SN Computer Science (2022) 3:262
https://doi.org/10.1007/s42979-022-01153-6

ORIGINAL RESEARCH

An Iterative Process for the Evaluation of a Mobile Application Prototype

Eliza R. Oliveira[1] · André C. Branco[1] · Daniel Carvalho[1] · Eveline R. Sacramento[1] · Oksana Tymoshchuk[1] · Luis Pedro[1] · Maria J. Antunes[1] · Ana M. Almeida[1] · Fernando Ramos[1]

Received: 14 September 2021 / Accepted: 13 April 2022 / Published online: 30 April 2022

Abstract

This study is part of a research conducted under the CeNTER Program, an interdisciplinary project that aimed to promote the development of the Centro Region of Portugal. The general contribution of this paper is the evaluation of a mobile application prototype to promote the collaboration between the various agents involved in the areas of Tourism, Health and Wellbeing. It presents two assessment stages, conducted with experts and end-users. For the evaluation with experts, an inspection technique based on a heuristic evaluation was employed. For the evaluation with end-users, different methods were used, which included the collection of quantitative and qualitative data. Quantitative data were obtained through two User Experience evaluation tools (SUS and AttrakDiff) and from usability metrics of effectiveness and efficiency, which are key factors related to the usability of a product. Qualitative data were obtained using the Think-aloud protocol, which allowed the collection of immediate feedback on their experience of interacting with the prototype. Although there are still improvements to be addressed, the prototype was positively evaluated. Furthermore, the obtained opinions showed that the CeNTER application is a sustainable and timely contribution, with an interesting potential to foster community-led initiatives. Finally, this article offered an understanding for the evaluation of mobile applications, which foster the same subject approached in this study.

Keywords User experience · Mobile application prototype · Usability · Design and evaluation · Communities-led initiatives · Territorial based Innovation

Introduction

Digital media promotes the communication between local regional agents and boosts the dissemination of information regarding local products and activities for an unlimited number of people online [1]. Thus, it can facilitate

This article is part of the topical collection "Enterprise Information Systems" guest edited by Michal Smialek, Slimane Hammoudi, Alexander Brodsky and Joaquim Filipe.

✉ Eliza R. Oliveira
elizaoliveira@ua.pt

André C. Branco
acastello@ua.pt

Daniel Carvalho
daniel.carvalh@ua.pt

Eveline R. Sacramento
eveline.sacramento@ua.pt

Oksana Tymoshchuk
oksana@ua.pt

Luis Pedro
lpedro@ua.pt

Maria J. Antunes
mariajoao@ua.pt

Ana M. Almeida
marga@ua.pt

Fernando Ramos
fernando.ramos@ua.pt

[1] Digital Media and Interaction Research Centre, University of Aveiro, Aveiro, Portugal

collaborative processes among local citizens, valuing endogenous resources and promoting assets associated with a specific territory [2]. It also allows to recreate a "virtual proximity" among the different agents involved in the territory's development process [3]. In this context, a digital platform (mobile application) was designed to promote collaboration between the various agents (community-led initiatives, public and private entities, networks, and citizens), involved in territorial-based innovation processes in the Centro Region of Portugal [4].

The main goal of this article is to present an iterative evaluation process of a mobile application prototype, designed under the scope of the CeNTER Research Program. A two-stage assessment will be presented. The first stage of the assessment was conducted with digital technologies experts and specialists in the Tourism, Health and Wellbeing areas, while the second encompassed user experience tests with end-users, considered as potential application users. As continuous feedback from users in the early phases of development is crucial to detect possible problems of a system, both testing stages were carried out in a laboratory context. Therefore, such tests included the appreciation of the prototype at different stages of evolution, and from various perspectives, enabling a complete and thorough assessment.

In the first phase of evaluation, two groups of specialists carried out the heuristic evaluation of the prototype. Local experts are essential sources of information, as they have knowledge and experience about the final scenario of product use, helping to understand the needs of the end-user and the local restrictions on the development of the system. For the effectiveness of the tests, it was necessary to choose the evaluators according to their expertise, considering their professional area and experience with similar systems. Likewise, it was important to use methods that allowed a global view of the usability of the interface, and that allowed access to the main functionalities available in the prototype.

So, the first panel consisted of five experts in the Digital Technologies field, with knowledge and experience in developing applications and interfaces. The second panel consisted of five experts in Tourism, Health, and Well-being, who have knowledge of the domain and are involved in different community projects. This evaluation allowed us to identify and correct 50 usability problems, providing more engaging versions of the application [5]. Through the tests conducted in this phase, the prototype presented an innovative solution from the experts' point of view. It was also possible to obtain relevant data that gave rise to substantial changes in the preliminary prototype and was also important as a concept validation of the application.

The second stage of assessment of the mobile application prototype was carried out with potential end-users and it was presented in a previous article [6]. Selecting a sample of potential end-users was crucial to generate validity data with a usability test [7]. The involvement of the end-user in producing a system can facilitate the creation of mobile application interfaces that correctly correspond to the mental models created by the users about possible usage situations. Therefore, in the second stage of evaluation, this sample consisted of people representing the end-users who, potentially, will use the mobile application in the future. For this stage, the profile of the final evaluators was defined, whereby four use cases were defined: Community Initiatives, Public Entities, Individual Participation and Entity Networks.

As already stated, this paper presents both evaluation stages. Through the tests carried out, the development of a community-based territorial innovation application proved to be a sustainable and timely proposal, since the concept of the "CeNTER" prototype was positively evaluated, though there are still improvements to be addressed. These considerations will be presented with details along the text.

One of the main challenges that the mobile application developer faces is the high dropout rate after the first use. Therefore, this study was based on a "User-Centered Design" (UCD) approach, which defines the process necessary to develop products easier to use and better fulfil the objectives related to usability [8]. In this context, it is of utmost importance to maintain a UCD approach, involving the users in all phases of the interface development, to understand their needs, which should predominate in the interface design [9].

This study was also supported on a User Experience (UX) theoretical basis, which provided significant knowledge to elaborate on the mobile application prototype evaluation in the CeNTER Program scope.

This paper is organized as follows. "Theoretical background" briefly reviews some important concepts used in this research. "Methodology" addresses the adopted methodology and "Prototype" presents the mobile application prototype. "Evaluation reults" shows the results collected from the experts and end-users' evaluation tests. "Discussion" presents future research and challenges to be faced, due to some limitations of the tests. Finally, "Limitations and future work" contains the main conclusions and compares the results obtained in both evaluation tests.

It is important to say that this work is part of a larger research effort, part of it presented at the 2021 edition of the International Conference on Enterprise Information Systems [6]. This work extends the previous one by adding new ideas in the introduction and theoretical background sections, deepening the presentation of test results with end-users, and including a new section about the evaluation results with experts in technologies, tourism, health and wellbeing.

Theoretical Background

The purpose of UCD is to define the iterative process to develop products focusing on the users' needs and better fulfil the usability's goals [8]. Therefore, it implies the active engagement of users throughout the product or service development process, to prevent digital systems from failing due to a lack of communication between developers and users [10]. For these authors, design professionals need to follow a set of guiding principles in developing a product, so they can adapt it to the needs of each user. Still, according to these authors, compliance with these principles makes it possible to develop a product or service that is entirely user-centered. In order to understand users' desires and needs, it is necessary to gather as much observable data as possible in the entire design process and make a comparative analysis of these data to determine what similarities are found. Different evaluation methods are used, which include the collection of qualitative and quantitative data.

User Experience refers to how the end-user feels about the products created [11]. Bernhaupt and Pirker [12] state that the concept of UX is related to positive emotions and emotional results, such as joy, fun and pride. For Knight [11], creating an experience is not just about how the product is designed, which structures were implemented or whether state-of-the-art technology is used. It is about how the product helps users accomplish their tasks, achieve their goals and how they feel when they use and get involved with the product. For example, in the case of digital solutions, intentions are turned into products, which real people will use.

A mobile application's usability allows it to work as expected, enabling users to achieve their goals effectively, efficiently, and pleasantly [13, 14]. As Jones and Pu [15] mention, usability is not a purely one-dimensional property of an interface. It consists of a subset of user experiences associated with the effectiveness, efficiency, and satisfaction with which users can perform a specific set of tasks in a given environment. Usability is one of the key factors that affect software quality [16].

Heuristic evaluation, as Nielsen [17] mentioned, is an usability engineering method for finding the usability problem in a user interface design so that it can be attended to as part of an interactive design process. Heuristic evaluation involves having a small set of evaluators examine the interface and judge its compliance with recognized usability principles (the "heuristics").

This heuristic is supported by a checklist and verification items, which allows the evaluators to investigate and identify usability problems in the system interfaces, not directly involving the end user. After initial usage, Nielsen heuristics [9] were updated in 2020 [18], to add more current explanations and examples, but keeping the core intact as they apply to the development of a user interface from 1994.

A set of usability heuristics and the MATCH-MED checklist were developed to assess the usability of mobile health systems on smartphones. Based on the systematic review of the literature and the mapping of heuristic sets [19, 20], a set of heuristics was proposed based on the generic Nielsen heuristics with the addition of device-oriented heuristics mobile phones and mHealth apps. A checklist was developed from this set of heuristics to operationalize the usability evaluation [21].

For heuristics of effectiveness, efficiency and satisfaction, as Sauro and Lewis [22] mentioned, tests usually contain some combination of completion rates, errors, task times, task-level satisfaction, test-level satisfaction, help access, and lists of usability problems (typically including frequency and severity).

In this context, efficiency is seen as "the quickness with which the user's goal can be accomplished accurately and completely and is usually a measure of time" [, p.4]. Effectiveness refers to "the extent to which the product behaves in the way that users expect it to and the ease with which users can use it to do what they intend" [, p.4]. Effectiveness is usually measured quantitatively with an error rate. According to these authors, satisfaction refers to "the user's perceptions, feelings, and opinions of the product, usually captured through both written and oral questioning" [, p.4].23, 23, 23

Therefore, interfaces with good usability are characterized by their ability to offer a practical, easy, appreciable, and satisfying user experience [13]. As pointed out by Sharp et al. [24], users' cognitive processes must be considered when developing an interface, since these processes are constantly present in their daily activities. Therefore, when interfaces are not designed with users in mind, the most "inexperienced can suffer hours of frustration and may never be successful in their task" [, p.13].25

In this sense, to certify that a product has a satisfactory level of usability, it is essential to carry out tests that provide direct information about the problems that users encounter, allowing researchers to obtain precise recommendations on what should be modified in an interface [; 9, 26–28].

Methodology

This study was carried out under the CeNTER Program, developed at the University of Aveiro. One of the objectives of this program is the prototyping and validation of a mobile application for territorial innovation, with a focus on the areas of Tourism, Health, and Well-Being. In a prior stage to the development of the prototype, state-of-the-art

and requirements phases for the future platform were carried out. These processes included several complementary steps: a systematic review and a narrative literature review; a mapping of the innovative initiatives of the Region Centro of Portugal; a benchmarking of websites and social networks; interviews with leaders of four initiatives, as well as two focus groups attended by representatives of initiatives that stand out for their dynamism in the Region Centro [4–6]. Concerning the prototype development, the project team outlined its evaluation according to the iterative design methodology, with the evaluation of the medium-fidelity prototype by experts and laboratory tests with end-users [29].

The first stage of evaluation of the CeNTER prototype consisted of heuristic evaluation by two panels of experts. The first panel consisted of five experts in digital technologies and the second panel consisted of five experts in the fields of Tourism, Health, and Well-being.

The tests with experts took place at the University of Aveiro facilities, with the first group on October 28–31, 2019, and the second group on November 11–29, 2019. In total, the experts evaluated 78 screens of the developed mobile application prototype. This evaluation consisted of two phases: in the first phase, the experts freely explored the prototype and commented on their interaction with the prototype, and in the second, the experts filled a table of design-oriented heuristics for mobile applications.

The project team created a heuristic checklist to determine the severity of usability issues of the application prototype, based on two different references.

The first one was Nielsen's 10 heuristics [9] namely: (1) visibility of system status, (2) match between system and the real world, (3) user control and freedom, (4) consistency and standards, (5) error prevention, (6) recognition rather than recall, (7) flexibility and efficiency of use, (8) aesthetic and minimalist design, (9) help users recognize, diagnose, and recover from errors, (10) help and documentation.

The second one was three heuristics of the MATCH-MED scale [30] namely: (1) minimization of human–computer interaction; (2) physical interaction and ergonomics; (3) readability and Quick View.

It is important to note that the researchers chose to add these items of the MATCH-MED scale since Nielson's heuristics do not consider specific characteristics and limitations of mobile devices [31]. A total of 44 items were identified in this checklist, which was employed to evaluate the prototype usability.

In addition, the project team also employed the Think-aloud Protocol [7] to obtain immediate feedback from experts about their experience of interacting with the prototype. The application of this method allowed the qualitative evaluation of the prototype based on the experts' verbal comments.

According to the iterative design methodology, the researchers made improvements to the prototype after reviewing the results obtained on each evaluation panel. For example, a back button has been added to the tutorial to allow users to return to the previous page, search and save icons have been standardised to provide consistency across the platform. Experts suggested these modifications during the free exploration of the prototype in the first evaluation stage. Therefore, the prototype version tested in the second stage was different from the first, as it already included these main improvements proposed by the experts.

In a second stage, the prototype was evaluated by people with different profiles representing the main end-user groups of this application. With the intent to cover the most significant number of usage scenarios by each group of regional actors, such as citizens, community-led initiatives, public and private entities and networks, four different hypothetical Use Cases were prototyped. These cases correspond to common scenarios elaborated with 10 ordered tasks to be performed by three distinct participants that composed each group.

In the second stage, the collected data was also based on qualitative and quantitative information. Quantitative data were obtained through the combination of two UX evaluation tools, and metrics of effectiveness and efficiency, which are key factors related to the usability of a product. The evaluation instruments were the System Usability Scale (SUS) [32], and the AttrakDiff [33]. SUS is a widely used instrument for identifying usability's issues of a system, while AttrakDiff also comprises emotional and hedonic aspects of a product, embracing other important UX factors in the evaluation.

For each task performed by the users, the researchers identified whether they finished the task successfully or with assistance. The completed tasks are those in which users have accomplished their objectives without any help. Tasks that required help were pointed out as "Needed some help" and were not considered for the computation. Based on this result, a percentage of effectiveness is calculated for each use case. This indicator was based on the Nielsen [34] success rate usability metric. The effectiveness metric is a percentage of completed tasks divided by the total number of tasks (ratio).

The efficiency metric considered the time that each evaluator took to complete the tasks. According to different authors [; 22, 34], the evaluator with the best average time is considered as the reference for the use case that he belongs to. The time was measured in seconds, and it was counted from the user's first touch on the screen. Then, the percentage obtained from the best evaluator was calculated and compared to the average of the two other evaluators for each task.

Tasks that have the highest difference ratio between the time, in seconds, from the best participant to the average, will have usability problems since they present a significant variation in their execution times and need to be reviewed.

Qualitative data were obtained through a dialogue with the evaluators, captured on video throughout the test.

The test session began with a presentation of the CeNTER Program, the reading and collection of a free and informed consent document and an explanation of the test. The evaluation started with a free exploration of the prototype by the evaluator. Then a researcher dictated the tasks for the expert to perform. A Guided Exploration Task Guide, or Cognitive Walkthrough [35], was used, being this an inspection method based on performing a sequence of actions to complete a task. In addition, the Think aloud Protocol [7] was also employed, which encourages users to think out loud while exploring and /or performing a set of tasks.

Afterwards, the instruments (SUS and AttrakDiff) were presented to users, fulfilling the three phases of the test: Introduction, task execution and application of the instruments. All tests were recorded for later analysis by the team, to obtain more qualitative data through the evaluators' comments.

Considering the dynamic evaluation process presented, the tests encompassed the following goals:

- Measure indicative aspects of the prototype's usability, such as efficiency, effectiveness and satisfaction;
- Collect other important UX factors, such as hedonic qualities and an overall perception regarding the interface's look and feel;
- Verify the acceptance of the CeNTER prototype concept;
- Gather suggestions for improvement.

The second stage of the evaluation sessions occurred in October and November of 2020. These sessions took place at locations and times that varied according to the preference of each evaluator. Some tests were carried out at the University of Aveiro, while others took place at the institution or even at the participants' residence.

The evaluations were carried out individually, with evaluators who met the inclusion criteria within the different agents in the territory. In total, the researchers' team performed 12 tests. The researchers defined four Use Cases (UC): UC1—Community-led initiatives—involved evaluators representing community-based initiatives in the Centro Region of Portugal; UC2—Public Entities—tests were carried out with City Councils, Health Centers, and Parish Councils representatives; UC3—Citizens—people gathered as individual participation; and UC4—Networks—tests were performed with representatives of the Networks. This study's participants represented different profiles in terms of education, age, gender, and role performed in society, presenting distinct learning curves concerning the use of digital technologies.

Finally, after two months of testing, the researchers' team verified the UX assessment instruments results for data analysis. In parallel, qualitative data obtained from the careful observation of the videos were gathered, collecting comments and suggestions from the evaluators during the test.

Use Cases

This section presents the use cases based on user requirements in detail. These requirements were achieved from phases mentioned earlier, which comprised interviews with leaders of four initiatives, as well as two focus groups attended by representatives of initiatives that stand out for their dynamism in the Region Centro. In this sense, each use case was composed of a sequence of 10 pre-established tasks.

Use Case 1 (Community-led Initiatives) encompassed the following tasks: (1) See examples of higher-ranking events; (2) Add new event; (3) Select a specific date in the register; (4) Request a specific volunteer in the event register; (5) End registration (detailed event screen appears); (6) Share event on Facebook; (7) See on the map the location of the event; (8) Check on the map if there are events nearby; (9) See settings/configurations; (10) Change user preferences.

Use Case 2 (Public Entities) implied the following tasks: (1) Search initiatives that are happening in a specific place; (2) Read and participate in an initiative; (3) Identify the organization that organizes this initiative; (4) In this initiative, browse the existing events (Identify the place, date and time of the event); (5) Browse partners for this event; (6) Request to be an event partner; (7) Go back to the home screen; (8) Create a new resource offering; (9) In the definitions, see initiatives created by you; (10) Open an initiative created by you and change its location.

Use Case 3 (Individual participation) presented the following tasks to be accomplished: (1) Search events occurring in a specific place; (2) Search the classification of an event; (3) Participate in an event; (4) Create a profile (choose the option register yourself); (5) Save an event; (6) On the home page, consult and delete an event that has already taken place; (7) Browse the notifications; (8) Contact the organizers of a given event to clarify a doubt by email; (9) Ask to be a volunteer and (10) Consult the ideas section and insert an idea.

Finally, the fourth and final Use Case (Networks) requested the realization of the following tasks: (1) Add an Initiative; (2) Request a resource; (3) Request partners; (4) Consult events on the agenda; (5) Change user preferences; (6) See on the map the volunteers available in a geographic area; (7) Consult information about a volunteer; (8) Contact

a volunteer; (9) Comment on an idea; (10) Consult the participation of the user.

The Use Cases were elaborated by the CeNTER team, considering the results from previous research [36, 37] that allowed the identification of the potential regional agents highly involved in territorial innovation. Therefore, the outcomes achieved in this study may help to identify whether the CeNTER accomplishes the relevant functionalities for territorial development.

Prototype

To create a product that fulfills the ultimate goal of the CeNTER project, a prototype was developed under the User Centered Design (UCD) approach. In this sense, the elaboration of the CeNTER prototype had three iterative phases: (1) the Conceptual phase; (2) the Component phase; and (3) the Interactive phase. At the Conceptual phase, ideas were discussed and the problems to be solved by the platform were exposed. Brainstorming activities helped to identify the features that should be present in the prototype. Hereafter, a user flow was created, allowing us to understand which screens to develop and how they should be ideally connected to each other. From this user flow, several paper sketches (Fig. 1) were built to understand how the application should look in the future. Low-fidelity mockups were created, tested, and discussed among the team members for validation.

The Component phase included the conversion of the lo-fi prototypes to wireframes, using the "Sketch" application. From the wireframes, it was possible to understand how the texts, lines, and other graphic elements of the layout of each screen should be organized. After this, more improvements were made, giving rise to wireframes that already resembled the look and feel of the final product.

In the Interactive phase, the screens developed from the wireframes (Fig. 2) were later used to create a medium-fidelity prototype. The "Principle" application was used, allowing the development of a prototype with more complex interactions, such as dragging on a map, swiping on a carousel menu, or tapping on a menu to collapse visible content [6]. This medium-fidelity prototype was later shared with users to be evaluated to make final improvements. A mobile application was developed to promote interactions among local agents, facilitating processes of communication and collaborating, encouraging the joint creation of new ideas and activities, while taking benefits from existing mediation strategies.

Figure 2 shows the application's main screen, presenting a grid with six primary tabs: initiatives, events, entities, volunteers, resources and highlights, which are starting points of the application. The user finds the information displayed in a carousel mode as it opens a tab, and each card represents different units of content. These cards have basic information, like image, date and time and location, and can be manipulated with gestures, such as swiping, for discarding a card or saving it as a favourite. Distinct actions are possible when a card is presented, such as viewing the element on the map, adding a new element, or making specific searches within each tab.

The prototype header presents agenda features, search tools across the platform, and also provides access to the application settings. The menu in the footer includes other functionality options, such as accessing the user profile and ideas, visiting saved items, viewing notifications, and general exploration on the map. The navigation in the application is done with a minimum number of gestures.

Besides that, the CeNTER mobile application prototype presents a small tutorial, helping the users to easily understand how to interact with the platform.

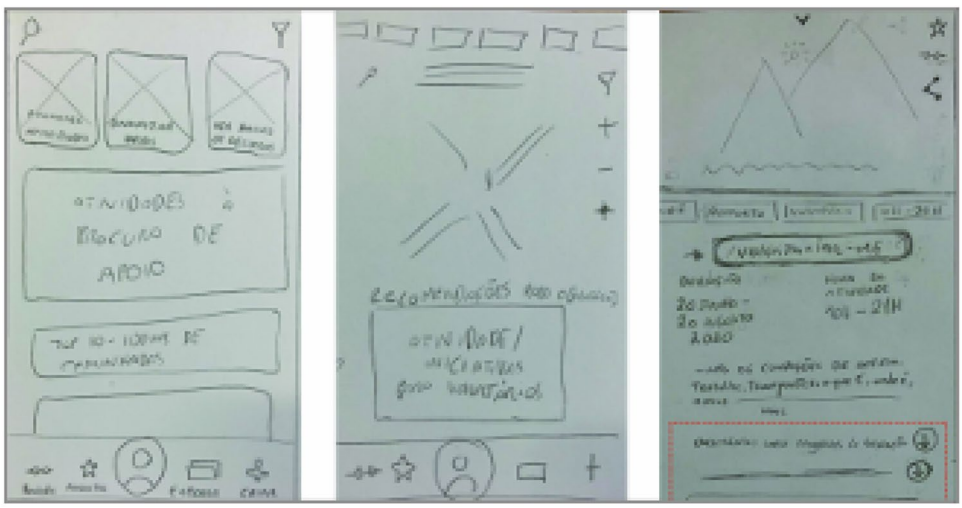

Fig. 1 Sample of sketches from the CeNTER platform (from left to right): Main menu, map and details of an event selected by the user

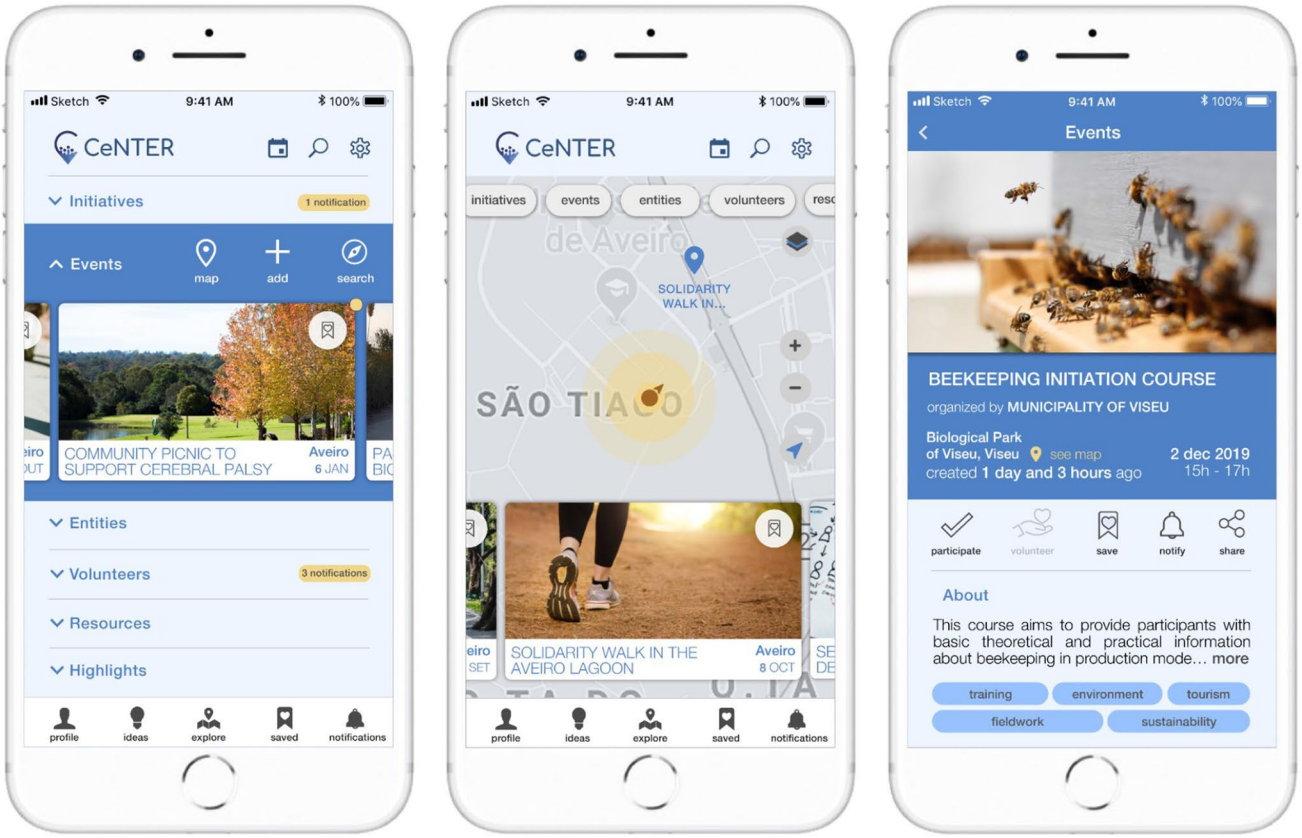

Fig. 2 Screen samples from CeNTER Prototype: Main screen, map screen and event screen

Evaluation Results

This section presents the main results obtained through the tests carried out with experts and potential end-users, which provided relevant quantitative and qualitative results, regarding instrumental and non-instrumental characteristics of a medium-fidelity mobile application prototype.

Experts Evaluation Results

After performing the heuristic evaluation and filling in the checklist, the general opinion about the prototype was positive. The prototype was evaluated as having a simple interface: the page layout was easy to use, did not require great effort by the user; and no incongruities were detected in the form of presentation, although some errors were indicated in the passage of links and in the use of some specific terminology in the area of Health and Welfare.

The heuristic assessment of the prototype was conducted through a comparative analysis between the results of two evaluators groups. To consolidate the heuristic validation results performed by both groups, a grid was created to gather all results. Next, these grids were analysed and compared. Table 1 shows the number of problems identified by each group of evaluators, regarding the total number of problems and the average severity.

As presented in Table 1, the group of technology experts identified a higher number of usability problems (108) than the other group (46). Regarding the severity of such problems, Group 1 identified: 46 problems with score 1 (visual problem only), 39 problems with score 2 (small usability

Table 1 Problems and average severities identified by the two groups of evaluators

Group	Evaluators	Total number of problems	Severity of problems				Severity average
			1	2	3	4	
1	Technology Experts	108	46	39	21	2	1.78
2	Tourism, Health and Wellness Specialists	46	14	26	6	0	1.53

problem), 21 problems with score 3 (main usability problem), and 2 problems with score 4 (usability catastrophe). These last problems refer to the heuristic "*Correspondence between the system and the real world*" and are related to the items "*The proposed interactions in the application are similar to real actions*" and "*Information appears in a logical and natural order*". There were no problems reported with score 4 by the second group.

With the highest Mean Severity, Group 1 identified the following heuristics: Help and Documentation (2.75), Recognition rather than Reminder (2.5), and Interaction between person and application (2.33). Group 2 reported the highest Mean of Heuristic Severities: Flexibility and Efficiency (2.25), Help and Documentation (2.2) Recognition rather than Reminder (2) and Interaction between person and application (2). Therefore, both groups were quite consistent in assessing the severity of the heuristic violation. The Mean Severity in the two groups was low, with 1.78 by Group 1 and 1.53 by Group 2.

The difference in results between these groups also occurred in the type of problems identified (Table 2). The problems identified by Group 1 were more focused on prototype design, robustness and accessibility. The problems identified by Group 2 were related to the suitability of the prototype for each application domain, as well as the efficiency and type of functionality available to users.

The maximum number of heuristic non-observations, identified by Group 1, englobe: Readability and layout (14, mean severity 1.5), followed by User controls and freedom (13, mean severity 1.54) and Aesthetics and minimalist design (12, medium severity 1.66). Group 2 identified the maximum number of heuristic non-observations: User controls and freedom (11, mean severity 1.45), Avoid errors (7, mean severity 1.86), and Matching the system to the real world (6, mean severity 1.5). Identifying these heuristic non-observations will make it easier to prioritize issues that need urgent attention before the final deployment of the application.

Regarding the inputs collected from the evaluation, a total of 174 inputs were obtained during the free exploration of the prototype by the experts, 121 collected from technology experts and 53 from specialists. Forty-one inputs were considered as prototyping errors, 124 were considered suggestions for improvements and nive were interpreted as prototyping errors and suggestions for improvements. Prototyping errors correspond to inconsistencies in the use of the prototype and the user interface, such as different icons representing the same function, or lack of an icon that should be present by default. An example of a suggestion that enabled the identification of a prototyping error by experts in technology is: *"Standardize the 'save' icons on the right side of the ideas screen with the rest of the application"* (Panel 1—Expert 1). An example of suggestions for improvement is: "*In 'Ideas', add year information and not just day and time*" (Panel 2—Expert 2). In addition, an example of prototyping error given by the specialists is: "*A person can get lost while switching between tabs, not sure how to go back when opening one tab and how to minimize another*" (Panel 2—Expert 3). An example of suggestions for improvement is: "*Change 'sport' by 'physical and sport activity. Sport is interpreted as competitive action"* (Panel 2—Expert 4).

Table 2 Type of violated heuristics and average severities identified by each group of evaluators

Nº	Heuristic	GR 1—Technology experts		GR 2—Tourism, Health and Wellness Specialists	
		Number of problems	Average of severities	Number of Problems	Average of severities
1	Visibility of system status	11	2.18	1	1
2	Match between system and the real world	11	2.27	6	1.5
3	User control and freedom	13	1.54	11	1.45
4	Consistency and standards	7	1.57	3	1.66
5	Error prevention	5	1.6	7	1.86
6	Recognition rather than recall	4	2.5	1	2
7	Flexibility and efficiency of use	7	1.85	4	2.25
8	Aesthetic and minimalist design	12	1.66	1	1
9	Help users recognize, diagnose, and recover from errors	10	1.1	4	1.9
10	Help and documentation	4	2.75	5	2.2
11	Interaction between person and application	3	2.33	1	2
12	Physical interaction and ergonomics	8	1.75	0	0
13	Readability and layout	14	1.5	2	1
	Total	108	1.78	46	1.53

It is noteworthy that experts of Group 1 highly commented about technology issues. In contrast, experts of Group 2 mainly commented about the concept of the CeNTER prototype, functionality and effectiveness for the end-user, while technical factors were not so mentioned. Some inputs are worth highlighting since they were pointed out by both panels. Overall, it was consensual that the Drag and Drop interaction system is a challenge when it refers to usability.

Only the experts of the ICT area were able to explore this interaction mode without any clue. Still, in terms of interaction, the swipe between the cards had good acceptability among the evaluators, but according to most of them, it won't be possible to use the carousel with a high amount of information in the database of the application. According to some experts, an alternative could be using a list format instead of cards. It is important to emphasize that the concept of the CeNTER application prototype has been well evaluated among the experts of both panels. In this aspect, the "Resources" and "Volunteer" components were highlighted as being the most innovative features of the platform.

Finally, the test with experts provided welcome hints to modify the prototype according to the experts' opinion,which gave rise to improvements accomplished before the tests with end-users, which will be presented in the next section.

End-Users Evaluation Results

Results from Effectiveness and Efficiency

This section presents the results of effectiveness and efficiency tests with potential end-users of the CeNTER prototype. The usability metric of effectiveness (whether the user performed the task, with or without help, or did not perform it) and efficiency (time of execution of each task), provided cues on how intuitive the design is; how frequent errors were committed while performing a specific task or action; and the required learning curve to use the platform. The effectiveness usability metric measured each user´s success rate in performing 10 tasks, totalizing 30 tasks performed in each use case. The results can be seen in Table 3.

On average, an effectiveness rate of 87% was obtained. However, it is important to highlight the lowest and highest effectiveness index obtained in the use cases, being 80% for Use Case 2 and 97% for Use Case 4. According to the metrics pointed out by Nielsen [33], an index above 80% is considered good, and it is not necessary to reach a higher value in time for a project prototype [22, 34]. These values correlate to the average obtained in verbal help, so the use case with the highest effectiveness index had a lower average in verbal help and vice versa, i.e., when the evaluator needed assistance to perform a task, this contributed to the decrease in the effectiveness index. Henceforth, the total average of the four use cases was made (Table 3), making it possible to understand that approximately every evaluator needed verbal help in at least one in ten tasks.

Table 3 Global results regarding effectiveness and efficiency

Use case	Effectiveness	Verbal help (average per end-user)	Efficiency (average per task)
UC1	87%	6.66%	16 s
UC2	80%	20%	15 s
UC3	83%	16.66%	18 s
UC4	97%	10%	17 s
Average	87%	13.33%	16.5 s

The average was made according to the number of times an end-user needed verbal help during the execution of the 10 tasks. Afterwards, in the same use case, the average obtained from all end-users was determined. Finally, the average obtained from the total use cases was calculated.

It is possible to conclude that the results obtained in the efficiency analysis were satisfactory. It is noteworthy that the efficiency metrics were obtained according to the time difference that the distinct evaluators took to perform the same task. It was also observed that the average time of execution of each task was around 16.5 s (Table 3), with low variation between the average of each use case, which demonstrates a high efficiency in terms of prototypes' usability.

In addition, three evaluators revealed some difficulties in carrying out tasks that required content creation (creating a profile or event with a date and time) and browsing tasks (such as finding the existing initiatives or reading the ideas' screen and subsequently creating a new idea). These outcomes were directly influenced by the learning curve of users, and their experience in using similar mobile applications. Thus, the usability evaluation of the CeNTER application prototype provided good results in terms of learnability, effectiveness and efficiency.

Results from the SUS and AttrakDiff Instruments

The main results concerning the application of SUS and AttrakDiff in all Use Cases are shown in Table 4.

In terms of usability characteristics, the SUS results show that the prototype is at an excellent level according to the opinion of the evaluators of the first use case (85 points). According to Sauro [38], the average System Usability Score is 68 points. In this sense, if the score is less than this value, the product probably faces usability problems, since it is under the average [39]. Therefore, a score between 80 and 90 in SUS corresponds to excellent

Table 4 Global results from SUS and AttrakDiff

Use cases	Instrumental qualities	Non-instrumental qualities			
	SUS (0 to 100)	AttrakDiff (-3 to 3)			AttrakDiff (-3 to 3)
		PQ	HQ-S	HQ-I	ATT
UC1	85	1,57	2,10	1,76	2,19
UC2	87,5	1,00	1,52	1,52	1,71
UC3	95	1,76	2,24	1,90	2,67
UC4	75,83	2,05	1,67	1,67	2,43
Average	85,83	1,60	1,88	1,71	2,23

usability [39], reflected in the case of the CeNTER mobile application prototype, with a global result of 85,83 scores.

The results obtained through the SUS administration in all use cases show an overall agreement among the participants, reinforcing the value of excellence, which is between 80 and 90 points, relative to the usability criteria measured by this evaluation instrument within the CeNTER platform.

Although the value related to SUS reinforces a high usability index, the value of Pragmatic Dimension (PQ), which encompasses usability and product functionality, obtained lower results (1,60), with oscillations between the Use Cases. The higher value was achieved in Use Case 4, while the participants of Use Case 2 gave the lower scores. However, the global average value remained positive (between − 3 and 3), so it is possible to consider that the prototype has a favourable index in the criteria of effectiveness, efficiency, satisfaction and ease of learning.

Concerning the results obtained from the AttrakDiff scale, the average values of the four dimensions were calculated, all of which had high scores, being possible to achieve scores between − 3 to 3. The general value is related to the prototypes' aesthetics "ATT"—Attractiveness (2,23), followed by the Hedonic Quality – Stimulation (HQ-S—1,88), which is strictly related to the desire to understand and develop skills for using the product. Afterwards, the biggest score is from the Hedonic Identification (HQ-I—1,71), which are attributes alluding to the level of user identification with the system. Finally, as previously said, the lowest score corresponds to the Pragmatic Quality (PQ—1,60), which is correlated to usability issues. Figure 3 shows the average values obtained in the other dimensions, highlighting the aspect related to the prototype aesthetics (ATT), which presented a value significantly higher than in the other dimensions, in agreement with the previous results.

Also, coherently with the rest of the results, HQ-S obtained a higher value than HQ-I, showing that the aspects referring to the desire to understand and develop skills for using the product are more evident than those related to the level of user identification with the system.

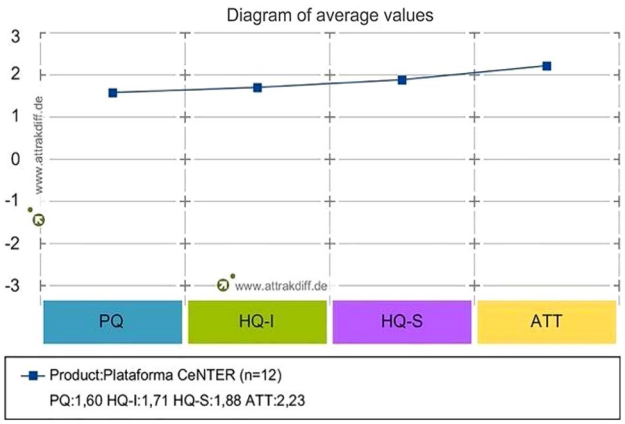

Fig. 3 Diagram of the global average of values of the four dimensions of AttrakDiff

Figure 4 shows that the pair of words which received the negative result in AttrakDiff was the topic "cheap—premium", in the HQ-I dimension, with no other negative average values among all items in the other dimensions.

However, it is essential to emphasise that, under the CeNTER project, none of the opposites in "cheap—premium" has an essentially negative connotation. Thus, a quality of "cheap" might mean that the Platform is accessible to all social fringes, consolidating the intention to democratize digital technologies in all strata of the population. Likewise, "cheap" can refer to a low complexity of the platform, indicating the desired ease of use within the scope of CeNTER. This point of view is consistent with the fact that the punctuation for the "simple—complicated" opposites are significantly more inclined towards the simple than for its reverse, and with the fact that the usability score measured by SUS, which refers to the ease of use, have shown to be substantially high.

Concerning Fig. 5, the general results achieved from AttrakDiff positioned the confidence rectangle in the "desirable" quadrant, assuming the perceptions of PQ (1.60) and HQ (1.80). According to the Attrakdiff methodology, the smaller the difference between the two rectangles, the greater is the confidence level of the results, indicating that participants maintained good affinity among their responses. Moreover, in the CeNTER scope, the confidence rectangle extends within the "desired" or "desired" area. Therefore, it can be clearly classified as a desirable product. This value, and all the other graphs presented above, were generated according to the AttrakDiff methodology.

An accurate analysis of the quantitative results in each use case, separately, shows that the participants in UC2 had more difficulty in performing the tasks, considering that this was the group that most needed verbal help. In the meantime, the results obtained from AttrakDiff showed the lowest values scored by participants. In this sense, the global results

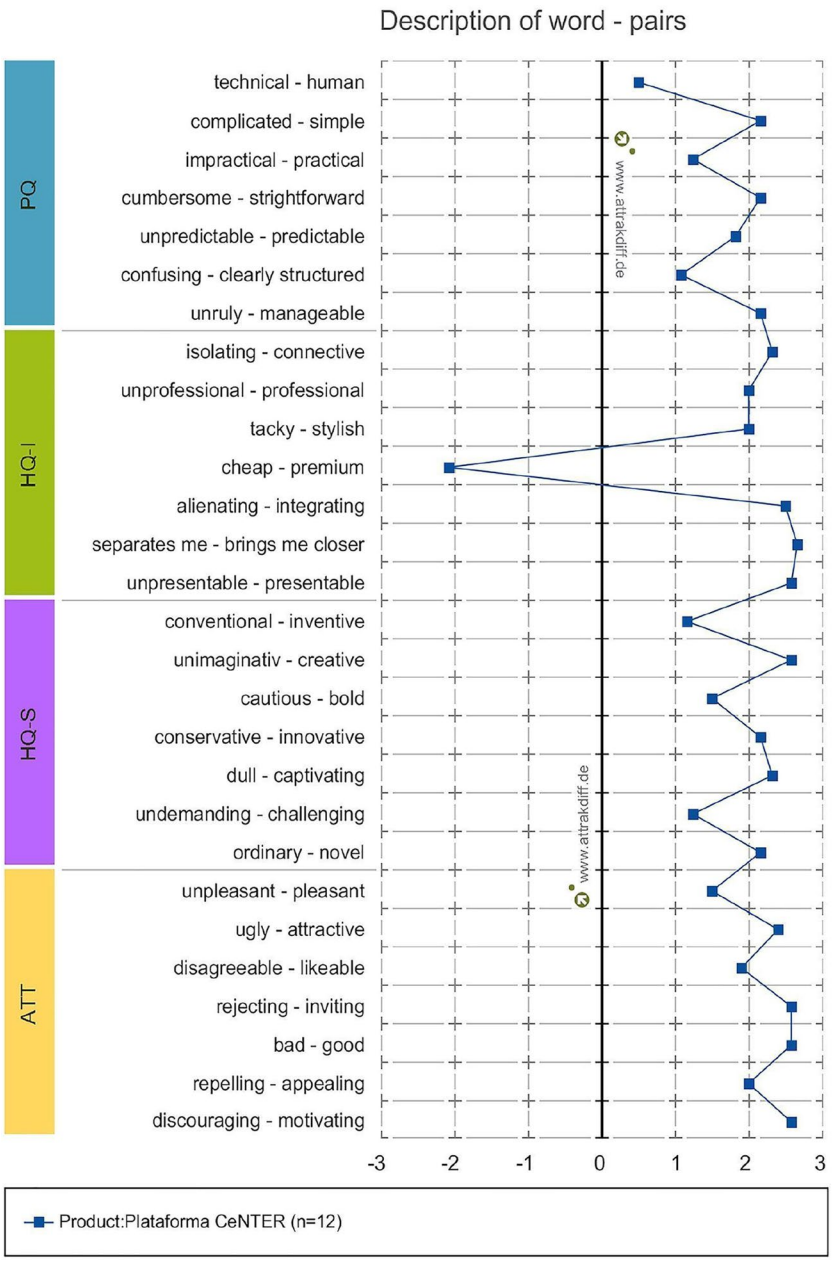

Fig. 4 Diagram of the description of word pairs. Global average of measured items

indicate that the UC2 tasks (public entities) were challenging for the local agents, reflecting the evaluation of AttrakDiff. Additionally, it is noteworthy that the UC4 presented higher scores in effectiveness, while the UC2 had better values in terms of efficiency. Regarding SUS and AttrakDiff, the higher average punctuation was given by the end-users of the UC3, showing a higher level of satisfaction concerning the CeNTER prototype.

Qualitative Results

The Think-aloud protocol was used to obtain immediate feedback from end-users about their experience of interacting with the prototype. The application of this method allowed the qualitative evaluation of the prototype based on the users' verbal comments. The inputs were divided according to each corresponding screen to relate user comments to the main screens tested. Table 5 shows the number of inputs for each screen, showing that the main screen stood out (13/46), as well as the details screen (9/46).

Forty-six (46) inputs were reported during the free exploration by the end-users, 36 of which were considered by the team as suggestions for platform improvements, seven as prototype usability errors, and three were interpreted as suggestions for improvement and usability errors. Usability errors correspond to inconsistencies in the interface's use,

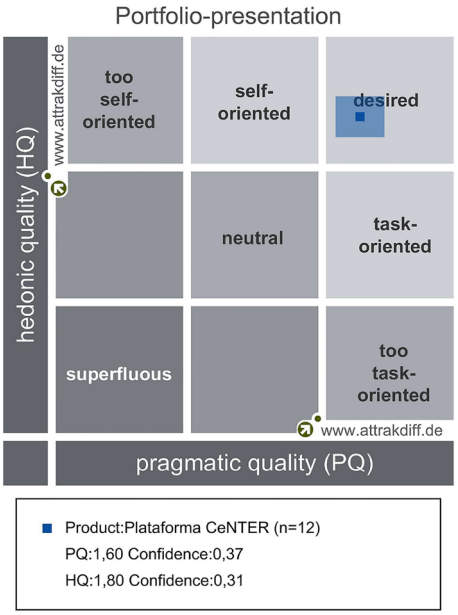

Fig. 5 Confidence rectangles of the evaluation with end-users

Table 5 Inputs according to the prototype interface

Interfaces	Nº of inputs
Tutorial	1
Main screen	13
Profile	5
Register of an initiative or event	1
Ideas	4
Maps	4
Agenda	2
Saved	2
Notifications	0
Details of an event/initiative/entity	9
Others	5
Total	46

such as the lack of feedback on acting, the need to do more than three steps on one of the screens to return to the home screen, and the difficulty moving the cards on the carousel. It should be noted that some of these problems were related to the limitations of the software used in the prototyping process, for example some difficulties that the evaluators felt in the movement of the cards. Improvement suggestions were related to the possibility of changing the main screen according to user preferences; apply search filters to the schedule; replace the title "Ideas" with a more dynamic one, such as: "Get your idea moving".

The largest number of suggestions for improvement was related to the suggestion of new features (10/39). As an example of these suggestions, we can mention the suggestions of "include supply/demand for an employee, in addition to a volunteer"; "Generate certificate of participation of volunteers"; "To be able to invite participants who have participated in previous events". It is important to note that these suggestions are precious for developing this mobile application and future digital solutions aimed at community initiatives. In addition to these inputs, the project team also collected 34 positive comments on the mobile application under development. These comments showed that users have fully understood the purpose and objectives of the platform. For example: "I liked the fact that I could cross similar initiatives, access it, either by map or by theme. I liked the possibility of creating a synergy between the partners"; "The synergies created within the application allow us to create new forms of interactions, which current applications would not yet allow".

Discussion

Usability in mobile devices is essential and should be considered when launching a new application, as it can make a difference in its acceptance by users. As reported in previous studies, there is a need for new usability assessment methods or, at the very least, usability assessments should be particularised for touchscreen-based mobile devices [19, 20, 40].

The techniques used for the evaluation, with both experts and end-users, allowed to collect valuable data, inquiring persons from different areas of knowledge. In this sense, a set of 13 specific usability heuristics was applied for expert testing. The initial validation proved its usefulness and potential. However, it demonstrated the need to carry out tests with end users as well [40]. As Jaspers [7] points out, "usability experts apparently seldom have enough knowledge of the users' work domain to evaluate, for example, whether the system interaction structure follows users' task flow" (p.348). Further, the importance of using protocols to assess user experience in the field of digital technologies has been previously presented in the literature, specifically the use of Attrakdiff and SUS has been reported [41]. In this direction, Attrakdiff has been reported as a feasible tool for evaluate the User Experience, differing from more traditional assessments of perceived usability by virtue of their more direct focus on the perceived subjective user experience (e.g., hedonic plus ergonomic/pragmatic quality). For example, the work in [33] recently used a version of AttrakDiff to explore aspects of experience-oriented and product-oriented evaluation. Moreover, SUS evaluation seems to complement the Attrakdiff UX, since they are commonly used together and provide ideas regarding what users think about the usability of the product, as well as a practical statistical evaluation which makes it easy to apply and obtain quantitative

outcomes. As reported by [42] the SUS, originally developed to be a "quick and dirty" measure of satisfaction with usability, has become one of the most popular post-study standardized questionnaires with practitioners, and recent research indicates that although it is quick, it is far from "dirty." [p. 186].

Furthermore, as we said before, metrics of effectiveness and efficiency are key factors related to the usability of a product [23]. While the efficiency metric considers the time that each evaluator took to complete the tasks (time execution), the effectiveness metrics indicates whether the user performed the task, with or without help, or did not perform it. Studies about efficacy and effectiveness have been published. For instance, [43] obtained significant results from evaluating a mobile application that uses machine learning, highlighting the relevance of these measures when developing digital technologies products. Therefore, the importance of using effectiveness and efficacy tests in product development has been proven. Concerning CeNTER application, the evaluation using these metrics provides information regarding the most challenging paths to accomplish a specific task since there are several ways to realise them. Thus, it hints at the functions and functionalities that should be worked on within the project.

Our methods can serve as a model of design best practises for other investigators seeking to develop applications that promote community empowerment. As already mentioned by several studies, to maximise the effectiveness and efficiency potential of this type of applications, the researcher's team must understand the needs and preferences of the communities in the use of these tools [44, 46]. A simple and intuitive interface, easily adaptable to different contexts of usage [47, 48] is an important tool for the users, represented by different agents of territorial development context (communities, entities, networks and citizens).

Also, the nature of the experts' area oriented the character of the collected suggestions, with the most technical issues naturally being pointed out by the professionals of the technological area. In this respect, Almeida et al. [41] stands out the importance of using iterative phases for the evaluation of a digital platform. Almeida et al. [49] also emphasise the experts' suggestions about UI design and interaction as a key factor for the development of a satisfactory and disruptive user interface.

Our iterative intervention design process, which was conducted in collaboration with different local agents (communities, public entities, small companies), allowed us to determine the critical features and functions to be integrated into the developed application. These features, such as sharing resources, discussing ideas, making volunteers available, were highlighted as valuable and innovative by end-users. Other studies prove that this type of intervention promotes viability and acceptability of mobile applications [24, 43].

Limitations and Future Work

This study had limitations related to the relatively small sample size, as it was relatively short, which restricted the generalisation of the results. However, it seemed to be sufficient for the execution of usability tests. Another restriction is related to the fact that the Principle software does not allow some types of interactions, such as the pinch gesture (pinching to zoom in and zoom out on the map of a mobile touchscreen application), insert personalised data by the user (the prototype only simulates the information that the user entered) or some limitations in gestures such as drag and drop (it is not possible to use the same graphic object to perform two different drag functions). However, these limitations were not an impairment for a good user experience evaluation. The main positive results obtained from the evaluation tools are a good indicator of the acceptance and a pleasant experience concerning the use of the prototype.

As a future work, it is aimed to develop a fully functional platform, which will allow the experimentation and evaluation of the application in the context of community-led initiatives. Nonetheless, it is intended to study the adoption, use and impact of the application in promoting processes of articulation and approximation between local agents, as well as in the construction and diffusion of knowledge and innovations.

Conclusions

As the development of the CeNTER prototype application required an iterative design process, in order to detect usability issues and identify important improvements to be made, the performance of the early tests was a crucial part of the system design. Furthermore, the iterative method of production of the application supported the methodology oriented to the needs and requirements of users, while being anchored in a continuous spiral of evaluation-correction of the prototype, resulting in improved versions of the pilot platform. This became evident, with the satisfactory results that emerged from the evaluation with end-users. Therefore, the developmental characteristics made it possible to build highly usable products. Simultaneously, the evaluation model made it possible to create a product with a more attractive look and feel.

Usability tests proved to be an effective way to acquire information that contributes to significantly improving the interface of a future mobile application, thus favouring the user experience. The user-centred design approach, used in all stages of the CeNTER prototype development, contributed strongly to understanding the users' needs.

The application of the Cognitive Walkthrough method and the Think-aloud protocol, together with the SUS and AttrakDiff, allowed the integration of quantitative and qualitative assessment approaches in this study. The different analysis methods with metrics of usability provided a multifaceted understanding of what local agents expected and how they intend to interact with the mobile application during their community and/or professional activities. Instrumental and non-instrumental characteristics of the prototype allowed us to obtain information, in addition to the usability data, providing results on aesthetic and emotional aspects related to the platform.

It is essential to highlight that the number of usability problems identified during the end-users' tests has significantly decreased when compared to tests with experts. As previously mentioned, in usability tests conducted with experts 50 problems were identified and the vast majority of those were soon corrected. After that, only seven usability problems were identified in the 12 tests applied with end-users.

The analysis of the data collected indicates good usability and high values of acceptance and satisfaction from the different local agents with the developed prototype. This tends to demonstrate the relevance of the end-user-centered approach to the development of tools dedicated to territorial based innovation. The sample size was composed by three evaluators per use case, demonstrating that the prototype had a good efficiency index, since it obtained a classification of 80% or more in all use cases.

The difficulties in carrying out tasks that required content creation and consultation by some evaluators were influenced by the learning curve of users, as well as their experience in using similar mobile applications. Although the design of this study does not allow us to generalise its results, they reflect the user experience of the regional agents previously selected, providing evidence of what is important for a mobile application in territorial-based innovation.

User tests positively highlighted several platform features, such as sharing resources and volunteers, collaborative development of events, sharing ideas and creating new initiatives based on these ideas. Also, many users reported that these are innovative features, which increase the relevance of the CeNTER platform as an original and useful option.

As a final conclusion, it was possible to learn several essential lessons throughout this collaborative process, which can be useful for other researchers who develop digital solutions in the same subject area: (1) include community initiatives in the entire design process to better tailor the solution to their needs; (2) be flexible to meet the preferences of the community and of the stakeholders; (3) incorporate mixed methods in design and assessment tests, which provide valuable information to produce an acceptable and well-designed solution.

Acknowledgements This article was developed under the support of the Research Program "CeNTER—Community-led Networks for Territorial Innovation" Integrated Research Program (CENTRO-01-0145-FEDER-000002), funded by Programa Operacional Regional do Centro (CENTRO 2020), PT2020."

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

1. Encalada L, Boavida-Portugal I, Ferreira C, Rocha J. Identifying tourist places of interest based on digital imprints: Towards a sustainable smart city. Sustainability. 2017;9(12):2303–17. https://doi.org/10.3390/su9122317.
2. Bonomi S, Ricciardi F, Rossignoli C. Network organisations for externality challenges: how social entrepreneurship co-evolves with ICT-enabled solutions. Int J Knowl Based Dev. 2017;8(4):346–66.
3. Saint-Onge H, Wallace D. Leveraging communities of practice for strategic advantage. Routledge; 2012
4. Tymoshchuk O, Almeida AM, Pedro L, Antunes MJ, Ramos F, Oliveira E, Carvalho D. Digital technologies as tools to promote tourism and territorial development: design of a mobile application for community-led initiatives. In: Dinis MG, Bonixe L, Lamy S, Breda Z (eds). Impact of New Media in Tourism. IGI Global; 2021. pp. 268–291. https://doi.org/10.4018/978-1-7998-7095-1.ch016
5. Branco AC, Carvalho D, Sacramento, E, Tymoshchuk O, Oliveira E, Antunes MJ, Pedro L, Almeida AM, Ramos F. Prototyping and evaluating a mobile app to promote territorial innovation. In: Proceedings-Conference Communities and Networks for Territorial Innovation. 2021. pp. 73–81.
6. Oliveira E, Tymoshchuk O, Branco A, Carvalho D, Sacramento E, Antunes MJ, Pedro L, Almeida AM, Ramos F. End-user evaluation of a mobile application prototype for territorial innovation. In: Proceedings of International Conference on Enterprise Information Systems (ICEIS). 2021. pp. 495–504.
7. Jaspers MW. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. Int J Med Informatics. 2009;78(5):340–53.
8. Fonseca MJ, Campos P, Gonçalves D. Introdução ao Design de Interfaces. 2nd ed. FCA: Editora de Informática; 2012.
9. Nielsen J. Usability inspection methods. In: Conference companion on Human factors in computing systems. 1994. pp. 413–414.
10. Still B, Crane K. Fundamentals of user-centered design. Nova York: CRC Press; 2017.
11. Knight W. UX for developers. Northampton: Apress; 2019.
12. Bernhaupt M, Pirker R. Evaluating user experience for interactive television: towards the development of a domain-specific user experience questionnaire. Hum Comput Interact Lect Notes Comput Sci. 2013;8118:642–59.
13. Rogers Y, Sharp H, Preece J. Interaction design: beyond human-computer interaction. John Wiley and Sons; 2011.
14. Welfer D, da Silva RC, Kazienko JF. MobiCAP: A mobile application prototype for management of community-acquired pneumonia. In: 2014 IEEE 16th International Conference on e-Health Networking, Applications and Services (Healthcom). 2014. pp. 126–127.

15. Jones N, Pu P. User technology adoption issues in recommender systems. In: Proceedings of the 2007 Networking and Electronic Commerce Research Conference. 2007. pp. 379–394.
16. Dourado MAD, Canedo ED. Usability heuristics for mobile applications a systematic review. In: Proceedings of the 20th International Conference on Enterprise Information Systems. 2018. pp. 483–494.
17. Nielsen J. How to conduct a heuristic evaluation. Nielsen Norman Group. 1995;1:1–8.
18. Nielsen J. 10 Usability heuristics for user interface design. https://www.nngroup.com/articles/ten-usability-heuristics/ Accessed 15 Nov 2020.
19. Reolon M, Lacerda TC, Krone C, von Wangenheim CG, Xafranski J, Nunes JV, von Wangenheim A. Usability heuristics for evaluating health-care applications for smartphones: a systematic literature review. Florianópolis: INE, UFSC; 2016.
20. Lacerda TC, Nunes JV, von Wangenheim CG. Usability heuristics for mobile phone applications: a literature review. In: Emerging perspectives on the de-sign, use, and evaluation of mobile and handheld devices. IGI Global; 2015. pp. 143–157.
21. Wangenheim CG, Wangenheim A, Machado E, Lacerda T, Azevedo LF. MATCH-MED: Heurísticas e Checklist para Avaliação da Usabilidade de Aplicativos mHealth em Smartphones. Florianópolis: INE, UFSC; 2017.
22. Sauro J, Lewis JR. Quantifying the user experience. 2nd ed. Practical Statistics for User Research. Cambridge: Elsevier; 2016.
23. Rubin J, Chisnell D. How to plan, design, and conduct effective tests. Handbook of usability testing. Wiley Publishing, Inc; 2008.
24. Sharp H, Rogers Y, Preece J. Interaction design: Beyond human computer interaction. 2nd ed. New York: John Wiley & Sons; 2007.
25. Preece J. Online Communities: Designing Usability. Supporting Sociability: John Wiley; 2000.
26. Carroll C, Marsden P, Soden P, Naylor E, New J, Dornan T. Involving users in the design and usability evaluation of a clinical decision support system. Comput Methods Programs Biomed. 2002;69(2):123–35.
27. Nielsen J. Usability testing. Handbook of human factors and ergonomics. 1997;2:1543–68.
28. Muchagata J, Ferreira A. Visual Schedule: A Mobile Application for Autistic Children-Preliminary Study. In: Proceedings of ICEIS. 2019. pp. 452–459.
29. Lorenz T, Richter A, Köster F, Rataj J. Iterative design process for the development and testing of cooperative applications. In: Proceedings of the 2010 Spring Simulation Multiconference. 2010. pp. 1–6.
30. Salgado ADL, Pontin de Mattos Fortes R, Hung PC, Moreira DDA. A method for classifying usability findings to enhance validation of new heuristics. Revista de Sistemas e Computação-RSC. 2019;9(1):96–118.
31. Hashim N, Isse A. Usability evaluation metrics of tourism mobile applications. J Softw Eng Appl. 2019;12:267–77. https://doi.org/10.4236/jsea.2019.127016.
32. Martins AI, Rosa AF, Queirós A, Silva A, Rocha N. European portuguese validation of the system usability scale (SUS). Proc Comput Sci. 2015;67:293–300.
33. Hassenzahl M, Burmester M, Koller F. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: Mensch and computer. Vieweg Teubner Verlag; 2003. pp.187–196.
34. Nielsen J. Usability Metrics. 2001.https://www.nngroup.com/articles/usability-metrics/. Accessed 16 Jul 2021.
35. Wharton C, Rieman J, Lewis C, Polson, P. The Cognitive Walkthrough: A practitioner's guide. In J. Nielsen RL. Mack (Eds.), Usability inspections methods. New York: Wiley; 1994. pp. 105–140.
36. Silva PA, Antunes MJ, Tymoshchuk O, Pedro L, Almeida M, Ramos F. Understanding the role of communication and mediation strategies in community-led territorial innovation: a systematic review. Interact Des Arch J (IxD&A). 2020;44:7–28.
37. Tymoshchuk O, Renó D, Silva PA, Almeida AM, Pedro L, Ramos F. Mediação digital para a inovação territorial: um estudo de caso múltiplo em redes sociais digitais. OBS Observatorio. 2019;13(4):70–88.
38. Sauro J. Measuring usability with the system usability scale (SUS). Measuring U. 2011.
39. Barbosa A. Medindo a usabilidade do seu produto com System Usability Scale (SUS). CA Design. 2019.
40. Inostroza R, Rusu C, Roncagliolo S, Rusu V. Usability heuristics for touchscreen-based mobile devices: update. In: Proceedings of the 2013 Chilean Conference on Human-Computer Interaction. 2013. pp. 24–29.
41. Almeida P, Abreu J, Silva T, Varsori E, Oliveira, E, Velhinho A, Oliveira D. Iterative user experience evaluation of a user interface for the unification of TV contents. In: Iberoamerican Conference on Applications and Usability of Interactive TV. Springer, Cham. 2017. pp. 44–57. https://doi.org/10.1007/978-3-319-90170-1_4
42. Sauro, Jeff, and James R. Lewis. Quantifying the user experience: Practical statistics for user research. Morgan Kaufmann, 2016.
43. Shuba A, Markopoulou A, Shafiq Z. Nomoads: effective and efficient cross-app mobile ad-blocking. Proc Privacy Enhancing Technol. 2018;2018(4):125–40.
44. Brewer LC, Hayes SN, Caron AR, Derby DA, Breutzman NS, Wicks A, Patten CA. Promoting cardiovascular health and wellness among African-Americans: community participatory approach to design an innovative mobile-health intervention. PLoS ONE. 2019. https://doi.org/10.1371/journal.pone.0218724.
45. Stowell E, Lyson MC, Saksono H, Wurth RC, Jimison H, Pavel M. Designing and evaluating mhealth interventions for vulnerable populations: a systematic review. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. Montreal QC, Canada. ACM; 2018. p. 1–17.
46. Veinot TC, Mitchell H, Ancker JS. Good intentions are not enough: how informatics interventions can worsen inequality. J Am Med Inform Assoc. 2018;25(8):1080–8.
47. Feijó VC, Gonçalves BS, Gomez LSR. Heurística para avaliação de usabilidade em interfaces de aplicativos smartphones: Utilidade, produtividade e imersão. Design e Tecnologia. 2013;3(06):33–42.
48. Kumar BA, Mohite P. Usability of mobile learning applications: a systematic literature review. J Comput Educ. 2018;5(1):1–17.
49. Almeida P, Abreu J, Oliveira ER, Velhinho A. Expert evaluation of a user interface for the unification of TV contents. In: Proceedings of the 6th Ibero-American Conference on Usability and Interactive TV Applications—jAUTI'2017. 2017. pp. 59–70.