (2020) Miro (Version 2.0.1). <https://miro.com/>
- Aldag R, Kuzuhara L (2015) Creating high performance teams: applied strategies and tools for managers and team members. Routledge, London
- Bafoutsou G, Mentzas G (2002) Review and functional classification of collaborative systems. *Int J Inf Manage* 22(4):281–305. [https://doi.org/10.1016/S0268-4012\(02\)00013-0](https://doi.org/10.1016/S0268-4012(02)00013-0)
- Baker K, Greenberg S, Gutwin C (2001) Heuristic evaluation of groupware based on the mechanics of collaboration. *Lect Notes Comput Sci (Including Subser Lect Notes Artif Intell Lectu Notes Bioinform)* 2254:123–139. https://doi.org/10.1007/3-540-45348-2_14
- Baker K, Greenberg S, Gutwin C (2002) Empirical development of a heuristic evaluation methodology for shared workspace groupware. In *Proceedings of the 2002 ACM conference on computer supported cooperative work - CSCW '02*, 96, <https://doi.org/10.1145/587091.587093>
- Bartel CA, Saavedra R (2000) The collective construction of work group moods author (s): Caroline A. Bartel and Richard Saavedra Published by: Sage Publications, Inc. on behalf of the Johnson Graduate School of Management, Cornell University Stable URL: [http://www.jstor.com/s, 45\(2\), 197-231](http://www.jstor.com/s, 45(2), 197-231).
- Berkman MÍ, Karahoca D, Karahoca A (2018) A Measurement and structural model for usability evaluation of shared workspace groupware. *Int J Hum Comput Interact* 34(1):35–56. <https://doi.org/10.1080/10447318.2017.1326578>
- Bevan N, Carter J, Harker S (2015) ISO 9241–11 revised: what have we learnt about usability since 1998? In *Human-computer interaction: design and evaluation: 17th international conference, HCI international 2015, Los Angeles, CA, USA, August 2–7, 2015, Proceedings, Part I* 17, pp 143–151
- Bringas S, Duque R, Nieto-Reyes A, Tîrnăuică C, Montaña JL (2021) A framework for identifying sequences of interactions that cause usability problems in collaborative systems. *Electronics (switzerland)* 10(4):1–16. <https://doi.org/10.3390/electronics10040388>
- Christensen B, Ellingsen G (2016) Evaluating model-driven development for large-scale EHRs through the openEHR approach. *Int J Med Informatics* 89:43–54. <https://doi.org/10.1016/j.ijmedinf.2016.02.004>
- Collazos CA, Gutiérrez FL, Gallardo J, Ortega M, Fardoun HM, Molina AI (2019) Descriptive theory of awareness for groupware development. *J Ambient Intell Humaniz Comput* 10(12):4789–4818. <https://doi.org/10.1007/s12652-018-1165-9>
- Cugini J, Damianos L, Hirschman L, Kozierok R, Kurtz J, Laskowski S, Scholtz J (1997) Methodology for evaluation of collaboration systems. In *The evaluation working group of the DARPA intelligent collaboration and visualization program, Rev. 3*
- Damianos L, Hirschman L, Kozierok R, Kurtz J, Greenberg A, Walls K, Scholtz J (1999) Evaluation for collaborative systems. *ACM Comput Surv* 31(2):15. <https://doi.org/10.1145/323216.323362>
- Dew K, Turner AM, Desai L, Martin N, Kirchoff K (2015) Evaluating groupware prototypes with discount methods. In *Proceedings of the 18th ACM conference companion on computer supported*

- cooperative work and social computing - CSCW'15 companion, 183–186, <https://doi.org/10.1145/2685553.2699002>
- Ellis CA, Gibbs SJ, Rein G (1991) Groupware: some issues and experiences. *Commun ACM* 34(1):39–58. <https://doi.org/10.1145/99977.99987>
- Eppler MJ (2006) A comparison between concept maps, mind maps, conceptual diagrams, and visual metaphors as complementary tools for knowledge construction and sharing. *Inf vis* 5(3):202–210. <https://doi.org/10.1057/palgrave.ivs.9500131>
- Fuks H, Raposo AB, Gerosa MA, Lucena CJP (2005) Applying the 3C model to groupware development. *Int J Cooperat Informat Syst* 14(2–3):299–328. <https://doi.org/10.1142/S0218843005001171>
- Geszten D (2021) Teammunkát támogató szoftverek team-szintű használhatósági vizsgálati módszerének fejlesztése. Doctoral dissertation, Budapest University of Technology and Economics. Retrieved from <https://repozitorium.omikk.bme.hu/server/api/core/bitstreams/5a689069-c57b-41da-9fd3-6498b15b3dcd/content>
- Geszten D, Hámornik BP (2023) Can team communication obscure severe usability problems? *Information Society/Információs Társadalom (InfTars)*, 23(2)
- Geszten D, Hámornik BP, Hercegfı K (2019) Usability evaluation of a collaborative design software in the wild. In: 2019 10th IEEE International conference on cognitive infocommunications (CogInfoCom). IEEE, pp 101–106
- Geszten D, Hámornik BP, Hercegfı K (2021) Empirical study of team usability testing: a laboratory experiment. *Cogn Technol Work* 23:1–15
- Geszten D, Hámornik BP, Hercegfı K (2023) Finding treasure in the wild: a field study of a collaborative design software. In: *Accentuated innovations in cognitive info-communication*. Springer, pp 185–202
- Grigera J, Gardey JC, Fernandez A, Garrido A (2021) Tycho: facilitation support for groupware user tests. *Interact Des Architect*. <https://doi.org/10.55612/s-5002-049-003>
- Grudin J (1988) Why CSCW applications fail: problems in the design and evaluation of organizational interfaces. In *Proceedings of the 1988 ACM conference on computer-supported cooperative work, CSCW 1988*, 85–93. <https://doi.org/10.1145/62266.62273>
- Gumienny R, Gericke L, Wenzel M, Meinel C (2013) Supporting creative collaboration in globally distributed companies. In *Proceedings of the 2013 conference on computer supported cooperative work*, 995. doi: <https://doi.org/10.1145/2441776.2441890>
- Gutwin C, Greenberg S (1996) The effects of workspace awareness support on the usability of real-time distributed groupware. In *Proceedings of the 1996 ACM annual conference on human factors in computing systems - CHI '96*, 6, 3, 511–518. <https://doi.org/10.1145/345190.345222>
- Gutwin C, Greenberg S (1998) Effects of awareness support on groupware usability. In *Proceedings of the SIGCHI conference on human factors in computing systems - CHI '98*, 511–518. <https://doi.org/10.1145/274644.274713>
- Gutwin C, Greenberg S (2000) The mechanics of collaboration : developing low cost usability evaluation methods for shared workspaces. In *Proceedings of the 9th IEEE international workshops on enabling technologies: infrastructure for collaborative enterprises*, 98–103.
- Gutwin C, Greenberg S (2002) A descriptive framework of workspace awareness for real-time groupware. *Comput Support Coop Work* 11(3–4):411–446. <https://doi.org/10.1023/A:1021271517844>
- Gutwin C, Roseman M, Greenberg S (2004) A usability study of awareness widgets in a shared workspace groupware system. In *Proceedings of the 1996 ACM conference on computer supported cooperative work*, 258–267. <https://doi.org/10.1145/240080.240298>
- Hassenzahl M (2007) The hedonic/pragmatic model of user experience. *Towards a UX Manifesto* 10:2007
- Hassenzahl M, Tractinsky N (2006) User experience—a research agenda. *Behav Informat Technol* 25(2):91–97. <https://doi.org/10.1080/01449290500330331>
- Haynes SR, Purao S, Skattebo AL (2005) Situating evaluation in scenarios of use, (May 2014). *Proceedings of the 2004 ACM conference on computer supported cooperative work*, 92. <https://doi.org/10.1145/1031607.1031624>
- Heikkilä P, Honka A, Kaasinen E, Väänänen K (2021) Quantified factory worker: field study of a web application supporting work well-being and productivity. *Cogn Technol Work* 23(4):831–846. <https://doi.org/10.1007/s10111-021-00671-2>
- Herskovic V, Ochoa SF, Pino JA (2009) Modeling groupware for mobile collaborative work. In 2009 13th international conference on computer supported cooperative work in design, IEEE, pp 384–389.
- Ignat C, Oster G, Fox O, Shalin VL (2015) ECSCW 2015. In *Proceedings of the 14th European Conference On Computer Supported Cooperative Work*, 19–23 September 2015, Oslo, Norway, (September 2015), 19–23. <https://doi.org/10.1007/978-3-319-20499-4>
- ISO (2015) Ergonomics of human-system interaction—part 11: usability: definitions and concepts (ISO 9241–11: 2015). ISO, Geneva
- Kirkman BL, Mathieu JE (2005) The dimensions and antecedents of team virtuality. *J Manag* 31(5):700–718
- Kutlu B, Aggul YG, Atası I, Kaymaz Z (2021) A meta-analysis of studies on groupware for collaborative work environments. *Proceedings* 74:9. <https://doi.org/10.3390/proceedings2021074009>
- Larsen RJ, Diener E (1992) Promises and problems with the circumplex model of emotion. In: *Clar MS (ed) Review of personality and social psychology: emotion and social behavior*. Sage, Newbury Park, pp 25–59
- Lazar J, Feng JH, Hochheiser H (2017) *Research methods in human-computer interaction*. Morgan Kaufmann
- Leso BH, Cortimiglia MN (2022) The influence of user involvement in information system adoption: an extension of TAM. *Cogn Technol Work* 24(2):215–231. <https://doi.org/10.1007/s10111-021-00685-w>
- Lopez G, Guerrero LA (2017) Awareness supporting technologies used in collaborative systems. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing - CSCW '17*, 808–820. <https://doi.org/10.1145/2998181.2998281>
- Marks MA, Mathieu JE, Zaccaro SJ (2001) A temporally based framework and taxonomy of team processes. *Acad Manag Rev* 26(3):356–376
- Marlow J, Carter SA, Good N, Chen J-W (2016) Beyond talking heads: multimedia artifact creation, use, and sharing in distributed meetings. In *Proceedings of the 19th ACM conference on computer-supported cooperative work and social computing*, 1701–1713. <https://doi.org/10.1145/2818048.2819958>
- Martins LL, Gilson LL, Maynard MT (2004) Virtual teams: What do we know and where do we go from here? *J Manag* 30(6):805–835. <https://doi.org/10.1016/j.jm.2004.05.002>
- Nielsen J (1994) Usability inspection methods. *Conf Compan Hum Fact Comput Syst CHI* 94 25(1):413–414. <https://doi.org/10.1145/259963.260531>
- Nielsen J, Landauer TK (1993) A mathematical model of the finding of usability problems. In *Proceedings of the INTERACT'93 and CHI'93 conference on human factors in computing systems*, pp. 206–213.
- Parnell KJ, Banks VA, Allison CK, Plant KL, Beecroft P, Stanton NA (2021) Designing flight deck applications: combining insight from end-users and ergonomists. *Cogn Technol Work* 23(2):353–365. <https://doi.org/10.1007/s10111-020-00629-w>

- Pinelle D (2000) A survey of groupware evaluations in CSCW proceedings. In CSCW proceedings
- Pinelle D, Gutwin C (2002) Groupware walkthrough. In Proceedings of the SIGCHI Conference On Human Factors In Computing Systems Changing Our World, Changing Ourselves - CHI '02, (1), 455, <https://doi.org/10.1145/503376.503458>
- Pinelle D, Gutwin C (2008) Evaluating teamwork support in tabletop groupware applications using collaboration usability analysis. *Pers Ubiquit Comput* 12(3):237–254. <https://doi.org/10.1007/s00779-007-0145-4>
- Pinelle D, Gutwin C, Greenberg S (2003) Task analysis for groupware usability evaluation. *ACM Trans Comput Hum Interact* 10(4):281–311. <https://doi.org/10.1145/966930.966932>
- Pipek V, Wulf V (1999) A Groupware 's Life. In Proceedings of the sixth European conference on computer-supported cooperative work, (September), 199–218.
- Rubin J, Chisnell D (2008) Handbook of usability testing [electronic resource] : how to plan, design, and conduct effective tests (2nd ed.). Wiley Pub, Indianapolis. <https://doi.org/10.1007/s13398-014-0173-7.2>
- Rusu C, Rusu V, Roncagliolo S, Quiñones D, Rusu VZ, Fardoun HM, Collazos CA (2016) Usability heuristics: reinventing the wheel? In Social computing and social media: 8th international conference, SCSM 2016, held as part of HCI international 2016, Toronto, ON, Canada, July 17–22, 2016. Proceedings 8, pp 59–70
- Salomón S, Duque R, Montaña JL, Tenés L (2019) Modeling users behavior in groupware applications. In International conference on cooperative design, visualization and engineering Springer, Berlin, pp 11–21.
- Solano A, Collazos CA, Rusu C, Fardoun HM (2016) Combinations of methods for collaborative evaluation of the usability of interactive software systems. In Advances in human-computer interaction, 2016, <https://doi.org/10.1155/2016/4089520>
- Steves MP, Morse E, Gutwin C, Greenberg S (2001) A comparison of usage evaluation and inspection methods for assessing groupware usability. In Proceedings of the 2001 ACM international conference on supporting group work 125, <https://doi.org/10.1145/500286.500306>
- Szokolszky Á (2020) A pszichológiai kutatás módszertana. Osiris Kiadó, Budapest
- Tang JC, Isaacs EA, Rua M (1994) Supporting distributed groups with a montage of lightweight interactions. In Proceedings of the 1994 ACM conference on Computer supported cooperative work, pp 23–34.
- Thurmond VA (2001) The Point of Triangulation. *J Nurs Scholarsh* 33(3):253–258. <https://doi.org/10.1111/j.1547-5069.2001.00253.x>
- Veer Van Der G, Welie Van M (2000) Task based groupware design : putting theory into practice. In Proceedings of the 3rd conference on designing interactive systems: processes, practices, methods, and techniques, 326–337.
- Vyas D, van der Veer G, Nijholt A (2013) Creative practices in the design studio culture: collaboration and communication. *Cogn Technol Work* 15(4):415–443. <https://doi.org/10.1007/s10111-012-0232-9>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.