



HAL
open science

Usability inspections based on heuristics: towards a state of practices and requirements for support

Eliott Dutronc, Sybille Caffiau, Sophie Sdc Dupuy-Chessa

► To cite this version:

Eliott Dutronc, Sybille Caffiau, Sophie Sdc Dupuy-Chessa. Usability inspections based on heuristics: towards a state of practices and requirements for support. EICS '24: The 16th ACM SIGCHI Symposium on Engineering Interactive Computing Systems, Jun 2024, Cagliari Italy, France. pp.49-56, 10.1145/3660515.3661330 . hal-04617625

HAL Id: hal-04617625

<https://hal.science/hal-04617625v1>

Submitted on 19 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Usability inspections based on heuristics: towards a state of practices and requirements for support

Eliott Dutronc
Univ. Grenoble Alpes, LIG
Grenoble, France
eliott.dutronc@univ-grenoble-
alpes.fr

Sybille Caffiau
Univ. Grenoble Alpes, CNRS,
Grenoble INP, LIG
Grenoble, France
sybille.caffiau@univ-grenoble-
alpes.fr

Sophie Dupuy-Chessa
Univ. Grenoble Alpes, CNRS,
Grenoble INP, LIG
Grenoble, France
sophie.dupuy-chessa@univ-
grenoble-alpes.fr

ABSTRACT

User interface evaluation using heuristics is one of the most widely used evaluation methods in Human-Computer Interaction design. However, the quality of heuristic evaluations is highly dependent on the experts who perform them. To limit this bias, recommendations and tools to support some of the evaluation steps are proposed. Considering these recommendations and tools, this article investigates how the heuristics evaluations are really carried out and can be improved. Two studies are performed. First, an online survey investigates how practitioners perform heuristics inspection and studies the tools they used. The answers shows inequalities in the way the different tasks of the inspection process are realized. In particular, there is few well matured practices for proposing feasible solutions to usability problems and transferring inspection results to designers. These tasks are the moments when the knowledge of the experts who carried out the inspection is passed on to the people who will implement the interface. As such, they are particularly dependent on the communication quality. Our second study focused on this communication aspect by studying the use of a serious game, JADE, for usability evaluation. The JADE game has been successfully proposed in teaching to promote communication between learners. Two workshops investigates whether this game can also be a useful tool for experts in their inspection tasks, particularly as a support for exchanges. JADE seems to encourage communication between experts in particular to include non-ergonomic experts in the evaluation process.

CCS CONCEPTS

• **Human-centered computing** → **Heuristic evaluations.**

KEYWORDS

heuristic inspection, evaluation method, JADE

ACM Reference Format:

Eliott Dutronc, Sybille Caffiau, and Sophie Dupuy-Chessa. 2024. Usability inspections based on heuristics: towards a state of practices and requirements for support. In *Companion of the 16th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS Companion '24)*, June 24–28, 2024, Cagliari, Italy. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3660515.3661330>

1 INTRODUCTION

Inspection is the most common evaluation method used for external services in usability [7]. To carry out an inspection, heuristics can be used [14]. A set of usability experts inspect the same interface separately, following the same set of heuristics, before comparing their points of view and producing the evaluation report. The quality of the inspection therefore depends on the experts who carry it out.

This dependence can be justified first and foremost by the differences in the experts' knowledge and expertise. Works [9, 11–14] have been carried out to provide experts with knowledge during their inspection. These supports relate to the heuristics to be used, the protocol to follow and the application domain (context and user profile). Our first contribution aims at knowing actual practices to identify whether the existing supports are appropriate. To this end, we conducted an online survey covering the applied inspection methods and the tasks performed by experts carrying out inspections. The results show that the moments when the knowledge of the experts who carried out the inspection is passed on to the people who will implement the interface are less supported. So communication between experts is an important point to study.

In their study about usability evaluation practices, Folstad et al. [7] report that 74% of respondents collaborated during their last evaluation. Communication is particularly important when it comes to justifying the identification of a problem and specifying a solution. Some heuristic inspection support tools take this dimension into account by proposing collaborative task completion. However, none are designed to encourage communication between experts.

Direct interpersonal interaction during an inspection is one of the components of the JADE game (in French Jeu d'Apprentissage de l'Ergonomie [10]). This game, created by the University of Lyon (France), aims to teach ergonomic heuristics to computer science students using a playful approach. Considering its relevance for teaching, we explored whether this game could also be useful as tool for experts in their inspection tasks, particularly as a support

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

EICS Companion '24, June 24–28, 2024, Cagliari, Italy

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0651-6/24/06

<https://doi.org/10.1145/3660515.3661330>

for exchanges. The second contribution of this article is a preliminary study on the use of JADE for ergonomic inspection in two design iterations of a business application user interface. During the two workshops, we observed the use of the game as a support for exchanges and integration into the iterative process. JADE seems to achieve its goals in terms of communication between experts in particular it seems to permit including non-ergonomic experts in the evaluation process.

The next section presents the materials available to evaluators for carrying out a heuristic inspection. The sections 3 and 4.2 respectively present the two studies: the online survey and the heuristic inspection workshops we implemented. The data collected during these studies enable us to draw up some initial directions to provide effective help to support the experts' knowledge and improve the communication between them (section 5).

2 STATE OF THE ART: MATERIALS FOR HEURISTIC INSPECTION

A heuristic inspection is a non-user evaluation carried out by a group of experts (ideally 3 to 5 according to [14]) with the aim of detecting a set of usability problems on mock-ups or interfaces already developed [7]. Each expert carries out an inspection individually, before pooling their efforts to analyze and produce a report outlining the problems identified. In practice, this report can also contain some design solutions to resolve the problems identified [7]. To conduct the individual inspections, the experts can rely on heuristics (a set of rules) that identify points to be observed on the interfaces.

An heuristic evaluation is one of the most commonly used evaluation methods, notably because it is 50% cheaper than user-test evaluation [5, 7]. However, the main criticism of this evaluation method is the strong influence of the inspectors' level of expertise on the quality of the obtained results [4, 12]. Increasing the inspector's experience implies a better structure of the analysis [7], more thoroughness [5] and a better quality of the solutions solving the problems raised [5]. A number of tools - software or not - have been proposed to help with heuristic evaluations (and thus be less dependent on the level of expertise). R-IDE [12] is an on-line tool designed to help define a protocol using the DECIDE framework. Following this framework, various aspects of protocol definition are addressed, including planning, costing and ethical issues. Other tools provide assistance at specific stages.

The step of choosing the heuristics that structure the inspection has been the subject of several contributions, both software and methodological. R-IDE [12] and OPEN-HEREDEUX [13], for example, include modules to help choosing (and customizing) heuristics. The choice of heuristics is justified by the context (user profile, execution platform, etc.). To take these specificities into account, numerous sets of heuristics have been defined [2] and a methodology has been proposed to enable experts to create their own set of heuristics [11].

Carrying out the (individual) inspection stage is also highly dependent on domain expertise. To help experts orienting their inspection to the constraints of the domain, a frequently-applied approach is the use of persona [9]. Personas can be used to project a particular user profile, as in the case of usability inspection for

elderly users [3]. According to Friess et al. [9], this method also makes it possible to practice user-centered communication to expose usability problems and proposed solutions.

To permit the total independence from inspectors' expertise, Phetcharakarn & Senivongse [15] offer an automatic heuristics-based evaluation tool for Android applications. This tool analyzes code to detect usability issues corresponding to a few heuristics (19/146 evaluation questions). This approach saves time in identifying some problems, but cannot be generalized to all heuristics. Moreover, it does not propose any solutions to correct the problem detected.

However, as Fernandez et al. point out in [6], although the proposed tools make it easier to carry out a heuristic inspection, processing and comparing the results are the most difficult aspects. OPEN-HEREDEUX proposes to introduce a notion of score to help in the comparison of inspections by providing quantifiable (and therefore comparable) knowledge [13]. This addition provides a structuring element for exchanges, but does not replace them. Moreover, the analysis includes suggestions for solutions/modifications in addition to a summary of the problems detected and their importance [7]. These can be provided by others, with different skills [7]. Analysis thus integrates a human communication dimension.

While the literature reports on appropriate techniques and tools for conducting heuristic evaluations, it is difficult to know which are actually used in practice to explain the differences in the quality of heuristic evaluations. Moreover the human communication dimension is little studied. This dimension was taken into account when developing the JADE game [10]. This game aims to enable students to acquire persistent knowledge in usability. It also encourages exchanges between students, particularly during the phases of solution production and explanation of detected usability problems. In the context of university teaching, these objectives have been achieved, but no study has been carried out to identify whether its use by professionals could have the same benefits.

3 ONLINE STUDY OF PRACTICE

To understand the possible sources of differences in the quality of heuristic evaluations, we submitted a questionnaire about practices. Our study aims at confirming (or not) the needs of practitioners identified in the state-of-the-art works. We defined 3 research questions:

QA1. Do the tools proposed by the state of the art correspond to the experts' practice concerning the heuristic evaluation method?

QA2. Do the tools proposed by the state of the art correspond to the experts' practice to characterize heuristics?

QA3. Do the tools proposed by the state of the art correspond to the experts' practice to take into account the application domain?

3.1 Implementation of the only survey and participants

According to these objectives our participants were asked 18 questions. Following the topics, these questions¹ are grouped in 5 pages:

¹The questions in the questionnaire are all in French, and have been translated for the purposes of this article.

- Meeting and non inclusion
 - What is your current occupation?
 - As part of your professional activities, have you ever carried out at least one user interface evaluation by heuristic inspection (using a set of heuristics or ergonomic recommendations)?
 - Give examples or a definition of what a set of heuristics or ergonomic recommendations is.
 - How many heuristic inspections have you carried out in the past year?
 - What gender do you identify with?
 - How old are you?
- General information on heuristic evaluation practice
 - During the heuristic inspections you took part in, what steps were carried out?
 - What do you need to start a heuristic inspection?
 - What software tools do you use for this activity?
 - How many people were involved in this inspection (including you)?
- Identifying the set of heuristics
 - In the course of your career, how many sets of heuristics have you used to inspect user interfaces?
 - Have you ever had to define/create a set of heuristics for an inspection?
 - Have you ever carried out an inspection using a set of heuristics you hadn't been trained in?
 - Was the choice of heuristics for your last inspection imposed on you?
- Proposed solutions
 - During your last inspection, was there a step to identify solutions?
 - During your last inspection, who proposed the proposed solutions?
 - What form might the proposed solutions take?
- Evaluation reporting
 - What does an inspection report contain?

The questionnaire consisted of general questions followed by specific questions based on the literature. Each part included questions based on the participant's work, e.g. on the last evaluation carried out. The questions of what constitutes a set of heuristics and what tools are used to carry out an inspection evaluation are open questions. All other questions were presented with answer options (yes/no or multiple choice or selection).

The questionnaire was edited with Framafom, a free online tool and the questionnaire was distributed by a mailing list of French-speaking ergonomics and HCI practitioners. These practitioners, must not be researchers. 23 practitioners responded, including 19 who have already carried out at least one heuristic evaluation. 4 respondents to our questionnaire were disregarded as they did not match our profiles. The results presented below come from the analysis of the responses of the 19 participants. Participants are 7 males and 12 females with an average age of 36.5 years (standard deviation: 9). 13 are ergonomists, 3 are designers, 2 are HCI-computer scientists and 1 does not carry out any professional activity.

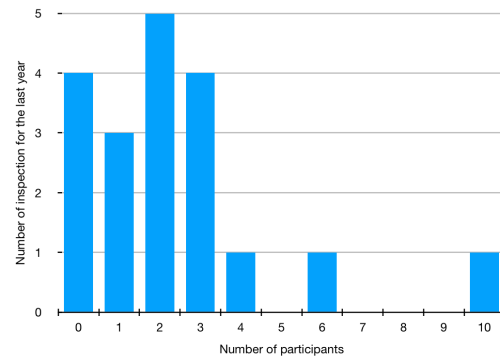


Figure 1: Distribution of number of heuristics inspection for the last year

3.2 Results

3.2.1 General knowledge of heuristic inspections. To identify our participants' knowledge and experience on heuristic inspections, we asked 2 questions. Firstly, we asked them to define what a heuristic is. All 19 participants correctly defined heuristics. Secondly, we asked them how many inspections they had participated in over the past year. 4/19 had not carried out any inspections in the past year, and 3 only one. All the others have done at least 2, which tends to show that this is a frequent activity for our participants (Figure 2).

3.2.2 QA1: Do the tools proposed by the state of the art correspond to the experts' practice concerning the heuristic evaluation method? There are four steps involved in carrying out a heuristic evaluation. First, a set of heuristics adapted to the application domain must be identified or defined. Secondly, based on this set, experts inspect for usability problems. Third, from this data, analysis can be performed. Analysis is defined by Følstad & al. [8] as "the process of turning the initial observations and hunches about difficulties into prioritized, coherent descriptions of usability problems, including a description of causes, implications and potential solutions to the problems.". Last, as with any evaluation, the results must be included to the artefact design.

The first, second and last steps are followed by all. The analysis step - such as defined by [8] - is partially performed. Only 12 participants include solution proposals during the inspection process. 8/19 also provide follow-up/accompaniment in the implementation of corrections to identified problems. In contrast to these numbers, 17/19 indicate that they were involved in designing a solution during the last inspection they attended. Improving inspected interfaces by implementing solution suggestions seems to be a step that is not systematically realized by practitioners who carried out the heuristic inspections and that is not considered as a part of the heuristic evaluations.

Following the inspection, practitioners are expected to produce a report. Consistent with the previous assumption, only 15/19 include solutions into the inspection report. One of the common difficulties in designing solutions and producing an evaluation report is the need to exchange/communicate with other stakeholders [16]. Practitioners use a variety of modalities to transfer their suggestions for

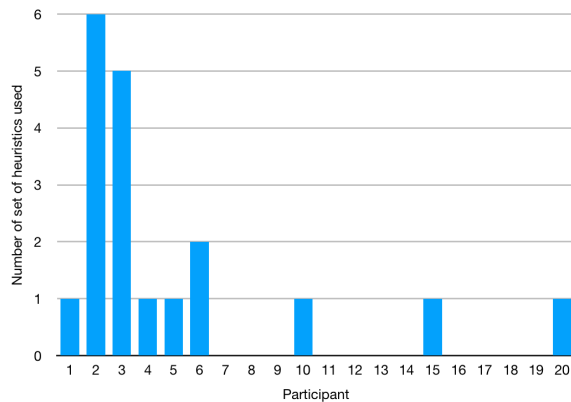


Figure 2: Distribution of number of used heuristics

improvement. The most frequently cited modality is text (17/19), followed by graphics (including screenshot modifications) (14/19), prototypes (12/19) and, finally, dynamic mock-ups (11/19). Less frequently, wireframes (1/19) and verbal recommendations (1/19) are used. These presentation methods are accessible to all areas of expertise (including, project manager, user and developers).

7 of the practitioners who took part in the design of solutions during their last heuristic inspection report that this step was performed by including experts who had not been involved in problem detection. In addition to the inclusion of other experts, Nielsen&Molish [14] recommends that 3-5 experts participate in the problem detection step. This recommendation could lead to communication difficulties between usability experts, particularly during analysis. Only 8/19 experts follow this recommendation to carry out their last inspection (6/19 with 3 experts and 2/19 with 4 experts). 5/19 carried out their last inspection alone, 6/19 with two experts.

3.2.3 QA2: Do the tools proposed by the state of the art correspond to the experts' practice to characterize heuristics? Depending on the field of application, the interaction modality and the type of interface, different heuristics may be applied and prove more relevant than others. With 3 exceptions, the practitioners surveyed used between 1 and 6 heuristic sets, and 12/19 between 1 and 3 sets (Figure 2). Despite this, 12/19 also claim that they have already carried out an inspection using heuristics for which they had no prior training. This number, as well as the fact that 7/19 have already defined their own set of heuristics for performing an inspection, shows that they are not afraid to acquire new knowledge if the case under study requires it. Furthermore, 18/19 indicate that they were free to choose (or participated in) the heuristics for their last inspection. The tools offered to help choosing among existing (but not necessarily known) heuristics, or to create new sets, can therefore be a real help in practice. However, none of the participants uses specific tools, such as OPEN-HEREDEUX [13], to search for a set or create a new set. As observed in [6], the used tools are pencil and paper, standard office tools and/or mock-up tools.

3.2.4 QA3: Do the tools proposed by the state of the art correspond to the experts' practice to take into account the application domain?

To be able to start an inspection, practitioners all need knowledge of the application domain of the interface they are going to inspect. We asked them what they might need to start an inspection. The most frequent answers were: a working prototype or developed application (18/19), screenshots (14/19), usage scenarios (13/19) or non-functional mock-ups (12/19). These resources correspond to the productions of the development phases identified in [7]. Only 6/19 indicated that they might need personas to represent target profiles, although studies have shown that the use of personas for this activity is beneficial [9].

3.2.5 Identification of a communication challenge. This study seems to identify differences between experts' practice. In particular, they do not all seem to share the same opinion on the place of correction suggestions: are they part of an inspection or not? Nor does the need for collaboration within this activity seem to be shared by all. 5/19 carried out their last inspection alone. Under these conditions, the quality of heuristic evaluation is bound to vary from one person to another. In addition to the impact on the quality of the inspection itself, these issues are also damaging to the integration of inspection within the design process. How can developers code improved interfaces from usability problems detected using heuristics of which they are not experts? Communication between different experts (in usability, IT and the domain) can be beneficial for problem identification and analysis, especially when it comes to find solutions that fit the field and the technical possibilities. No tools have been proposed to promote communication between experts during heuristic evaluation. The existence of a tool that supports communication between experts could perhaps encourage him to move from a recommendation to practice.

4 USE OF JADE TO SUPPORT EXPERTS' COMMUNICATION

In accordance with what has been observed previously (in [7, 16] and in our online survey), the transition from inspection analysis to the identification of solutions and their implementations is difficult.

In order to propose and design implementable solutions and to present an inspection, these tasks require personal and collective qualities to communicate. In a pedagogical way, University of Lyon (France) has developed a game (JADE) to learn heuristics inspection methods and supporting the inter-personal communication qualities. We conducted an exploratory study with pairs of ergonomics experts and software experts, with the aim of answering 3 questions concerning the use of JADE:

QB1. Does the introduction of the JADE game make it possible to follow the heuristic inspection stages as identified by studies of the practice of this evaluation mode? More specifically, does the introduction of JADE help to structure and justify the usability analysis, and then to propose corrective solutions?

QB2. Does the use of JADE allow the different areas of expertise to express themselves? Do usability experts not impose their point of view? Is the use of JADE accessible to software experts without initial training?

QB3. Is JADE suitable for use in iterative user interface design? Does the use of JADE enable the identification of problems and relevant solutions for further design? Can JADE be used to detect problems on developed prototypes?



Figure 3: Explanation using the prototype in Workshop 2

4.1 JADE: a learning game for ergonomics

JADE is a board game [10] that pits two teams against each other. Taking turns, the teams roll a dice to determine which squares on the board are accessible to them. Each accessible square determines an ergonomic concept (or heuristic) whose implementation (good or bad) the team must identify on the interface to be evaluated. Since our study focused on the implementation of heuristic inspection, the choice of heuristics was of no consequence. That's why we chose to use the game board linked to general usability concepts (Miller's Magic Number, Hick's Law, 2-Second Principle, 3-Click Principle, Bird Syndrome, Fitts' Law, Potentiality (Affordance), Proximity Law, Similarity Law, Colors, Reading Direction, Text).

4.2 Definition and implementation of JADE workshops

The study took the form of two workshops, each one a game of JADE (see section 4.1) to conduct a heuristic inspection of the user interface of a business application for tracking and managing product life cycles (see section 4.2.1). In each workshop, two teams (each made up of two participants, a usability expert and a software expert) competed to detect problems on design outputs (mock-ups and functional prototypes) (see section 4.2.2).

The usability experts took part in both workshops, while the software experts were different for each workshop. The usability experts were trained in heuristic inspection (during their initial training) and had already carried out heuristic inspections during their professional careers (during the last year they performed two heuristics evaluation). They are two French women aged 26 and 45. They are ergonomists and they work together. They correspond to the profile of participants to our online survey. They are not researchers. The software experts have never carried out a heuristic inspection and were not familiar with JADE prior to the workshops. The software experts taking part in Workshop 1 know each other and sometimes work together, while those in Workshop 2 also know each other and sometimes work together, but do not know those who took part in Workshop 1. The two workshops took place 12 months apart (September 2022 and September 2023). This corresponds to the development time required from the mock-up. Both workshops followed the same procedure: 30 minutes of training on one of the game sets available with the board (paper models) before the evaluation phase. The evaluation phase was stopped after 12 rounds of play (approximately 1 hour per session). Finally, at the end of each workshop, a phase of free feedback from the

participants enabled us to gather their opinions on the workshop methodology.

4.2.1 Use case: design of a personalization interface for a business application. We based our study on a real-life case of user interface design. This corresponds to an evolution of an existing software solution already in use at ST Microelectronics. This application is one of the internal applications used by employees to monitor and manage the life cycle of products manufactured by the company (Figure 4 [a]). To allow employees to adapt the interface of their business application, an interface for customizing operations and data presentation is currently being designed. The heuristic inspection workshops required for our study took place during two of the design iterations of this personalization interface. The aim of the first iteration was to produce paper mock-ups (Figure 5 [a]). The first heuristic inspection workshop took place at the end of this iteration. Based on these mock-ups and feedback from the workshop, a second iteration was carried out to produce a working prototype of the interface (Figure 5 [b]). The second heuristic inspection workshop was carried out on this prototype. These two types of support (mock-ups and functional prototype) correspond to those commonly used for inspections [7].

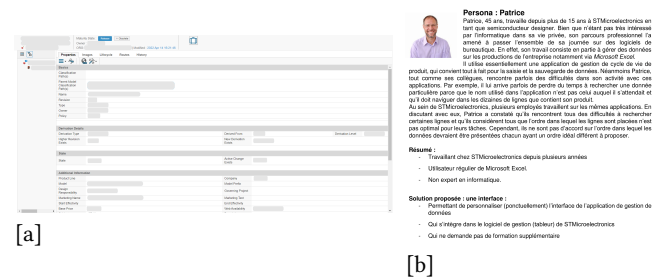


Figure 4: [a] Screenshot of the business application before evolution [b] Persona of business application

4.2.2 Material used during the two workshops. The two workshops were carried out at the DOMUS experimentation platform², which made it possible to record (video and sound) all the workshops (Figure 3). In addition to the game board, counters and dice (needed to play JADE), a persona (Figure 4 [b]) and a document containing the definition of each of the game board's criteria were distributed to each participant. The persona was created with the help of a team familiar with the business application and its users. For Workshop 1, a page with paper mock-ups was produced (Figure 5 [a]). Workshop 2 required a working prototype (Figure 5 [b]). In this workshop, participants had to present the problems detected on the prototype. In each round, the user interface was returned to its initial state. Participants performed the manipulations on the same machine (Figure 3).

4.3 Results

This study involved a small number of participants and cases. Consequently, the results are based on a descriptive analysis of the data

²<https://www.liglab.fr/fr/recherche/plateformes/domus>

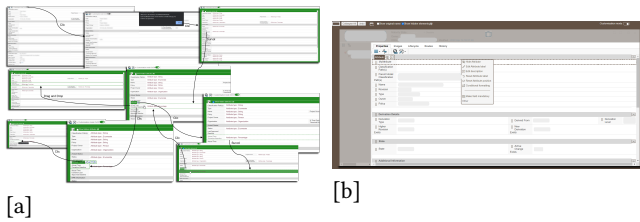


Figure 5: [a] Mockups and [b] Screenshot of improved prototype to be evaluated

(time, principles, problems and solutions) and on informal (qualitative) feedback from participants. In addition, only the most striking results are presented here, while others remain to be confirmed (or not) in further studies.

4.3.1 QB1: Does the introduction of the JADE game make it possible to follow the heuristic inspection stages as identified by studies of the practice of this evaluation mode? For this question, we observed the completion, during the workshops, of the stages of individual problem identification (in our case "individual" means "within a team") and analysis (through presentation to others and solution identification). In practice, the analysis stage includes a structuring phase (prioritization of problems, etc.) which cannot be observed in the workshops, as the JADE rules have been defined so that only one heuristic is studied in each round of the game, regardless of what was done in the previous rounds. As a result, participants did not have to prioritize the problems they identified. The rules of the game are also designed to encourage the exploration of the greatest number of heuristics (the team that explores the most heuristics wins bonus points). Although this rule (like all the others) was given by the facilitator at the start of the game period, the participants in both workshops chose to play collaboratively, with the common aim of achieving a complete inspection. By the end of the play period (54 minutes for Workshop 1 and 1 hour 15 minutes for Workshop 2), their common goal had been achieved.

The analysis phase of a heuristic involves argued exchanges to present the problems identified by one or more of the experts. In our workshops, these exchanges take place between teams. The average time taken by one team to identify a problem was 2:25 minutes (standard deviation: 1:23), and the average time taken to present a solution to the other team was 2:09 minutes (standard deviation: 2:01). These times indicate that there were exchanges between team members and between teams. At the end of the presentation of the problem and its solution, all participants had agreed (except for one solution out of the 23 proposed).

In line with what was observed in [9], participants used the persona (Figure 4 [b]) both in their internal deliberation phase (exposing their ideas to their teammate) and in external deliberation (defending their idea). Different persona characteristics were used in the argument: age, experience, Microsoft Excel usage habits...

The introduction of JADE made it possible to follow the analysis steps required for a heuristic inspection, with the exception of prioritizing problems and solutions.

4.3.2 QB2 : Does the use of JADE allow the different areas of expertise to express themselves? We have chosen to leave participants

free to choose the length and distribution of time during their exchanges, to ensure that they are realistic. The consequence of this decision is that we are unable to establish individual speaking times. As a result, we cannot draw any conclusions about the weight given to each expert's opinion during the discussions. However, we can provide an answer to the question of whether JADE allows software domain experts (who know neither JADE nor heuristics before the inspection) to contribute to a heuristic inspection.

Software experts detected 6/12 problems in Workshop 1 and 7/12 problems in Workshop 2 (Table 1). The average internal discussion time (in minutes) was 2:09 (standard deviation: 1:35) for Workshop 1 and 2:41 (standard deviation: 1:09) for Workshop 2.

These numbers seem to show that the use of JADE, although unknown to the software experts at the start of the session, allow them to identify usability problems, and that they are then able to discuss them with the ergonomics experts.

Workshop	Ergonomists	Software experts	Totals
Workshop 1	6	6	12
Workshop 2	5	7	12

Table 1: Problems detected only by one team member (ergonomist or software domain expert)

4.3.3 QB3 : Is JADE suitable for use in iterative user interface design? We carried out two heuristic inspection workshops for two design iterations. After Workshop 1, the solutions proposed to correct the problems raised were implemented in the functional prototype. In Workshop 2, none of the problems identified in Workshop 1 were identified again. Re-inspection using the same heuristics enabled other problems to be highlighted or their diagnosis to be refined. For example, in Workshop 1, the addition of a button to reduce sections was suggested. In the prototype, this addition was the subject of an affordance problem detected in Workshop 2.

The use of a functional prototype during Workshop 2 enabled us to identify problems related to the dynamic aspects of the interface (display on hovering over an element, difficulty in reaching a button, etc.) or to the implementation of the graphic charter (contrast), whereas in Workshop 1 the problems identified were related to static aspects (icon sizes, number of elements displayed on the screen, for example). Indeed, with the material used during Workshop 1 (mock-ups), participants only had access to what was presentable within the constraints of space. In Workshop 2, participants use the prototype themselves, allowing them to explore the interface and its interactions.

Iterating two heuristic inspections with JADE, when it follows a user interface concretization process, found different problems, even with the same set of heuristics.

5 DISCUSSION AND FUTURE WORK

As we discuss in session 4.3, this study involved a small number of participants and cases. A larger number could make the conclusion more accurate and representative. Despite this, we can draw up a few preliminary guidelines for future work, based on our initial results.

5.1 Adapting material to professional needs

Our study shows that there are tools available that are rarely used by practitioners. One possible explanation is that practitioners are unaware of the existence of these tools. Since they come from the academic world, we can assume that they have not been the subject of any particular publicity campaign aimed at industrial practitioners. However, the online study also showed that practitioners integrate heuristic search (or even the creation of a set of heuristics) when they need to. The question of finding tools to support their activity is therefore not linked to the knowledge of a tool (or of a particular need), but perhaps to a lack of knowledge of the need itself. This could be the reason why so few practitioners use a persona, even though this tool is familiar to them and studies show its value in heuristic evaluations.

Another possible explanation is that the tools do not fit in with their current practice and/or the constraints imposed on them in their professional environment. During the workshops, our observations have shown that the use of persona fits in well with the use of JADE, and provides an interesting complement for practice: "The persona enables us to better understand the user's point of view, especially when we're talking about a business application." (ergonomics expert, workshop 1). However, the participants also noted difficulties in presenting the persona: "I have trouble with the persona, it requires me to stay focused, after a while it gets out of my head." (ergonomics expert, Workshop 2). In our workshops, the personas were presented mainly in text form on a sheet of paper. Further work on the presentation of the persona could be carried out to identify how to integrate this source of knowledge for experts. Generally speaking, for all the tools offered to practitioners, the possible adaptations to their presentation need to be studied to facilitate their integration into experts' practices. These adaptations could take into account the experts' preferences as well as their experience on the task (with more or less guidance, for example).

5.2 Supporting communication between experts to perform analysis

In addition to the need for additional knowledge acquisition, our study also showed that one of the needs to improve the quality of heuristic inspections is to improve communication. A playful approach would provide support for communication during the analysis phase. To confirm and extend our primary results, as a first step, we could replicate the same study with more expert pairs to consolidate the observations made. In addition, this corpus would enable us to analyze the oral exchanges in greater detail (through discourse analysis, for example), in order to answer the question of the impact of each team member on decision-making. The analysis would focus on changes in exchange times and the frequency of interventions as a function of the participant and the progress of the game. The software infrastructure required for this data collection and (semi-)automatic data processing is currently being developed on the DOMUS platform (described in section 4.2.2).

These studies should also enable us to better identify some limitations of JADE for the practice of heuristic inspections by professionals. For example, as in the context of learning [1], participants in our workshops noted difficulties in identifying the "right" heuristic to apply. However, they also saw the structuring imposed by the rules

as a way of "[...] focusing on problem types, which helps to focus the discussion" (software expert, Workshop 2). With feedback from a larger number of experts, we would be able to better understand the extent to which we can take advantage of structuring while seeking to reduce the level of difficulty involved by its follow-up.

5.3 Supporting integration of heuristic evaluation analysis in iterative user interface design

Heuristic evaluation must be integrated into an iterative user interface design process. For this reason, communicating suggestions for improvement (analysis results) to developers is a key step for the usability of the interface. However, this communication is difficult [16]. A possible explanation is the lack of a common model or tool that supports this exchange. According to questionnaire responses, the suggestions for improvement are expressed mainly in textual or graphical forms without any notation. These modes of expression are accessible to everyone but do not correspond to the tools usually used by developers. Developers need specifications and use models and notations to express them. Improving the flow of evaluation results into the interface design cycle requires several additional studies on the practices of the two fields of expertise but also on the modes of communication and exchange. These studies could express solutions in the form of models used by developers (like UML models) or directly from tests (for the application of a Test Driven Development approach) or code (like certain online tools proposes to generate the css code corresponding to the designed graphic charter).

6 CONCLUSION

In this article, we have presented two complementary studies on the implementation of evaluation by heuristics: an online study with 19 participants and 2 workshops on the use of the JADE game for heuristic inspection by pairs of experts (ergonomists and software domain experts). Our objective is to identify factors to reduce differences in the quality of evaluations by heuristics. Our studies tend to show that although tools exist for this purpose (OPEN-HEREDEUX, R-IDE, persona method, recommendation to perform inspection by 5 experts...), they are not used by practitioners. They also highlight that communication between experts - although essential for this activity - is not appropriately supported in practice. We assume that this is the most important factor for improving the quality of heuristic evaluations and their integration into the design process. Nevertheless, the work presented here only concerns the observation from few experts' practice and two heuristic inspections on a single interface under design. To be able to conclude on all the questions we wished to investigate, further studies need to be carried out.

ACKNOWLEDGMENTS

The authors would like to thank the ergonomists Camille Roux-Gendron and Lilandra Farion from Floralis-UGA Filiale for taking part in this study. They would also like to thank COMUE UGA for its financial support for the IRGA EvalX project.

REFERENCES

- [1] Federico Botella, Cristian Rusu, Virginica Rusu, and Daniela Quiñones. 2018. How novel evaluators perceive their first heuristic evaluation. In *Proceedings of the XIX International Conference on Human Computer Interaction*. 1–4.
- [2] Eric Brangier, Michel Desmarais, Nemery Alexandra, and Sandrine Prom Tep. 2015. Évolution de l'inspection heuristique: vers une intégration des critères d'accessibilité, de praticité, d'émotion et de persuasion dans l'évaluation ergonomique. *Journal d'Interaction Personne-Système 4*, Special Issue" PISTIL" (2015).
- [3] Dana E Chisnell, Janice C Ginny Redish, and AMY Lee. 2006. New heuristics for understanding older adults as web users. *Technical Communication* 53, 1 (2006), 39–59.
- [4] Gilbert Cockton and Alan Woolrych. 2001. Understanding inspection methods: lessons from an assessment of heuristic evaluation. In *People and Computers XV—Interaction without Frontiers: Joint Proceedings of HCI 2001 and IHM 2001*. Springer, 171–191.
- [5] Gilbert Cockton and Alan Woolrych. 2002. Sale must end: should discount methods be cleared off HCI's shelves? *interactions* 9, 5 (2002), 13–18.
- [6] Jon Fernández and José A Macías. 2021. Heuristic-based usability evaluation support: A systematic literature review and comparative study. In *Proceedings of the XXI International Conference on Human Computer Interaction*. 1–9.
- [7] Asbjørn Følstad, Effie Law, and Kasper Hornbæk. 2012. Analysis in practical usability evaluation: a survey study. In *proceedings of the SIGCHI conference on human factors in computing systems*. 2127–2136.
- [8] Asbjørn Følstad, Effie Lai-Chong Law, and Kasper Hornbæk. 2010. Analysis in usability evaluations: an exploratory study. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*. 647–650.
- [9] Erin Friess. 2015. Personas in heuristic evaluation: an exploratory study. *IEEE Transactions on Professional Communication* 58, 2 (2015), 176–191.
- [10] Stéphanie Jean-Daubias. 2023. JADE: a board game to teach software ergonomics. *arXiv preprint arXiv:2308.03487* (2023).
- [11] Cristhy Jiménez, Hector Allende Cid, and Ismael Figueroa. 2017. PROMETHEUS: procedural methodology for developing heuristics of usability. *IEEE Latin America Transactions* 15, 3 (2017), 541–549.
- [12] Elizabeth Kemp and DT Setungamudalige. 2006. A resource support toolkit (R-IDE) supporting the DECIDE framework. In *Proceedings of the 7th ACM SIGCHI New Zealand chapter's international conference on Computer-human interaction: design centered HCI*. 61–66.
- [13] Lúcia Masip, Marta Oliva, and Toni Granollers. 2011. OPEN-HEREDEUX: open heuristic resource for designing and evaluating user experience. In *Human-Computer Interaction—INTERACT 2011: 13th IFIP TC 13 International Conference, Lisbon, Portugal, September 5-9, 2011, Proceedings, Part IV 13*. Springer, 418–421.
- [14] Jakob Nielsen and Rolf Molich. 1990. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 249–256.
- [15] Kwandee Phetcharakarn and Twittie Senivongse. 2018. Heuristic-Based Usability Evaluation Tool for Android Applications. *Applied Computing & Information Technology* (2018), 161–175.
- [16] Janice Redish, Randolph G Bias, Robert Bailey, Rolf Molich, Joe Dumas, and Jared M Spool. 2002. Usability in practice: formative usability evaluations—evolution and revolution. In *CHI'02 extended abstracts on Human factors in computing systems*. 885–890.