Sobrino-Duque, R., Carrillo-de-Gea, J. M., López-Jiménez, J. J., Nicolás Ros, J., & Fernández-Alemán, J. L. (2022). Usevalia: Managing Inspection-Based Usability Audits. *International Journal of Human–Computer Interaction*, *40*(3), 719–743.

Usevalia: Managing Inspection-Based Usability Audits

Raimel Sobrino-Duque, Juan Manuel Carrillo-de-Gea, Juan José López-Jiménez, Joaquín Nicolás*, José Luis Fernández-Alemán

Faculty of Computer Science, Department of Informatics and Systems, University of Murcia, Spain
Corresponding author. Faculty of Computer Science - 30100 Espinardo Campus, Murcia, Spain - Telephone: (+ 34) 868888525; fax: (+ 34) 868884151
Email-addresses: raimel.sobrino@um.es (Raimel Sobrino-Duque), jmcdg1@um.es (Juan Manuel Carrillo-de-Gea), juanjoselopez@um.es (Juan José López-Jiménez), jnr@um.es (Joaquín Nicolás Ros), aleman@um.es (José Luis Fernández-Alemán)

Abstract

Heuristic evaluation provides usability auditors with a structured flow of activities that enables them to obtain reliable, comparable and cost-effective usability evaluation results. Tools supporting heuristic evaluation are scarce in literature. Usevalia is an Internet-based tool that manages usability audits on the basis of heuristic evaluation. This paper presents the main features of Usevalia, together with a validation of the software by following the Technology Acceptance Model, which verified the users' perceptions of Usevalia as regards usefulness, ease of use, attitude and intention to use. A total of 22 students enrolled on a human-computer interaction course were involved in the validation. According to the results of the survey conducted with them, the proposed tool is easy to use (MD = 4.00 out of 5) and useful for conducting usability audits based on heuristic evaluation (MD = 4.00 out of 5). An expert-based validation was also carried out in order to thoroughly compare Usevalia's features with those of a spreadsheet-based tool that performs usability audits in the traditional manner. This comparative analysis made it possible to conclude that the Usevalia tool has a higher perceived usefulness (M = 4.43) and perceived ease of use (M = 4.13) than a traditional audit tool such as the Usability Datalogger spreadsheet (M = 2.43 and M = 3.03, respectively).

Keywords: Usability, Heuristic Evaluation, Usability Audit Tool, TAM

---

1. INTRODUCTION

One of the key attributes of software solutions that requires careful attention and evaluation is usability (Barnum, 2021). According to ISO 9241 (2018), usability is “the extent to which a system, product, or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use”. High usability (1) makes it possible to reduce task execution times, errors and learning times; (2) improves user satisfaction, and (3) leads to improved product acceptability, increased user satisfaction and better product reliability (Barnum, 2021; Güncan & Onay Durdu, 2021).

A 2016 study of 408 different companies found that the more companies invested in and focused on corporate website design, the more sales they achieved, the higher their customer retention and the faster they moved through their product cycles (Career Foundry 2016). All this was simply because these companies ensured that design, and more importantly, the user, were at the very core of their business. For example, the ease or difficulty confronted by users of e-commerce applications greatly determines their success or failure. Tharindu and Koggalage (2021b) suggest that poor usability is no longer tolerated by users and that more than 83% are likely to leave a website if they feel they have to make too many clicks to find what they are looking for (Tharindu & Koggalage, 2021).

The objective of usability audits is to enable users to make use of applications with the least number of usability problems. These audits are the “collection of methods, skills and tools used to study and analyse how users interact with a website” (Intechnic, 2016). The analysis is then used to make decisions and recommendations (based on facts, research and data) concerning how to tweak and optimize the applications.

In order to enable a continuous evolution of the quality of the application, the usability evaluation must be iterative and be present at all development stages (Krug, 2014). Literature describes several models, methods and techniques with which to ensure that usability issues are considered during the software development process and to avoid significant differences as regards identifying usability problems, as it has been demonstrated that there is great variability in the results of usability evaluations (Molich et al., 2004). The selection of these models, methods, and techniques depends on the development stage of the informatics solutions and the resources available (Morrissey, 2014).

Usability evaluation methods are divided into three categories, namely: (1) inspection methods, (2) inquiry methods and (3) testing methods (Hom, 1998). Inspection methods are carried out by a set of experts who examine the interface to verify its usability, while inquiry methods involve talking to users and observing them closely using the system or obtaining answers to questions verbally or in writing. Finally, testing methods seek the reason why users are interacting poorly with the interface. Usability testing can be performed in-lab or remotely, the latter either synchronously or asynchronously (Alhadreti, 2022). The inspection method contains several techniques that can be carried out on both a real interface and a prototype. These

techniques include (Hom, 1998): (1) heuristic evaluation; (2) feature inspection; (3) consistency inspections; (4) standards inspections; (5) formal usability inspections; (6) guideline checklists; (7) cognitive walkthrough, and (8) pluralistic walkthrough.

Heuristic evaluation is a well-known usability engineering technique for the faster and more cost-efficient evaluation of the usability problems in a user interface (Hollingsed & Novick, 2007). According to Nielsen and Molich (1990), heuristic evaluation involves having a small set of evaluators who examine the interface and judge its compliance with recognized usability principles. This is performed by individual evaluators, who each inspect the interface alone on the basis of the analysis of a set of design guidelines, which are nothing more than series of indicators with which a system must comply, and to which a priority or importance is assigned. Richardson et al. (2021), for instance, propose a mobile application user experience checklist called MAUX-C. The evaluation indicates the level of compliance with the checklist of guidelines. As a result of the evaluation, a list of usability problems with references to those usability principles that were, in the evaluator's opinion, violated by the design is obtained. It is not sufficient for evaluators to simply state that they do not like something; they should explain why they do not like it with reference to the heuristics or to other usability issues. The evaluators should attempt to be as specific as possible and should list each usability problem separately.

The use of automated tools to support the usability audit process is a useful complement and addition to standard evaluation techniques such as heuristic evaluation and usability testing. The automation of usability evaluation is of great importance because it has several potential advantages over non-automated evaluation. Some of these advantages are (Ivory & Hearst, 2001): (1) a reduction in the time spent on usability evaluation, and consequently the cost; (2) an increased consistency of the errors discovered; (3) predictability of time and error costs across an entire design, as it is not always possible to evaluate all aspects of an interface by means of non-automated evaluation, and the use of automated tools allows the expanded coverage of evaluated features; (4) a reduction in the need for the individual evaluators to have evaluation expertise, and (5) it is possible to compare alternative designs. Time, cost, and resource constraints lead to a situation in which usability evaluations often assess only one design or a small subset of features from several designs. Some automated analysis approaches, such as analytical modeling and simulation, allow designers to compare the expected performance of alternative designs; finally, another advantage is (6) that the use of automated tools allows evaluation to be incorporated into the design phase of user interface development, rather than after implementation. In particular, the automation of heuristic evaluation procedures facilitates the detection of usability defects, thus enhancing the effectiveness of usability evaluation (Sivaji et al., 2011). To best of our knowledge, there is a need to develop automated support tools that have the functionalities that auditors currently require in order to ease usability audits.

The objective of this paper is to present Usevalia, a tool that allows the management of audits on the basis of heuristic evaluation. Usevalia can help software engineers and HCI (Human-Computer Interaction) professionals perform usability audits on the basis

of heuristic evaluation, leaving little room for the traditional manual use of checklists. The structure of the remainder of this paper is as follows: Section 2 summarizes related work concerning the management of usability audits based on heuristic evaluation; Section 3 illustrates the design and main features of Usevalia; Section 4 includes the evaluation of this automated support, and finally, Section 5 highlights our conclusions and shows our future work.

2. RELATED WORK

The heuristic evaluation process has, to date, traditionally been organized in the form of checklists using spreadsheets. The procedure is performed on a computer or any other device capable of processing and editing spreadsheets, and the quantitative analysis is performed automatically according to the formulae entered in the corresponding cells. One of the templates employed to support this process is the Heuristic Evaluation-A System Checklist (WebCriteria, 2002), a simple list with four columns: guideline identifier, guideline, rating or level of compliance and optional comments. There are also tools containing predefined templates that allow the user to read each of the guidelines and evaluate them. In a literature review conducted by the authors, the following tools were found: (1) Usability Test Datalogger (Zazelenchuk, 2008); (2) Sirius (Suarez Torrente et al., 2016); (3) Prometheus (Chamba-Eras et al., 2017), and (4) SUIT (Systematic Usability Inspection Tool) (Ardito et al., 2006).

2.1. Usability Test Datalogger

The Usability Test Datalogger consists of a template oriented toward requirements management which can easily be modified to be used as an audit tool. The removal or addition of guidelines is simple, and it additionally has the possibility of including more than one auditor. This tool has been used in large companies such as AXA Insurance or Skype. Moreover, version 5.1.1 is currently available in English, although it was last updated in August 2008.

The right-hand side of Figure 1 shows that the tool allows the user to add all the information related to the audit to be performed, including dates and the name of the project to be evaluated. The tool supports audits with multiple evaluators and allows the user to indicate which auditors will be included in the final statistics. The tool also makes it possible to establish a customized scoring system, indicating which values are satisfactory and which are not. The left-hand side of Figure 1 shows how the guidelines to be evaluated in the audit can be added. The guidelines cannot be grouped into different sections, and neither description nor priority can be added for each guideline.

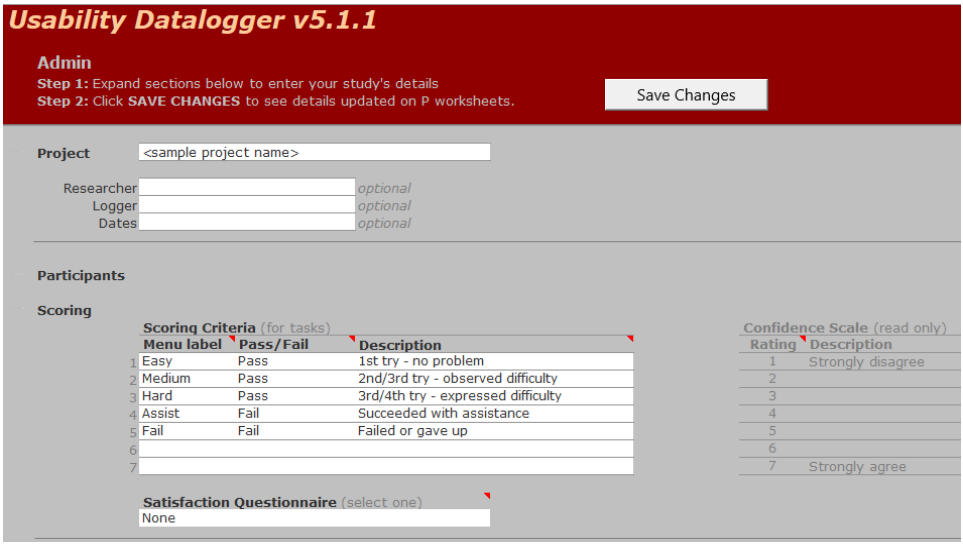


Figure 1. Usability Test Datalogger admin worksheet and configuration of the guidelines

One interesting feature of this tool when compared to the others on the market is that it includes a sheet with which to visualize statistics and graphs from the audit results. Some of the metrics that it provides are the average time needed by all participants to evaluate each guideline or the percentage of participants who have considered a guideline to be successful.

Usability Test Datalogger does not allow the downloading of a document with the result of the audit, including the scores awarded and the comments made by the evaluators. As the tool is supported by a spreadsheet, it does not allow the scoring systems or the list of guidelines to be saved for their subsequent use in another future audit, and it is thus necessary to make a copy of the sheet for each audit performed, which can be tedious and cause problems if the wrong sheet is modified.

2.2. Sirius

Suárez et al. (2016) proposed the Sirius tool (see Figure 2) in order to carry out audits of design guidelines following the Sirius evaluation method. Unlike Usability Test Datalogger, the Sirius tool focuses on a single user only, but its evaluation results are more detailed. It was last updated in 2011, although an English version was published in July 2019[1]. Each group of guidelines in the same catalogue is separated into several sheets of the document. The audit data and the evaluator are on the first sheet. This spreadsheet does not allow the creation of a customized scoring system or the establishment of the guidelines to be evaluated. Sirius has its own heuristics and scores with which to calculate the percentage of usability. Some guidelines will have greater or lesser relevance, depending on the type of website.

1 Accessible from https://olgacarreras.blogspot.com/2011/07/sirius-nueva-sistema-para-la-evaluacion.html

One drawback of Sirius is that it does not allow the downloading of documents showing the evaluation results, nor does it show graphics comparing each group of heuristics or show statistics regarding the evaluation. Sirius indicates only the usability percentage, together with certain calculations that are necessary in order to obtain that percentage. In general, Sirius is barely customizable, partly because it is an adaptation of a specific heuristic evaluation method.

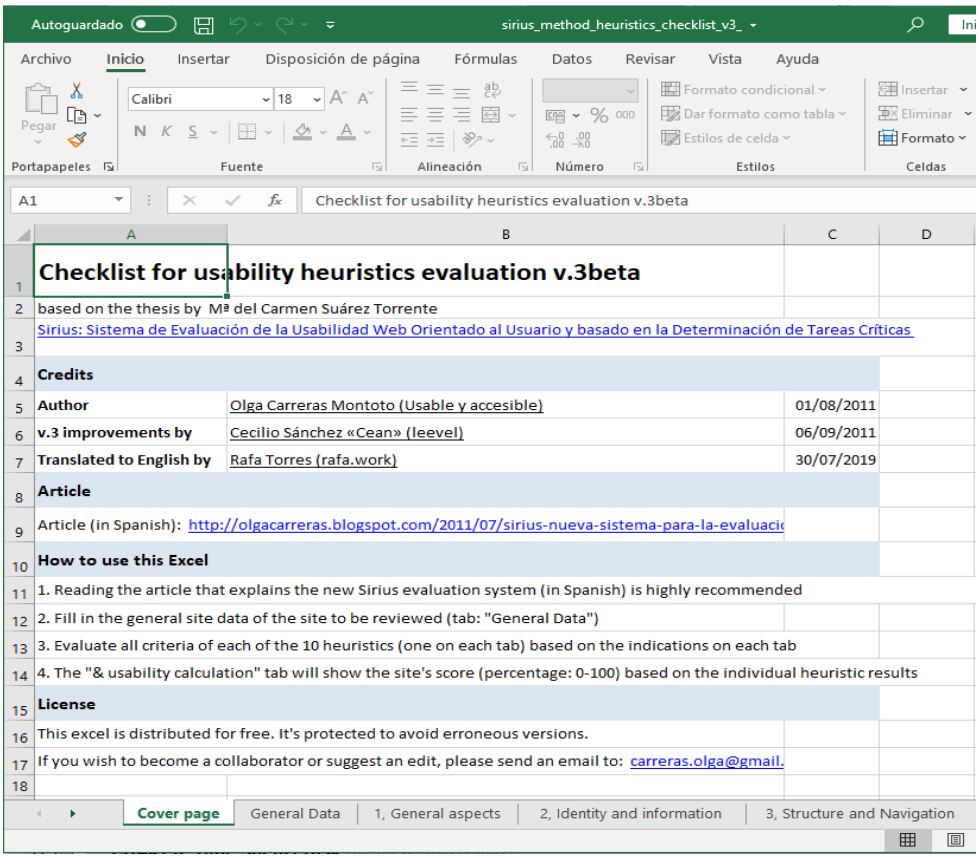


Figure 2. Configuration of an audit in Sirius

2.3. Prometheus

The Prometheus tool (Chamba-Eras et al., 2017) was developed to support the Sirius evaluation method. Prometheus has certain functionalities that the Sirius spreadsheet does not have, such as the possibility of establishing “hybrid websites”, indicating the percentage that each type of site listed in the tool includes. In this context, a hybrid website is defined as one that combines two or more of the 15 types defined by the Sirius tool, namely: (1) public administration/institutional; (2) online banking; (3) blog; e-commerce; (4) communication/news; (5) corporate/company; (6) downloads; (7) education/training; (8) collaborative environments/wikis; (9) virtual community/internet forum; (10) leisure/entertainment; (11) personal; (12) service portal; (13) image-based

interactive services; (14) non-image-based interactive services, and (15) webmail/mail (Suarez Torrente et al., 2016). This allows the calculation of the usability of the evaluated site to be more accurate when compared to the spreadsheet version of Sirius (see Section 2.2), which enables only the selection of a specific type for each application evaluated.

This tool does not have multi-evaluator support, but it does allow the results of the evaluations to be downloaded in an EARL (Evaluation and Report Language) report (W3C Working Group Note, 2017). The evaluation and score assignment process is like the spreadsheet version of Sirius, since it is a predetermined list of heuristics. There is also the option of opening the website being evaluated in a tab.

A notable difference with respect to the spreadsheet version of Sirius is the "owner" role. It is not the evaluator who adds the information concerning the website, but rather the owner who enters the sites to be evaluated into the system, with their respective information and type of site. This signifies that the evaluator has only to select the site to be evaluated from those available and assign the scores he/she considers appropriate, thus reducing the evaluator's work and allowing him/her to focus on the evaluation itself. Figure 3 shows a diagram depicting the users represented in the tool.

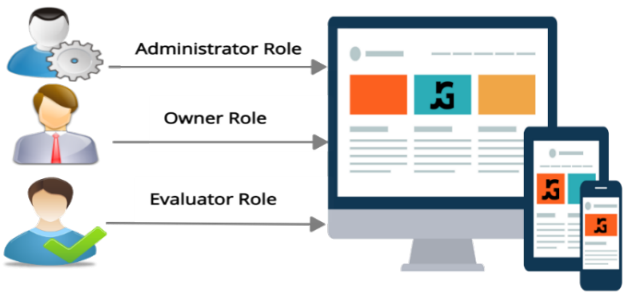


Figure 3. Prometheus role diagram

2.4. SUIT

SUIT (Systematic Usability Inspection Tool) (Ardito et al., 2006) was presented in 2006 to support a particular type of evaluation called SUE (Systematic Usability Evaluation) (Matera et al., 2002). This method has been denominated as SUE inspection and exploits a set of evaluation patterns, called Abstract Tasks (ATs), which guide the inspector’s activities, precisely describing which objects of the application to seek and which actions to perform during the inspection in order to analyze those objects.

Two phases are required during the AT inspection: (1) the Preparatory phase, which is performed only once for the kind of applications to be evaluated (for example, e-learning applications), and (2) the Execution phase, which is performed for each application to be evaluated. The Execution phase is coordinated by the so-called inspection manager, an expert inspector who decides how many inspectors should be involved, selects them, and coordinates their activities. The Execution phase consists of four stages: (1) Inspection planning; (2) Error detection; (3) Report generation, and (4) Discussion.

In the Report generation stage, the inspection manager creates a list of the problems of the application being evaluated by analyzing the inspector reports made available by SUIT. The inspector manager must integrate the problems detected by the evaluators. SUIT helps in this task by providing the possibility of sorting the list of problems with respect to various parameters. The inspection manager then manually highlights any problems on which two or more inspectors did not agree. If two inspectors give two different ratings to the same problem, an inspection team meeting is held, and an agreement should be reached in the final report. SUIT makes it possible to avoid face-to-face meetings by offering a discussion forum in which all inspectors, or a subset of them, participate. This attendance is decided by the inspection manager, who uses SUIT to send a notification e-mail to the inspectors involved. Figure 4 illustrates the way in which a usability problem is inserted into SUIT.

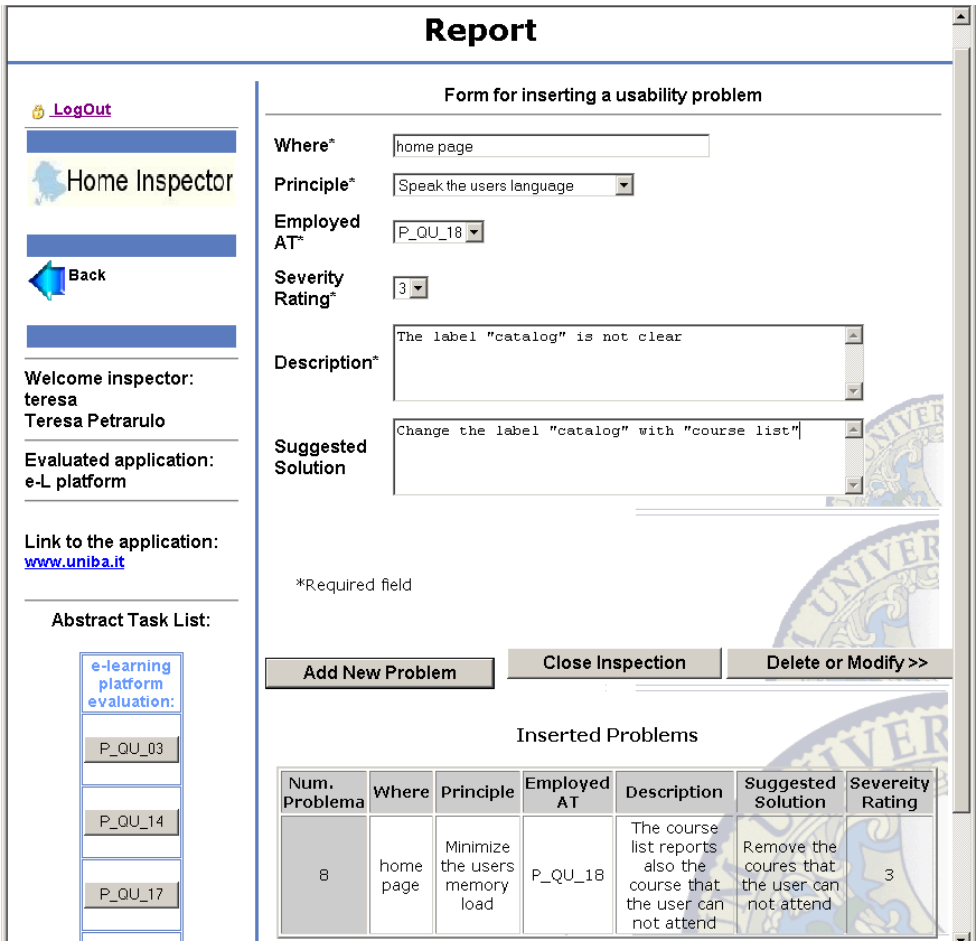


Figure 4. Form employed to insert a usability problem

When an evaluator reports a heuristic that is not met, he/she can indicate the level of severity of the failure, in addition to suggesting possible improvements by which to

solve it. Moreover, each evaluator can view the report of the others asynchronously in the discussion phase, and can then indicate whether or not the problems found are true by adding a message. This allows the evaluators to avoid collocated meetings in which to discuss the issues, and also allows them a certain flexibility of schedules, since they are able to work on the bug reports at the time that best suits them within the evaluation timeframe.

2.5. Comparative analysis of the tools

Table 1 presents a comparison between the tools described above together with Usevalia, which is presented in detail in Section 3. It is important to note that few usability audit tools are based on heuristic evaluation, and this could indicate the need to develop automated support tools that have the functionalities that auditors currently require in order to perform a usability audit. The comparison has been carried out by establishing a priority for each of the characteristics to be taken into account, based on the experience acquired during the expert evaluation presented in Section 4. In this context, the priority represents the relative level of importance of a given feature when compared to another characteristic. A three-point Likert-type scale (1 = High; 2 = Medium; 3 = Low) was used to define the priority. The decision was made to use a three-point Likert scale because Matell and Jacoby (1971) suggested that reliability and validity are independent of the number of response categories, and their results implied that collapsing data from longer scales into two-or three-point scales would not decrease the reliability or validity of the resulting scores.

In summary, Table 1 shows that Usevalia encompasses a set of functionalities that other tools do not have. For example, it enables the management of multi-user usability audits by allowing auditors to perform three types of evaluations (described in Section 3), while the others allow only the standard evaluation type. In addition, Usevalia supports the management of scoring schemes, which allow the type of value of the scheme to be specified. Another of the key functionalities of Usevalia that the other tools do not have is the possibility of importing customized guideline catalogues. Finally, although tools such as SUIT and Prometheus allow reports to be exported in different formats, Usevalia offers the possibility of adapting reports to the type of evaluation carried out by the auditors.

In addition to traditional functionalities, Non-Functional Requirements (NFRs, also known as quality requirements) (Cysneiros & Yu, 2004) establish constraints on the system and provide particular notions of the qualities that a system can have, such as accuracy, usability, safety, performance, reliability or security. Errors in the omission or inadequate treatment of NFRs are among the most costly and difficult to correct at a later date. NFRs such as internationalization and localization, energy efficiency, security and usability have been taken into account during the development of Usevalia (Cysneiros & Yu, 2004).

Table 1. Comparative analysis between tools.

Usability Audit Tools						
Features	**Priority**	**Usability Test Datalogger**	**Sirius**	**Prometheus**	**SUIT**	**Usevalia**
Users						
Role management	1	No	No	Manages three types of users	Manages two types of users	Manages three user types and manages user groups
Multi-auditor support	2	Yes	No	Yes	Yes	Yes
Concurrent multi-user connectivity	3	No	No	Yes	Yes	Yes
Audits						
Audit management	1	No	No	No	No	Yes
Customized scoring system	2	Yes	No	No	No	Yes
Types of usability evaluation supported	2	Standard	Standard	Standard	Standard	Basic, Standard and Task-based
Template for importing guideline catalogues	2	No	No	No	No	Yes
Reports						
Graphical comparison	2	Overall effectiveness and efficiency measures and users' satisfaction questionnaire results	Percentage of usability	Percentage of usability	List of issues	Total percentage of guidelines, percentage of guidelines by group, results by priority and results by guideline
Export report formats	3	No	No	EARL	PDF	PDF
Reports adapted to the type of evaluation	2	No	No	No	No	Yes
Others						
Internationalization and localization	2	No	Yes	Yes	Yes	Yes
Energy efficiency	2	No	No	No	No	Yes
Security and privacy	1	No	No	Yes	No	Yes
Usability	2	No	No	No	No	Yes

Usevalia is completely internationalized, and the language can be changed (English, Spanish) depending on the language preference of the user who is logged in. With

regard to energy efficiency, the Usevalia tool is deployed on a server that makes use of renewable energy sources and uses a MySQL database. The study carried out by Mahajan, Blakeney and Zong (2019) found that indexing in MySQL significantly improves performance and energy efficiency. Avoiding the use of the "SELECT *" command results in an increase in energy efficiency of more than 20%, which is significant owing to the ubiquity of the command among professionals. Security is another feature that is considered very important, and a comparative study of three CMSs (Joomla, WordPress, and Drupal) shows that Drupal has high security while the other CMSs have low and moderate security (Iqbal et al., 2020). The use of good security practices, such as using strong passwords, using SSL certificates, changing the administrator's root password and limiting login attempts, all contribute to ensuring data security and privacy. Finally, usability and responsive design were considered during the development of the tool. Responsive design has multiple advantages, such as enriching the user experience, improving statistical results, avoiding duplicate content, and reducing development and maintenance costs and time. Moreover, User-Centered Design was considered during the design of Usevalia because a well-designed user interface can make it easier for users to handle a system and lessen their cognition burden (Martin-Michiellot & Mendelsohn, 2000). The usability of the tool has been tracked during the successive stages of its development by means of an earlier version of the tool itself.

3. USEVALIA

Usevalia is a web tool that allows the management of usability audits on the basis of heuristic evaluation. Section 3.1 shows the functionalities of the tool, following a typical use scenario, and Section 3.2 then briefly describes the main design decisions and the Drupal-based architecture.

3.1. Workflow and features overview

Usevalia supports multi-user audits, thus allowing a team to work simultaneously on the same audit following a simple usability audit process that operationalizes the heuristic inspection. Although Usevalia does not prescribe any specific audit process, the flowchart in Figure 5 describes the activities of a typical Usevalia usability audit process. This process involves the participation of the so-called chief auditor, together with a team of auditors and the system itself. The audit starts when the chief auditor: (1) introduces the web or mobile application to be evaluated; (2) introduces the scoring schemes to be used; (3) introduces the deadline by which the audit must be completed, and (4) builds a team of auditors who will be in charge of performing the audit.

Scoring schemes can be created or deleted in the tool database. When creating a scoring scheme, the type of each of the values (e.g., high, medium, and low support) can be indicated. The deletion of a scoring scheme in the tool database is not permitted in case it is being used for an audit.

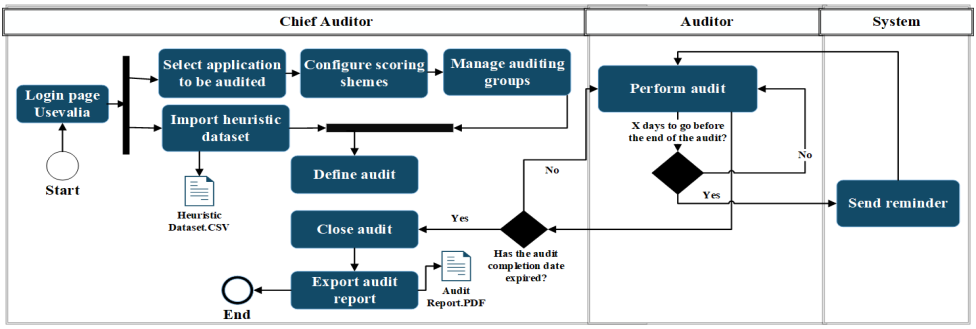


Figure 5. Execution flow of a usability audit

Usevalia allows the creation and deletion of teams of auditors. During the creation, the name of the team and the users who will play the role of auditor must be introduced, together with the description of the team. Auditor groups cannot be deleted if they have been linked to a heuristic dataset (i.e., to a guideline catalog) in an audit, until the audit is closed by the chief auditor.

The chief auditor imports the heuristic dataset that will be used to evaluate the application in the new audit (see Figure 6). Chief auditors have access to a repository of customizable templates that can be downloaded from the "Upload/Download catalog" page in order to create their own guideline catalogs. Guidelines to be added to the catalog can be defined from scratch or from other lists of heuristics, such as those defined in John Brooke's questionnaire (Brooke, 1996), Dix's fourteen usability principles (Dix et al., 2004) or Jacob Nilsen's ten heuristics (Nielsen & Molich, 1990).

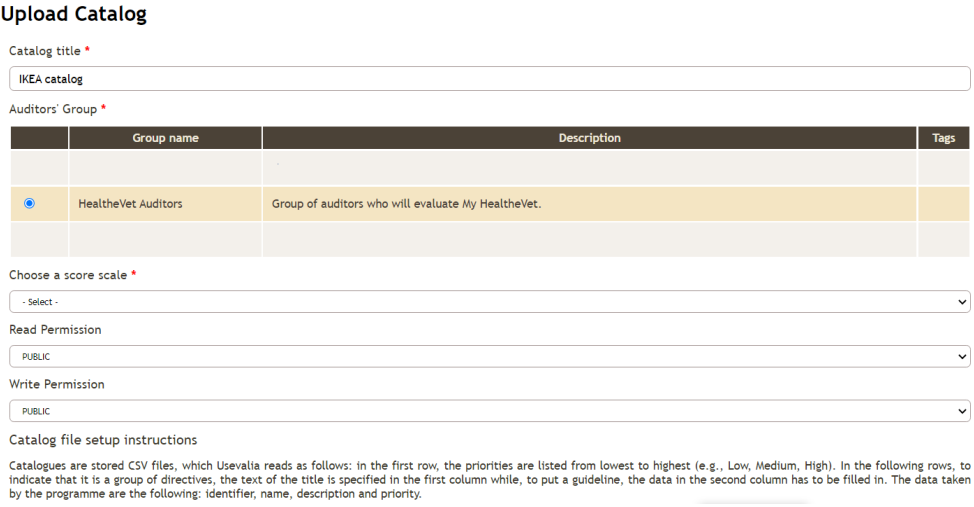


Figure 6. Import guideline catalog

Catalogs are stored as CSV files (see Figure 7), which Usevalia reads as follows: in the first row, the priorities are listed from lowest to highest (e.g., Low, Medium, High). In the following rows, the text of the title is specified in the first column in order to

indicate that it is a group of directives, while the data in the second column has to be filled in so as to insert a guideline. The data taken by the programme are the following: identifier, name, description, and priority. On the form, the title, the team to which it will belong, the scoring scheme, and the read/write permissions must be filled in before uploading the file containing the catalog. Chief auditors will be able only to delete the catalog from the application when it is not being used in any audits.

Low	Medium	High		
2	4	1		
Group 1				
	RU.1	Ease of learning (Dix)	The time required from unawareness of an application to its productive use should be minimal	High
	RU.2	Synthesizability (Dix)	The user should be able to evaluate the effect of previous operations on the current state.	High
	RU.3	Familiarity (Dix)	The knowledge required to use the system is related to the user's knowledge e.g. with the use	High
	RU.4	Concise and recognizable ma	If a help function is required it should be easy to locate and focused on the task the user is try	High
	RU.5	Communicate the purpose of	The purpose of the site is clear after a 5 second scan of the home page.	Medium
	RU.6	Prominent corporate name or	The company name or logo is prominently displayed in a reasonable size. The de facto standa	Low
	RU.7	Prominent value proposition/s	Include a slogan or value proposition of the website/company at the top of the page that clearl	Low
	RU.8	Emphasizing what makes the	In the value proposition emphasize what makes it unique for the user and different from the co	Low
Group 2				
	RU.9	Prominent call to action	The call to action is at the top of the page. You don't have to use the scroll bar to find it further	High
	RU.10	Easily findable company infor	It is easy to find the company information.	High
	RU.11	Easily reachable contact infor	It is easy to find the contact information. It is very common in the upper right corner.	Medium
	RU.12	Easily searchable newsletter	If there is a subscription form it is easy to find. It is common to be in the footer but it is not sta	High
	RU.13	Locatable social network icon	It is standard for social network icons to be in the footer.	High
	RU.14	Recognizable main navigation	It is easy to identify which is the main navigation. The de facto standard is in the header.	Low
	RU.15	Recognizable and consistent	Make links recognizable to the user at a glance. Indicate in some way whether a link has been	High
	RU.16	Easily searchable	The search engine within the web site is easy to access. A search icon or a search bar is usu	High
	RU.17	Navigation without invented w	Menus and all components that have to do with navigation should never have invented words.	Low
	RU.18	Navigation icons	Use icons for navigation categories only when they allow immediate recognition of a category	Medium
	RU.19	Ease of learning (Dix)	The time required from not knowing an application to its productive use should be minimal.	Low

Figure 7. Usability guideline catalog in CSV format

The chief auditor then defines the type of audit (standard, basic, task-based), after which the auditor or team of auditors begin to evaluate the previously uploaded guideline catalog. The types of audits are the following: (1) the standard audit consists of the evaluation of each usability guideline, one by one; (2) the basic audit evaluates an application on the basis only of the highest priority guidelines in the heuristic dataset; and finally, (3) the task-based audit performs the evaluation of each usability guideline on the basis of the different tasks that can be carried out in an application. A list of categories of applications is, therefore, provided in Usevalia (see Table 2), and each category incorporates a customizable set of typical tasks to be carried out. The categories of applications were the result of the harmonization of the proposals by Constantinides and Fountain (2008), Rossi et al. (2008), and Deshpande et al. (2002). These authors propose several taxonomies of web applications, each of which takes into account different characteristics of the website, such as its design, its development or the function it performs.

Table 2. Categories of applications predefined in a Usevalia task-based audit.

Category	Example	Usual, basic tasks
Search Engines/Portals	Google	register, log in, search with words, search with images, search in different languages, filter results, change search engine language, search prediction, search suggestions
Blogs/Informative	El País	register, log in, search with words, view latest news, share content, comment on content, view tags, rate content
Forums/Interaction	Stack Overflow	register, log in, search with words, view popular posts, view latest posts, post topic, quote topic/post, reply topic, view a user's profile, view a user's posts
Social networks	Facebook	register, log in, post content, share content, send private messages, view a user's profile, edit my profile, view tags, search for content
Transactional	BBVA	register, log in, make a transfer, view receipts, view balance, view my bank accounts, view my cards, apply for a loan, view account information
E-commerce	PcComponentes	register, login, word search, search by product category, search by product category, view product information, rate product, place order, cancel order, make complaint, add product to wishlist
E-mail	Outlook	register, log in, compose mail, filter mail, delete mail, create folders, reply to mail, block sender, word search
Entertainment	Netflix	register, login, search with words, search by category, view content/game, add to favourites, rate content/game, view latest news, view suggestions
Academic	Virtual classroom through a Learning Management System such as Moodle	log in, view announcements, view assignments, download material, contact a teacher, view exam marks, filter subjects, view subject planning/calendar, view subject planning/calendar
Collaborative	Google Docs	register, log in, search by words, view recent, create new document, filter documents, download document, add participants, communicate with participants
Downloads	Games Torrents	register, login, filter downloads, filter downloads, search by words, search by category, download content, share download, rate download, view description / summary download
Corporate/Public entities	University of Oxford	view company information, view rates, view services provided, view contact information, view links to social networks, change language of the page
Services	Google translator	register, login, use service, view help, view service information, view service information

Each auditor then performs the heuristic evaluation by checking the usability guidelines required, one by one, by specifying: (1) the score given to that guideline (according to the scoring scheme) in the context of the task or the whole application, depending on the type of audit; (2) (optional) any comments on the score, and (3) (optional) a set of suggestions to deal with the guideline. In this evaluation process, the auditor can take a screen grab of part of the relevant app or website and paste it onto the heuristic evaluation in order to justify the mark. It is usual for the application audited to evolve over successive deployments, and this mechanism of incorporating images can, therefore, help document the rationale behind each score (see Figure 8).

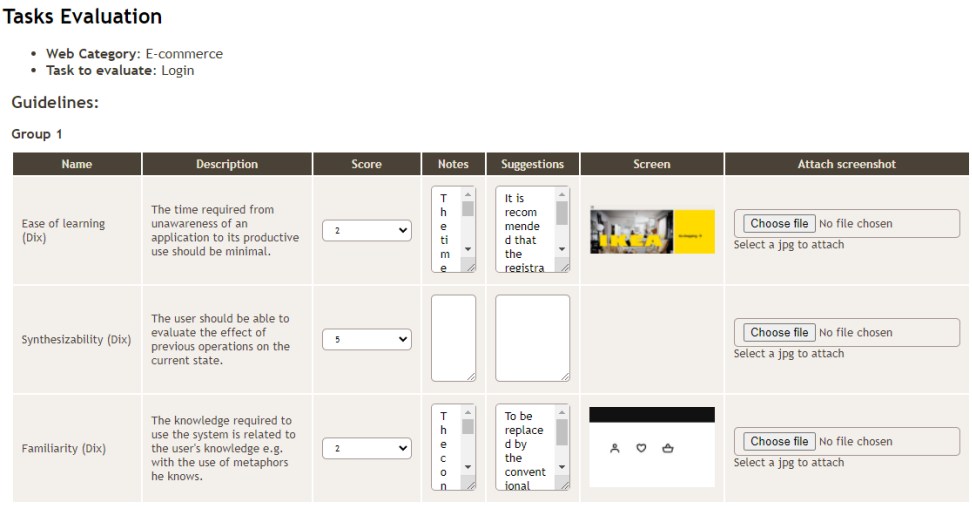


Figure 8. Upload screenshots to highlight problems encountered during the evaluation carried out using Usevalia

If there are x number of days left before the end of the audit (where "x" is a configurable parameter), the system sends a reminder message to the auditors who are performing the evaluation in order to make them aware that the deadline is close. Usevalia includes a field called "Evaluation status", which allows the chief auditor to see the evaluation status of the audit (pending or completed). When all the auditors have completed their assessments, an email is automatically sent to the chief auditor, who can explore the scores given to the guidelines by each auditor, together with their total averages. Once the audit time has expired, the audit is automatically closed and the chief auditor is informed.

Usevalia makes it possible to create, delete, evaluate, close, reopen, generate and view the statistics of an audit (see Figure 9). Some of the statistics displayed are the total percentage of passed and failed guidelines in the audit, the number of auditors who have indicated a guideline as passed/failed, or the number of passed/failed guidelines at each priority level. Once an audit has been selected in order to generate a report, the chief auditor will be presented with a link that will make it possible to download the

report. This file will be in PDF format and will contain all the information related to the audit selected.

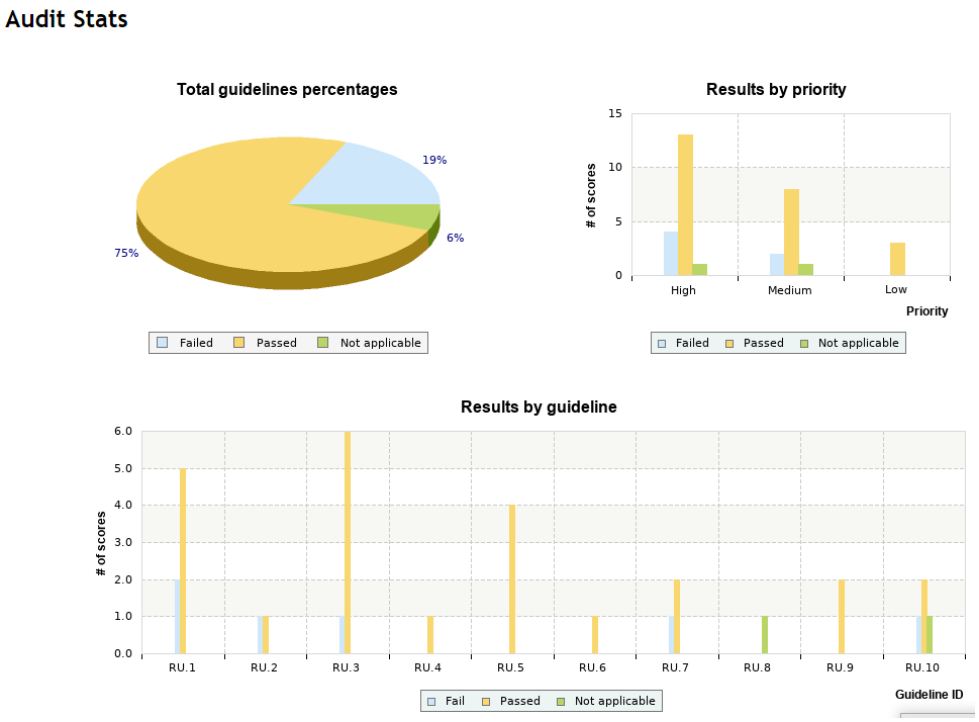


Figure 9. Statistics of a standard audit using Usevalia

Reports are intended to be the output of the evaluation, and they present the information and results of the audit formally, while graphs allow the most important data to be seen at a glance without having to read the whole document. Reports also provide the option of showing additional information, such as the number of guidelines approved by each auditor.

The tool is available to the HCI community at the following address: ***https://giis.inf.um.es:446/drupal2/usevalia***. It is in a state of continuous improvement, but the current version is stable and has been used on an HCI course for two years (see Section 4.3.1).

3.2. Architecture overview

Seven basic criteria have been considered when selecting the technology chosen for the development of Usevalia: (1) greater popularity in hosting and community portals; (2) the ease of its development and maintenance; (3) the degree of security; (4) the ease and quality of documentation; (5) the type of license; (6) the possibility of internationalization, and (7) flexibility in development. The most important criteria are those related to both the popularity of the technology and the security aspect.

The advancements in hardware and software technologies led to the evolution of the modern-day CMS (Content Management System) which allows anyone to design a website without the need for coding knowledge. CMS is defined as computer applications that allow the editing and publishing of content, along with maintenance from the user interface. Moreover, CMS has been developed as a modular system with the possibility of programming individual features as extensions (Blazek et al., 2019). Furthermore, open-source CMS such as Drupal have been the choice preferred for the development of several web applications. Many users employ Drupal for website development owing to multiple features, such as flexibility in content creation, easy administration of users, and the capability to handle complex workflows (Iqbal et al., 2020).

Analyzing the characteristics of an individual system can help choose the appropriate CMS for a specific web application. This is owing to the high variability of functions and features that make one particular CMS more suitable for developing an application than another (Vakode & Chaudhari, 2013). Drupal has multiple benefits, such as: (1) The user needs no technical knowledge other than connecting via FTP and installing the databases for installation; (2) the maintenance and updating of the site is easy, as it includes backup of the website and uses a web update interface to replace files. In addition, Drupal alerts the site administrator when an update is required; (3) the strength and contribution of the community, as it is very large in terms of its users and developers; (4) good usability, as administrators can easily access the site or section in visitor mode by simply clicking the edit button. Accessibility to the administration area requires some learning, as its default configuration does not have a refined look and feel. The main package does not include an editor like WYIWYG (What You See is What You Get), but one can be installed as a plug-in. In Drupal, the editing of pages or sections is the same, but when adding a new page, it may be necessary to link it by hand; (5) it is highly scalable with a high traffic handling capability. The user's websites are cached indefinitely as a default configuration, but can also be cached manually for a set amount of time; (6) security updates are published on drupal.org. Whenever a new update is released, users are informed via the update status plugin. The community is very active, and when security vulnerabilities are detected, they are remediated very quickly. In addition, references are available in order to guide the creation of a more secure site; and finally, (7) user roles and workflows are another of its main benefits, as it includes two sets of default roles (anonymous user and authenticated user). Beyond these, any number of user roles can be created and assigned different permissions depending on the type of content. Additional modules can be used to give more specific permissions to users on the basis of the content section and by using the taxonomy function.

Figure 10 shows a high-level architecture diagram of Usevalia. The diagram shows the main building blocks of the solution, namely: (1) a web server such as Apache; (2) MySQL, a relational database management system, and (3) the CMS, Drupal. All three tools are the result of major free and open-source software projects and are recognized as mature, popular and future-proof systems.

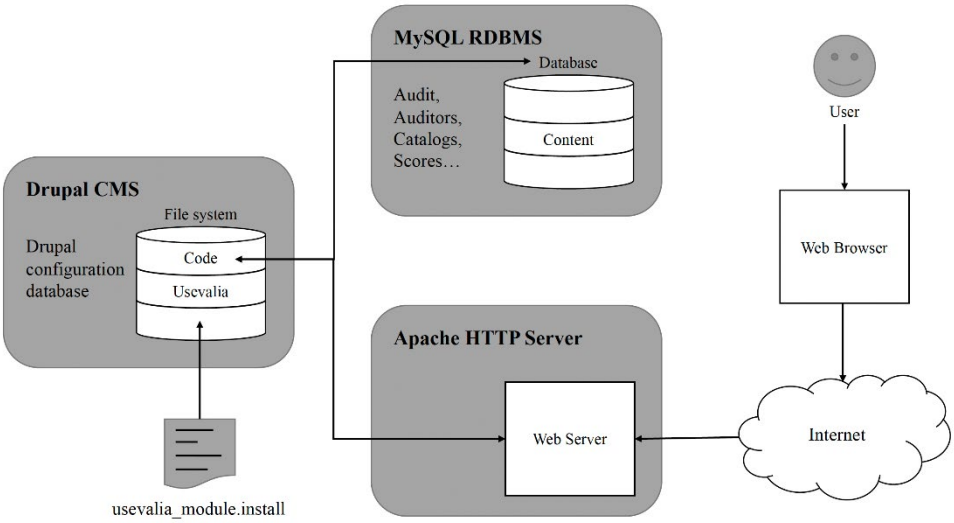


Figure 10. Architecture diagram of Usevalia using Drupal

Drupal provides functionality through the use of the so-called modules. A module in Drupal is a set of program files, images, data and style sheets organized into a directory structure that is stored in the server's file system. Usevalia was developed as a custom module. Table 3 provides a brief description of the purpose of the files included in the directory structure depicted in Figure 10.

Table 3. Directory structure of the Drupal automated support architecture module

Directory	Description
/	Root directory of the module, which contains the .info, .module and .install files
/Classes	Contains all the classes of the program model in .php files
/Classes/DAO	Contains the module's DAO handlers for program persistence in .php files
/Controller	Contains interface .php files that do not require user interaction
/EventSubscriber	Contains a service that restricts access to anonymous users
/Form	Contains the interface .php files that require user interaction
/Logo	Contains the program's logo and favicon
/Templates	Directory in which the two templates provided by the tool for catalogue creation are stored
/Service	Contains the main program controller
/Sql	Contains the file with the entity-relationship model tables
/Util	Directory containing the two libraries used for generating graphs and reports, along with the PHP classes that make use of these libraries
/Util/jpgraph	Source code of the jpgraph library that allows the creation of graphics
/Util/tcpdf	Folder with the source code from the tcpdf library, used to generate PDF files

In Drupal, the instructions required by a module in order to perform the actions for which it was conceived are defined in a .module file. Such a file is a script encoded in PHP. Together with the .module file, a .install file and a .info file are also written. The .install file is a PHP file that is run the first time the module is enabled to perform the setup procedures. This file specifically creates database tables and fields with which to represent the content types supported by Usevalia. The .info file, meanwhile, defines properties in key/value pairs in standard .ini file format. It stores metadata concerning

the module and is required in order for the Drupal system to recognise the existence of the module.

4. METHOD

This section presents the method followed in the empirical validation of Usevalia (see Figure 11). Both the overall objective and the research question of the study are stated in Section 4.1, while the hypotheses are formulated in Section 4.2. In Section 4.3, a process is described to validate the Usevalia tool by evaluating four webs and mobile apps in the e-commerce domain. The instruments and data analysis methods are described in Section 4.4.

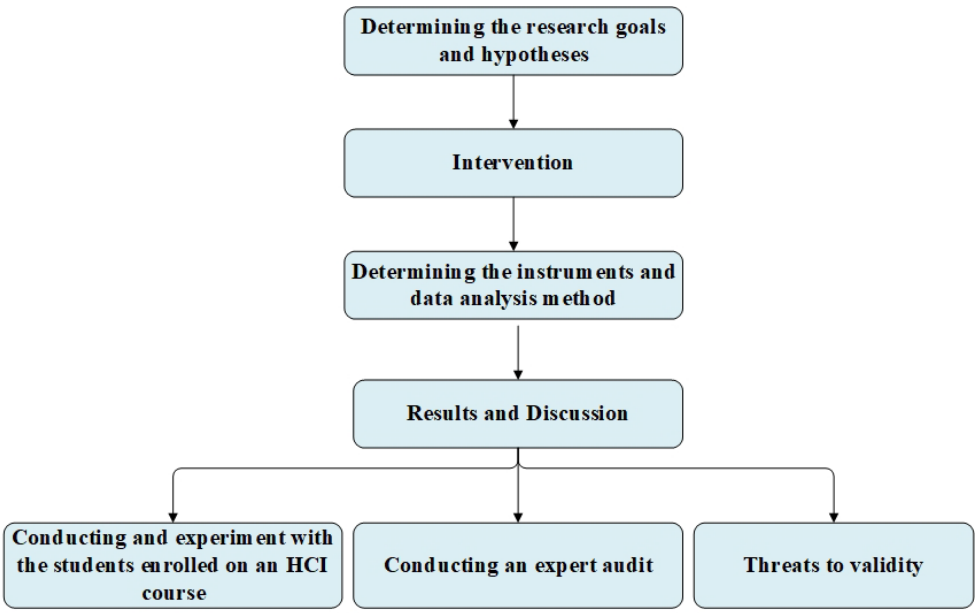


Figure 11. Workflow followed in the Usevalia empirical validation

4.1. Research Goals

The objective of our empirical evaluation is presented using **Basili and Rombach's (1988)** recommendations regarding the application of the Goal/Question/Metric (GQM) method. Our goal is thus defined as follows:

- Analyze the *performance of Usevalia, a usability audit tool based on heuristic evaluation*
- for the purposes of *assessing the intention to adopt the tool*
- in terms of *the adequacy of the usability audit tool*
- from the point of view of the *usability auditor*
- in the context of *the usability inspection based on heuristic evaluation*

The above was then considered as the basis on which to formulate the following research question:

- RQ. How does the Usevalia tool for heuristic inspection usability audit perform in practice?

4.2. Hypotheses

In order to answer the research question, we defined the following hypotheses:

- H1.1: Usevalia performs sufficiently well for it to be used in a real project.
- H1.2: Usevalia is better than a spreadsheet application.
- H1.3: Usevalia is useful.
- H1.4: Usevalia is easy to use and learn.
- H1.5: The users are willing to continue using Usevalia once they have tried it.
- H1.6: The users are willing to recommend Usevalia to others.

4.3. Intervention

The validation of the Usevalia tool was divided into two parts: (1) students enrolled on an HCI (Human-Computer-Interaction) course of the BSc. in Computer Engineering at the University of Murcia (Spain) in the academic year 2020/2021, followed by (2) a validation with five experts in usability audits. The abovementioned subject introduces students to HCI issues, including the development and auditing of usable and accessible user interfaces, paying special attention to the application of standards and style guides in web, mobile and desktop computer applications. The usability and accessibility of both web and mobile applications are the two key aspects of the course. The experts are the authors of this paper and were selected because they have extensive experience in conducting usability audits and previous knowledge of the tools evaluated. The method selected in order to collect data was a questionnaire. Questionnaires are popular means by which to acquire information on knowledge and perception (Bird, 2009).

The activity with students was performed between November 11th, 2020, and February 15th, 2021. A total of 22 students participated in the evaluation of Usevalia. Questionnaire 1 (see Table 4) was administrated to students in order to obtain their perception of Usevalia. Questions were answered from January 26th, 2021, to February 9th, 2021. The aim of this survey was to attain feedback concerning the students' perceptions of their experience with Usevalia after an in-class training session given by one of the experts. In this session, a theoretical and practical introduction to Usevalia was given and a hands-on experience with the features and capabilities of the tool was acquired by the students. A one-group post-test only design was used in which a treatment is implemented (in-class training session) and then a dependent variable (students' perceptions based on TAM) is measured.

Five experts responded to Questionnaire 2 (see Table 5) in order to validate Usevalia against a similar purpose tool (i.e., Usability Datalogger) using a task-based heuristic evaluation. Questionnaire 2 was completed from January 10th to January 24th, 2022.

Specifically, four usability audits of two web applications (IKEA[2] and ASOS[3]) and two mobile applications (Mercadona[4] and El Corte Inglés[5] apps) were conducted. The four applications evaluated are considered to be mature, well-known representatives of the e-commerce domain. The four applications were audited using Usevalia and the Usability Datalogger tool described in Section 2. The latter is the only alternative to Usevalia that is still available, given that Sirius, Prometheus, and SUIT are not currently available to the scientific community. Nevertheless, Usability Datalogger can serve to compare a kind of "traditional" or "manual" audit tool based on spreadsheets with the features of Usevalia, the audit management tool presented in this paper. Finally, the experts filled in the questionnaire in order to compare the tools evaluated: Usevalia and Usability Datalogger.

The evaluation method followed in this study is expert review (Tory & Moller, 2008), which can be used to evaluate software tools (Beecham et al., 2012). The analysis of the usability audit tools was carried out in the context of the heuristic evaluation of two websites and two mobile apps. In this type of evaluation, a small number of experts evaluate the interface separately (Nielsen, 1994). The heuristic evaluation process carried out by the reviewers provided them with useful information to assess the usability audit tools under study. Specifically, the reviewers completed a TAM-based questionnaire to evaluate Usability Datalogger and Usevalia.

There is no consensus on the optimal number of reviewers to perform usability evaluations (Hwang & Salvendy, 2010). According to different studies, the number of expert reviewers tends to be small, and range from two to over twenty reviewers (Olson, 2010). Nielsen and Molich (1990) reported that five evaluators found about 2/3 of usability problems using heuristic evaluation. In our specific case, however, it is worth noting that: (1) there are no very complex interfaces involved in the evaluation, and (2) all reviewers participating in the evaluation can be considered to have the same probability of encountering all usability problems (homogeneity assumption) (Hwang & Salvendy, 2010). Moreover, 3-hour evaluations of five reviewers are likely to find over 99% of the usability problems (Dumas & Sorce, 1995). For these reasons, five experts were chosen to perform the expert evaluation procedure in our study.

Both questionnaires employed a five-point Likert-type scale (1 = Completely disagree; 2 = Disagree; 3 = Neither agree nor disagree; 4 = Agree; 5 = Completely agree). The scale used was the same, since both questionnaires were designed based on the TAM model. For each block of questions, the mean of the responses to the questions in that block was calculated and used as the representative value given by the student to that block. In Questionnaire 1, the medians obtained for each question and for each block were compared with the hypothesized central value proposed by the researchers

2 Accessible from https://www.ikea.com/es/es/
3 Accessible from https://www.asos.com/es/
4 Available at https://play.google.com/store/apps/details?id=es.mercadona.tienda
5 Available at https://play.google.com/store/apps/details?id=com.elcorteingles.app

(MD = 3.00). In Questionnaire 2, the averages obtained for the questions related to each variable (perceived usefulness and perceived ease of use) were considered for comparative purposes.

4.4. Instruments and data analysis methods

The two questionnaires mentioned above were designed according to the Technology Acceptance Model (TAM) (Sauro & Lewis, 2016), which helps put the results into the perspective of technology adoption. The TAM highlights the need to be conscious of the socially constructed processes in which tools are deployed and used on a daily basis. This is because, when users are presented with new technology, several factors—and particularly the perceived usefulness and the perceived ease of use—influence their decisions regarding how and when they will use that technology.

The questionnaire filled in by the students (Questionnaire 1) contained 19 questions and was divided into four blocks of questions, namely: Block 1: Perceived utility; Block 2: Perceived ease of use; Block 3: Attitude of use, and Block 4: Intention to use. Block 1 was composed of questions, Q1-Q12, whose purpose was to answer H1.1, H1.2, and H1.3; Block 2 was composed of questions Q13-Q16 (H1.4); Block 3 was composed of questions Q17-Q18 (H1.5), and finally, Block 4 was composed of question Q19 (H1.6) (see Table 4).

With the data gathered with Questionnaire 1, a statistical test was run to determine whether the median of the answers was significantly different from 3 (neither agree nor disagree). The distribution of the mean values was not normal, except for Q18, which did have a normal distribution ($p = 0.066$) since the significance level is higher than $\alpha = 0.05$. The one-sample Wilcoxon signed-rank test was used for the questions that did not have a normal distribution, while the one-sample t-test was used for Q18, which did have a normal distribution. Notice that in our experiment, a difference has been considered significant if the p-value of the test is below 0.05 (i.e., $\alpha < 0.05$).

The expert's questionnaire (Questionnaire 2) contained 12 questions and was divided into two blocks of questions, namely Block 1: Perceived Usefulness (Q1-Q6), and Block 2: Perceived Ease of Use (Q7-Q12). It was divided into these two blocks because the two most important variables as regards predicting the development of a technological tool through the use of the TAM model are perceived usefulness and perceived ease of use. These two variables directly influence the users' intention to use a particular tool (Davis, 1989). The questions included in Block 1 corresponded to questions Q1, Q3, and Q6-Q11 in Questionnaire 1, while the questions included in Block 2 mapped onto questions Q5, Q12-Q16 in Questionnaire 1 (see Table 5).

In this work, inter-rater agreement was measured using Kendall's coefficient of agreement (W) (M. Kendall & Smith, 1939; M. Kendall & Gibbons, 1990). With this statistic it is possible to formally determine whether any consensus has been reached between raters and the relative strength of the consensus (Schmidt, 1997). The level of agreement finally obtained from the ratings issued by the five raters ($W = 0.75$; $p < 0.001$) can be considered strong (Moslem et al., 2019; Schmidt, 1997).

Table 4. Means, standard deviations, and Medians of students' perceptions (Questionnaire 1). "M": Mean;"SD": Standard deviations; "MD": Median.

Id	Associated hypotheses	Question	M	SD	MD
Block 1: Perceived utility					
Q1	H1.3	Using Usevalia would improve my performance on the HCI course.	3.40	1.14	4
Q2	H1.3	Usevalia would favor the teaching-learning process of the HCI subject.	3.54	0.80	3
Q3	H1.3	I consider that Usevalia represents a step forward in conducting usability audits.	3.72	0.93	4
Q4	H1.3	I would conduct future usability audits using Usevalia.	3.45	0.96	3.5
Q5	H1.1	The tool manages users and user roles sufficiently well for it to be used in a real project.	3.18	0.85	3
Q6	H1.3	The web categories included seem to be sufficient to represent all the websites/applications.	3.50	0.85	4
Q7	H1.3	The different types of predefined tasks associated with the web categories seem sufficient to represent all the websites/applications.	3.68	0.83	4
Q8	H1.3	The different types of evaluation (Basic, Standard, Task Based) in the tool seem sufficient to evaluate an audit.	3.86	0.83	4
Q9	H1.3	Usevalia streamlines the audit evaluation process.	3.64	0.95	4
Q10	H1.3	The reports generated by the tool help me understand the end result of an audit.	4.13	0.83	4
Q11	H1.3	I find Usevalia's graphs useful, as they provide interesting data on the audit.	3.81	1.09	4
Q12	H1.2	Usevalia is better than Microsoft Excel for performing usability audits, specifically audits with a heuristic evaluation.	3.59	1.00	4
Block 1	H1.1, H1.2 and H1.3	All the questions in Block 1	3.62	0.58	3.75
Block 2: Perceived ease of use					
Q13	H1.4	Learning to use Usevalia would be easy for me.	3.68	0.99	4
Q14	H1.4	I consider that my interaction with Usevalia would be clear and understandable.	3.77	1.11	4
Q15	H1.4	The steps to follow in the tool before creating an audit are intuitive.	3.54	1.05	4
Q16	H1.4	In general, I have had no problems using and understanding the operation of Usevalia.	3.45	1.18	4
Block 2	H1.4	All the questions in Block 2	3.61	0.98	3.75
Block 3: Attitude of use					
Q17	H1.5	I would use Usevalia if I needed to perform usability audits.	3.54	1.10	3.5
Q18	H1.5	I would use Usevalia to improve my performance on the HCI course.	3.36	1.04	3.5
Block 3	H1.5	All the questions in Block 3	3.45	0.96	3.50
Block 4: Intention to use					
Q19	H1.6	I would recommend Usevalia for usability audits.	3.54	0.96	3.50

Table 5. Expert evaluations of the Usevalia and Usability Datalogger tools (Questionnaire 2).

Id	Expert Questions	Usability Datalogger Tool					Usevalia Tool					Trace to Student Questions
Perceived usefulness		**E1**	**E2**	**E3**	**E4**	**E5**	**E1**	**E2**	**E3**	**E4**	**E5**	
Q1	Using this product enables me to accomplish tasks more quickly.	2	2	2	2	1	5	4	4	4	5	Q9
Q2	Using this product improves my current performance.	1	2	1	1	3	4	4	4	4	4	Q1
Q3	Using this product increases my productivity.	3	2	2	2	3	5	4	4	4	5	Q3
Q4	Using this product makes me more effective.	2	3	3	3	2	4	4	5	5	4	Q3
Q5	Using this product makes it easier to do my work.	4	3	2	3	3	5	4	4	4	5	Q6, Q7, Q8, Q10
Q6	I find this product useful.	3	3	3	4	3	5	5	5	5	5	Q11
Mean perceived usefulness of use		**2.43**					**4.43**					
Perceived ease of use												
Q7	Learning to operate this product was easy for me.	4	4	4	5	1	5	4	3	4	5	Q13
Q8	I found it easy to get this product to do what I wanted it to do.	2	2	2	3	2	4	4	4	4	5	Q12
Q9	My interaction with this product was clear and understandable.	4	3	4	4	2	5	4	4	5	4	Q14
Q10	I found this product to be flexible to interact with.	2	2	2	1	2	4	4	4	3	4	Q5
Q11	It was easy for me to become skillful at using the system.	4	4	4	3	3	4	4	4	3	5	Q16
Q12	I found the system easy to use.	4	3	4	4	3	5	4	4	4	4	Q15
Mean perceived ease of use		**3.03**					**4.13**					

5. RESULTS AND DISCUSSION

In this section, we validate Usevalia, a tool for managing inspection-based usability audits (see Figure 12). Firstly, we show and discuss the results of the experiment conducted with students enrolled in an HCI course (Section 5.1). Secondly, we report

the findings of the task-based usability audit performed by five experts (Section 5.2). Finally, we present the threats to validity (Section 5.3), along with attempts carried out to mitigate them.

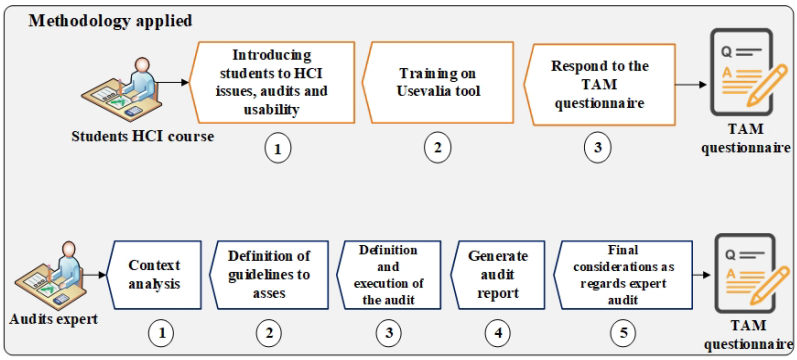


Figure 12. Methodology followed for the Usevalia validation

5.1. Conducting an experiment with students enrolled on an HCI course

The questions in Questionnaire 1 have been mapped onto the hypotheses presented in Section 4.2, and are used to structure this section. Table 4 shows the means, standard deviations, and medians of the students' answers to each question and for each block. Table 6 shows the median values of the responses to the student questionnaire per question and per block, together with the p-values returned by the Wilcoxon signed-rank tests and the one-sample t-test. Figure 13 presents the results of the student questionnaire by means of a stacked bar chart.

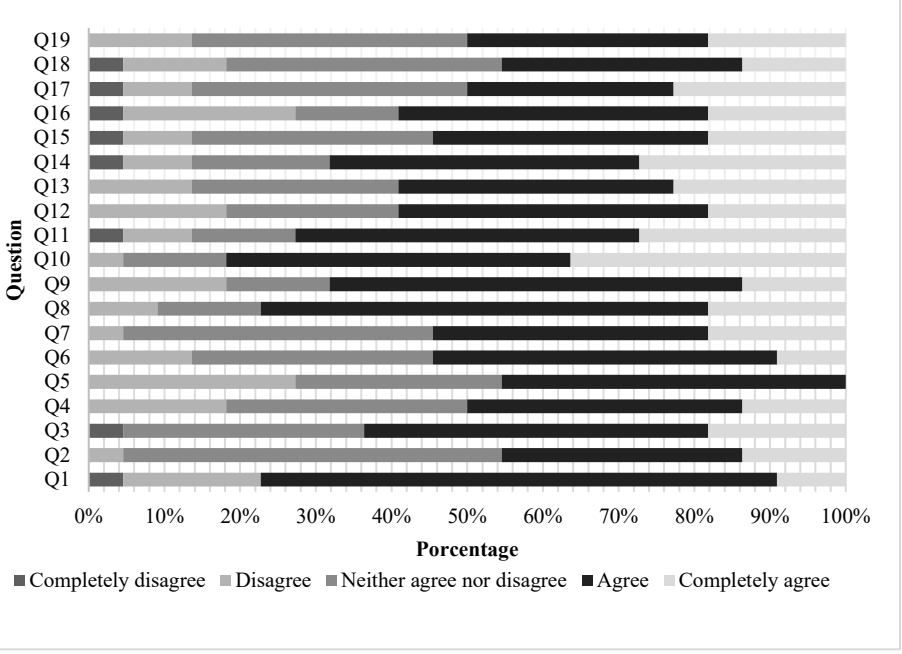


Figure 13. Summary of the participants' answers to each question

5.1.1. H1.1: Usevalia performs sufficiently well for it to be used in a real project

A Wilcoxon signed-rank test showed that there are no statistically significant differences ($Z = 85.000$, $p = 0.317$) to be able to claim that Usevalia's user management works sufficiently well to be used in a real project. Indeed, the median score rating was 3.0 (Q5). The problem with Q5 could be related to an inadequate assignment of roles and responsibilities by the Usevalia administrator during the implementation process. Gavrila and Barkley (1998) argue that role assignment helps administrators to maintain knowledge of the hierarchies that describe an organization, and hence the users working with the information systems. Athauda et al. (2008) state that one of the most common errors identified in user management in information systems is the granting of excessive permissions to users or groups, since access privilege requirements are often not specified for IT administrators. User and role management modules allow routine activities to be delegated to certain users with well-defined permission levels and limit access to systems to authorized users by applying a Role-Based Access Control (RBAC) approach. Usevalia was developed with the Drupal CMS, which includes a module with which to manage users and roles, and we are, therefore, of the opinion that the problem identified could be solved by reconfiguring the module in such a way that the roles and responsibilities among users are correctly defined.

5.1.2. H1.2: Usevalia is better than a spreadsheet application

The students consider that the Usevalia tool provides better functionality than the Microsoft Excel spreadsheet processing application when conducting usability audits (H1.2). Several studies have shown that most internal auditors perform their activities using MS Office Excel, and there are, therefore, indications that it is the most popular tool with which to perform data analysis. For example, Tang et al. (2017) concluded that most internal auditors use spreadsheet software (Excel), while another piece of research conducted by Bănărescu (2015) confirms that most survey participants use spreadsheet tools such as Microsoft Excel (65%) or database tools such as MS Access or MS SQL Server (43%). The results obtained by Ježovita et al. (2018) are aligned with the previous ones: 42% of participants use MS Excel, 36% of them use ACL, 9% use MS Access, and only 3% use the specialized internal auditing software IDEA[6]. Despite this, there are limitations associated with the use of Microsoft Excel as a data analysis tool in the context of auditing, which make it possible to argue that Usevalia offers better functionality when conducting usability audits. The main limitations associated with the use of Microsoft Excel are: (1) spreadsheet applications are often not designed for or capable of handling large data sets; (2) spreadsheets are prone to errors, especially when the auditor needs to clean large data sets that were acquired from multiple sources; (3) errors in formulas can negatively impact on the logic of the analysis derived from them;

6 IDEA® Data Analysis Software is a comprehensive, powerful, and easy-to-use data analysis solution designed by audit experts.

(4) the use of macros and multiple pivot tables may make audits and analysis time-consuming and complex; and finally, (5) extensive knowledge of programming languages is often required in order to, e.g. make use of routine or continuous data analysis (Ježovita et al., 2018; Smidt et al., 2019; Soileau et al., 2015).

Table 6. Median values of the answers to the survey together with the P values returned by the Wilcoxon Signed Rank test.

Id	MD	Sig.
Statistical analysis per question		
Block 1: Hypothesis 1.1		
Q5	3	(p=0.317)
Block 1: Hypothesis 1.2		
Q12	4	(p=0.015)
Block 1: Hypothesis 1.3		
Q1	4	(p=0.117)
Q2	3	(p=0.008)
Q3	4	(p=0.006)
Q4	3.5	(p=0.040)
Q6	4	(p=0.016)
Q7	4	(p=0.003)
Q8	4	(p=0.001)
Q9	4	(p=0.008)
Q10	4	(p=0.000)
Q11	4	(p=0.006)
Block 2: Hypothesis 1.4		
Q13	4	(p=0.007)
Q14	4	(p=0.008)
Q15	4	(P=0.034)
Q16	4	(p=0.086)
Block 3: Hypothesis 1.5		
Q17	3.5	(p=0.038)
Q18	3	(p=0.119)
Block 4: Hypothesis 1.6		
Q19	3.5	(p=0.018)
Statistical analysis per block		
Block 1	3.75	(p=0.000)
Block 2	3.75	(p=0.011)
Block 3	3.5	(p=0.041)
Block 4	3.5	(p=0.018)

5.1.3. H1.3: Usevalia is useful

According to the Wilcoxon signed-rank test carried out for the participants' answers, there is no statistical evidence to support the claim that the tool improves students' performance in the HCI subject ($Z = 171.000$, $p = 0.117$). The median score rating was 4.0 (Q1). However, the students had the perception that Usevalia would favor the teaching-learning process of the HCI (Q2) subject and represents a step forward in conducting usability audits (Q3). They also responded that they would conduct future usability audits with Usevalia (Q4). With regard to other more specific functionalities of the tool, the web categories included appear to be sufficient to represent all the websites/applications (Q6), while the different types of predefined tasks associated with the web categories would appear to be adequate to represent all the websites/applications (Q7). The types of

evaluation (Basic, Standard, Task-Based) in the tool cover an auditor's needs when evaluating an audit (Q8). This may justify the fact that the participants agreed that Usevalia streamlines the audit evaluation process (Q9), and the reports generated by the tool represent an aid for auditors as regards understanding the outcome of a particular audit (Q10). Finally, the participants consider Usevalia's graphs useful, as they provide interesting data on the audit (Q11).

5.1.4. H1.4: Usevalia is easy to use and learn

According to the results obtained, six participants (27.7%) indicated that they had problems when using and understanding the operation of Usevalia in general (Q16), and the null hypothesis related to Q16 cannot, therefore, be rejected ($p = 0.086$); this result is, however, remarkable given that the p-value is close to 0.05. We believe that the problem identified in this question may be owing to the students' insufficient previous interaction with the tool before answering the questionnaire, although the instructor had previously carried out an activity in which he explained, step by step, how the tool worked in order to ensure that the students would know the functionalities that Usevalia had and that this aspect would not, therefore, have a subsequent negative influence on the results of the questionnaire. However, according to Zanna and Rempel (2008), pre-adoption beliefs held by potential users are based principally on indirect experience with technologies and are, therefore, susceptible to change. Post-adoption beliefs held by experienced users are based mainly on experience. Empirical evidence suggests that beliefs and attitudes formed by direct experience are more enduring and better predictors of behavior than beliefs/attitudes formed by indirect experience (Fazio et al., 1982; Fazio & Zanna, 1981). Direct experience with a technology allows users to assess its effectiveness as regards meeting their needs more easily and confidently. In addition, attitudes and beliefs based on direct experience are more accessible in memory, resulting in stronger links between beliefs/attitudes and behavior. Thus, in the context of this research, we believe that if the students had interacted more with the tool, they would not have responded that they had problems using and understanding it in general. Despite the contradictory empirical results, the theory provides a consistent description of how the experience of using a given system moderates the relationship between perceived ease of use and users' intention to use it. We believe that the more users gain direct experience with IT, the easier they will find the tool to use. It is worth noting that the majority of respondents considered that Usevalia was easy to use and learn.

5.1.5. H1.5: The users are willing to continue using Usevalia once they have tried it

A Wilcoxon signed-rank test showed that the median of the respondents' answers as regards whether they would use Usevalia to conduct usability audits (MD=3.50) differed significantly from the hypothesized central value proposed by the researcher (MD=3.00). Indeed, the statistical results obtained were $Z= 84.500$, $p= 0.038$, signifying that the null hypothesis could be rejected (Q17).

The one-sample t-test run to determine the mean of the respondents' answers as regards whether they would use Usevalia to improve their performance on the HCI course showed that there were significant differences concerning the hypothetical value proposed by the researcher, defined as 3.0. The students' scores were normally distributed, as assessed by Shapiro-Wilk's test ($p > 0.05$). The mean student scores ($M = 3.363$, $SD = 1.048$) were higher than the normal score of 3.0, and there was a non-statistically significant mean difference of 0.363, a confidence interval of 95%, and $t(22) = 1.627$, $p = 0.119$. We cannot, therefore, reject the null hypothesis (Q18). Based on the above results, we have insufficient evidence to conclude that the students that participated in the study would continue using the Usevalia tool if they needed to.

5.1.6. H1.6: The users are willing to recommend Usevalia to others

A Wilcoxon signed-rank test showed that the median of the respondents' answers as regards whether they would recommend Usevalia for usability audits (MD=3.50) is significantly different from the hypothesized central value proposed by the researcher (MD=3.00). In this case, the results obtained were statistically significant, with a statistic $Z= 88.500$ and $p= 0.018$, thus allowing the null hypothesis to be rejected. This makes it possible to conclude that students would recommend Usevalia for usability audits (Q19).

We cannot be sure that the research participants are representative of the software development industry, but according to Carver et al. (2003), the results obtained through empirical studies with students constitute progress in software engineering. In this case, the use of students is considered feasible, as no industrial experience is required to perform the tasks. Furthermore, it has not yet been proven that there are real differences between students and industry professionals.

5.2. Conducting an expert audit

This section presents the evaluation of the Usevalia and Usability Datalogger tools carried out by an expert in usability audits (see Table 5). As stated in Section 3, Usevalia does not prescribe any process, and in this case, the expert performed the validation by following an audit process based on the USBAM audit method (USaBility Audit Method) resulting from other requirements engineering audit methods (Cruz Zapata et al., 2018). The expert subsequently answered a questionnaire based on the TAM model in order to compare both tools and discover what benefits and deficiencies one has when compared to the other.

The expert validation was carried out using the following process: (1) Context analysis: the expert examines the context of the applications to be audited: purpose, scope, user profiles, etc.; (2) Definition of tasks and guidelines to assess: the expert defines the set of tasks or functionalities of the applications to be audited and creates or reuses a catalogue of usability guidelines that will assess these tasks in all the audits; (3) Definition and execution of the audit: the expert defines an audit for each app and evaluates them using the guidelines established above; (4) Generate audit report: the

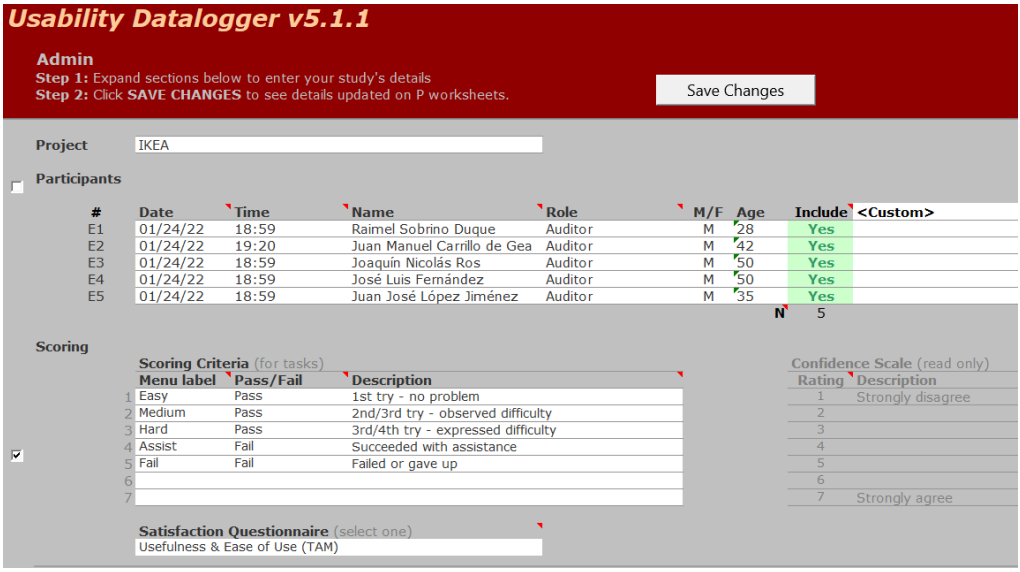
usability audit expert exports a report in PDF format with the results of the task evaluations.

5.2.1. Context analysis

In the first step of the process, the expert analyzed the context of the apps selected for audit. He then examined the official website of the two web apps and downloaded the two mobile apps via Google Play. This was done in order to extract the features or functionalities that would be evaluated later.

In this step, the expert completed the input fields in the two tools used to perform the audit. In the case of Usability Datalogger (see Figure 14), the details required for the audit were entered on the Admin worksheet before conducting a study. These details include the name of the project, names of the participants, and the dates and times of the sessions. Moreover, the date, time of the session, name, role, gender, and age of each audit participant were entered in the participants' section of the Admin worksheet. The scoring scheme was then defined in order to evaluate the tasks. Users can choose to define less than 7 scoring tags, depending on their needs. If they choose to define less than 7 scoring tags, they should leave the undefined cells empty in order to avoid cluttering the participants' worksheets and charts with unnecessary elements. With Usevalia users enter the data associated with the application to be evaluated, such as the name of the application, URL, web category to which it belongs, and a short description of the tool (see Figure 15). The group of auditors and the scoring scheme to be used in the audit are then introduced. In this case, the expert reused a scoring scheme that had already been defined in another audit. This feature was considered to have a medium priority for audit purposes (see Table 1).

According to the evaluation carried out when analyzing audit context, Usability Datalogger is not very flexible as regards the customized scoring system (score of 2.43 for perceived usefulness in Table 5), because an auditor has only the default scoring scheme available for redefinition. However, Usevalia gives the auditor the option of defining several scoring schemes and using them simultaneously in different audits (score of 4.43 for perceived usefulness in Table 5). Usability Datalogger provides support for multiple auditors, but the audit cannot be performed simultaneously, whereas with Usevalia, it can. This is because Usability Datalogger is based on an Excel sheet, the assessment must be completed by one auditor, and once completed, the next auditor continues until the assessment is finished. In general, Usability Datalogger allows the definition of several user roles associated with the auditors, but there is no support in the case of managing groups of auditors, as occurs with the Usevalia tool. The expert assigned high priority to the role management feature (see Table 1).



Usability Datalogger v5.1.1

Admin

Step 1: Expand sections below to enter your study's details
Step 2: Click **SAVE CHANGES** to see details updated on P worksheets.

Save Changes

Project IKEA

Participants

#	Date	Time	Name	Role	M/F	Age	Include	<Custom>
E1	01/24/22	18:59	Raimel Sobrino Duque	Auditor	M	28	Yes	
E2	01/24/22	19:20	Juan Manuel Carrillo de Gea	Auditor	M	42	Yes	
E3	01/24/22	18:59	Joaquín Nicolás Ros	Auditor	M	50	Yes	
E4	01/24/22	18:59	José Luis Fernández	Auditor	M	50	Yes	
E5	01/24/22	18:59	Juan José López Jiménez	Auditor	M	35	Yes	
						N	5	

Scoring

Scoring Criteria (for tasks)

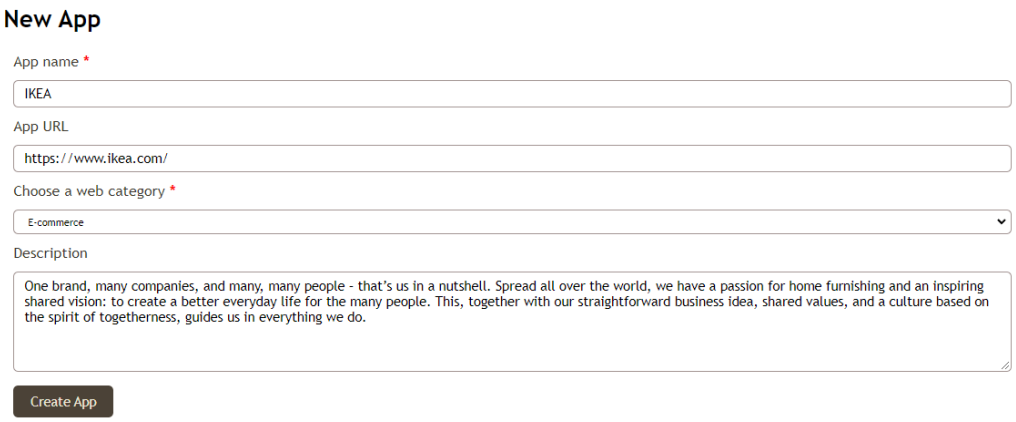
	Menu label	Pass/Fail	Description
1	Easy	Pass	1st try - no problem
2	Medium	Pass	2nd/3rd try - observed difficulty
3	Hard	Pass	3rd/4th try - expressed difficulty
4	Assist	Fail	Succeeded with assistance
5	Fail	Fail	Failed or gave up
6			
7			

Confidence Scale (read only)

Rating	Description
1	Strongly disagree
2	
3	
4	
5	
6	
7	Strongly agree

Satisfaction Questionnaire (select one)
Usefulness & Ease of Use (TAM)

Figure 14. Admin worksheet containing IKEA application audit data



New App

App name *
IKEA

App URL
https://www.ikea.com/

Choose a web category *
E-commerce

Description
One brand, many companies, and many, many people - that's us in a nutshell. Spread all over the world, we have a passion for home furnishing and an inspiring shared vision: to create a better everyday life for the many people. This, together with our straightforward business idea, shared values, and a culture based on the spirit of togetherness, guides us in everything we do.

Create App

Figure 15. Usevalia graphical interface containing IKEA application data

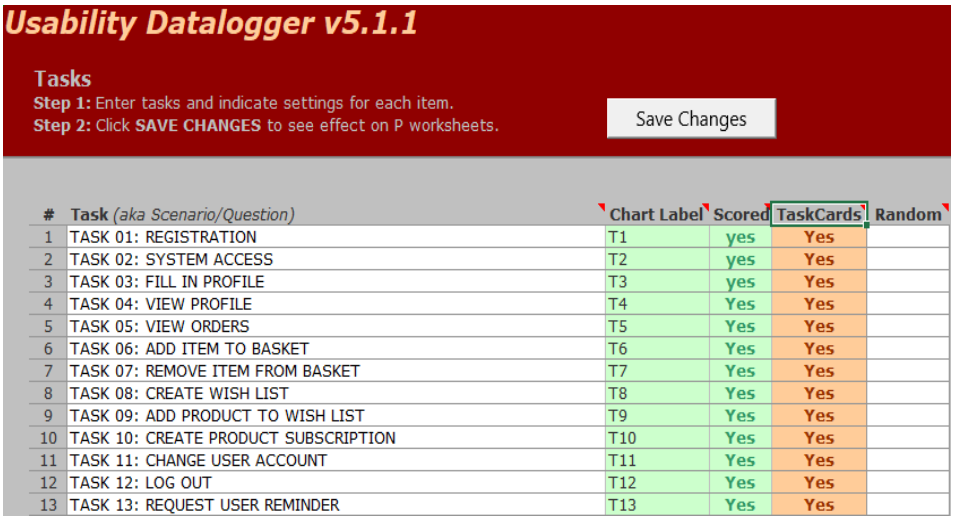
5.2.2. Definition of tasks and guidelines to assess

The expert employed the above analysis as the basis on which to define the following set of tasks to be assessed in each of the applications, together with defining a guideline catalog encompassing Jacob Nielsen's 10 usability principles. The tasks defined are typical of e-commerce applications (Dzulfikar et al., 2018), and are the following (see Figure 16):

- *Task 01: Registration*
- *Task 02: System access*
- *Task 03: Fill in the profile*

- *Task 04: View profile*
- *Task 05: View orders*
- *Task 06: Add item to basket*
- *Task 07: Remove item from basket*
- *Task 08: Create a wish list*
- *Task 09: Add product to wish list*
- *Task 10: Create product subscription*
- *Task 11: Change user account*
- *Task 12: Log out*
- *Task 13: Request user reminder*

At this stage, the auditor can create a catalog from scratch or reuse a catalog previously defined to evaluate the applications. In this evaluation, a new catalog was prepared, which was composed of Jacob Nielsen's 10 heuristic principles. The expert considers that Usevalia makes work easier (score of 4.13 for perceived ease of use in Table 5) than the Usability Datalogger tool (score of 3.03 for perceived ease of use in Table 5), because the former gives the auditor the option to reuse one of the catalogs defined in other audits or to import one of his/her catalogs. The expert assigned a medium priority to having a template for importing catalogs to be reused in an audit (see Table 1). Finally, it is possible to define several user roles associated with the auditors in Usability Datalogger, but there is no support that can be used to manage groups of auditors, as occurs with Usevalia.



#	Task *(aka Scenario/Question)*	Chart Label	Scored	TaskCards	Random
1	TASK 01: REGISTRATION	T1	yes	Yes	
2	TASK 02: SYSTEM ACCESS	T2	yes	Yes	
3	TASK 03: FILL IN PROFILE	T3	yes	Yes	
4	TASK 04: VIEW PROFILE	T4	Yes	Yes	
5	TASK 05: VIEW ORDERS	T5	Yes	Yes	
6	TASK 06: ADD ITEM TO BASKET	T6	Yes	Yes	
7	TASK 07: REMOVE ITEM FROM BASKET	T7	Yes	Yes	
8	TASK 08: CREATE WISH LIST	T8	Yes	Yes	
9	TASK 09: ADD PRODUCT TO WISH LIST	T9	Yes	Yes	
10	TASK 10: CREATE PRODUCT SUBSCRIPTION	T10	Yes	Yes	
11	TASK 11: CHANGE USER ACCOUNT	T11	Yes	Yes	
12	TASK 12: LOG OUT	T12	Yes	Yes	
13	TASK 13: REQUEST USER REMINDER	T13	Yes	Yes	

Figure 16. Task sheet showing the tasks assessed in the IKEA application.

5.2.3. Definition and execution of the audit

In this step, the expert defined an audit in Usevalia, as shown in (see Figure 17). The auditor selected Jacob Nielsen's 10 heuristic principles catalog prepared in the previous stage. When auditing applications, the expert considered that Usevalia helps to protect audits against possible changes to the applications being audited (score of 4.43 for perceived usefulness in Table 5), as applications are in a constant cycle of continuous improvement as part of the software development process itself. The expert's assessment was based on the fact that Usevalia allows auditors to persistently store screenshots that help clarify where the problems are when assessing the guidelines. Despite giving the auditor the option of commenting on the assessment, Usability Datalogger does not allow screenshots to be taken (score of 2.43 for perceived usefulness in Table 5).

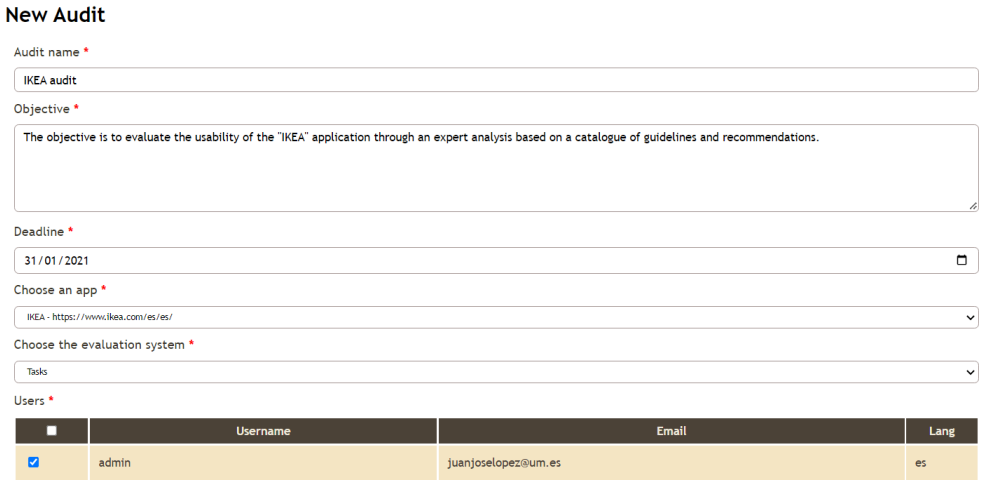


Figure 17. Usevalia graphical interface on the audit of the IKEA application is defined

5.2.4. Generate audit report

Finally, once the evaluation of the web and mobile applications had been completed, the expert generated a report showing the results of the audit. In this case, Usability Datalogger does not allow audit reports to be generated automatically, while Usevalia allows the generation of a detailed audit report in PDF format (see Figure 18). The expert assigned a medium and low priority to the availability of export formats and exporting reports adapted to the type of evaluation, respectively (see Table 1).

	AUDIT REPORT	Date: 17/01/2022

Audit Name:	IKEA AUDIT
Evaluated App:	IKEA
Evaluated Catalog:	IKEA catalog
Audit Leader:	rsobrino
Auditors Group:	admin - INTU24
Audit Date:	10/01/2022-24/01/2022
Evaluation System:	Tasks
Audit Objective:	The objective of the audit is to assess the usability of the "IKEA" application through an expert analysis based on a catalogue of guidelines and recommendations.
Exit Criteria:	Number of allowed failed scores on each priority level: Low: 2 Medium: 4 High: 1

Evaluation Results
According to the exit criteria presented on the audit information, the audit has not been successfully completed.

Figure 18. Beginning of the audit report of the IKEA application via Usevalia

5.2.5. Final considerations as regards expert audit

The two most important variables when predicting the development of a technological tool using the TAM model are perceived usefulness and perceived ease of use, which have a direct relationship with the end-users' intention to use the tool (Davis et al., 1989). Several studies (Chau & Hu, 2001; Dutta et al., 2018; Smoker et al., 1988) agree as to the positive and significant relationship between the perceived ease of use and perceived usefulness. The perceived ease of use affects the perceived usefulness, signifying that if users feel the system is easy to use, they will feel that the audit tool is useful. Moreover, they will perceive that they are prepared to use the technology. The causal relationship between these two variables has been confirmed by previous empirical research (Venkatesh & Davis, 1996).

The results of the evaluation by the usability audit expert (see Table 5) can be employed as a basis on which to state that, in general, the expert perceives the Usevalia tool to be more useful than the Usability Datalogger spreadsheet. For example, the expert thought that Usevalia allowed him to perform the tasks faster (Q1) when compared to Usability Datalogger. In the same vein, the expert strongly disagreed that Usability Datalogger improved his current performance (Q2), in contrast to his opinion of Usevalia (strongly agree), which allowed him to improve his current performance because it has the set of functionalities described in the process outlined in Section 4.5. Similar results were obtained when asked about the productivity generated when using such an

application (Q3), whether these tools make the user more efficient (Q4), and whether the tools facilitate the users' work during an audit process (Q5). On a general level, the expert considered that the tools facilitated his work in a usability audit process because they have a set of functionalities that contribute to this. However, Usevalia has several features that Usability Datalogger does not have, because it was not created with the specific objective of being used in audit processes, whereas Usevalia was conceived with this objective in mind from its initial development. Finally, the expert considered that Usevalia is useful (Q6) because it allowed him to perform an analysis of the audit results through the use of various static graphs on the audit. In the case of Usability Datalogger, the expert neither agreed nor disagreed as to whether it is useful because he considered that he did not have sufficient elements to be able to state this.

The perceived ease of use can significantly influence users' intention to use a system in some specific contexts, such as Using an Online Learning Community for the design of courses (Liu et al., 2010). However, in the context of conducting usability audits, an audit process support system such as Usevalia would be used by individuals for specific purposes, and individuals would, therefore, be primarily concerned with whether or not the services offered by the system were useful to improve and simplify the audit process. If individuals perceived that, although the system was easy to use, it did not improve their auditing activities, their attitude towards the use of Usevalia would not improve in any way.

With regard to perceived ease of use (see Table 5), the expert's evaluations indicated that both tools were easy to learn to use (Q7). This is because the expert has extensive experience in handling TIC. Another aspect rated in the questionnaire concerned whether it was easy to get the tools to do what he needed them to do (Q8). On this occasion, the expert disagreed that it was easy to get Usability Datalogger to do what he needed because it was developed with certain functionalities, and the user who needs it must adapt to what the tool offers. However, he agreed that it was easy to get Usevalia to do what he needed because its functionalities are typical of a usability audit process. The expert agreed and strongly agreed that his interaction with both tools was clear and understandable (Q9). With regard to whether the tools were flexible (Q10), the expert did not agree that Usability Datalogger was flexible, whereas Usevalia was. Finally, the expert indicated that he found it easy to become proficient in the use of the systems (Q11), and his perception was, therefore, that both tools are easy to use (Q12).

5.3. Threats to validity

The threats to the validity of the study are discussed in this section, along with the attempts made to mitigate them.

5.3.1. Construct and internal validity

In this case, construct validity refers to the extent to which the items selected for the questionnaires align with the underlying constructs that the questionnaires were designed to assess (Sauro & Lewis, 2016). The inadequate measurement of the quality aspects of

the Usevalia tool could have been a threat to construct validity. In order to mitigate this threat, we adapted an existing reliable test instrument: the TAM model. In addition, several iterative reviews of the content of the questionnaire were carried out by three authors, who are researchers in the field of usability engineering. This, therefore, allowed the appropriateness of the questions for the evaluation goals to be validated. The resulting questionnaire was used to obtain the students' perception of the Usevalia tool.

During the selection process, the students who filled in Questionnaire 1 received prior preparation on the Usevalia tool, thus counteracting the internal threat to validity that may be caused by the effect of the confounding variable related to prior knowledge and experience in this topic.

5.3.2. External validity

External validity may also be threatened when experiments are conducted with students; if professionals on usability engineering had been involved in the validation, the representativeness of the participants might have been improved. Nevertheless, controlled experiments provide insight into issues and problems that can later be considered in industrial case studies (Arisholm et al., 2006). As suggested by Carver et al. (2003), the results obtained through empirical studies conducted with students have relevance in the software engineering process (Salman et al., 2015).

The small sample size for the evaluation of Usevalia (n=22 students) reduces the statistical generalizability of the results. However, the consistent positive evaluation of the students participating in the study gives confidence that the method is understandable.

5.3.3. Conclusion validity

As stated previously, the sample size (n = 22 students) is too small to be able to produce acceptable statistical power. It is not, therefore, advisable to consider the results obtained as definitive, and it would be advisable to continue this study with a more representative sample. However, the consistently positive evaluation provided by the students who participated in the study gives confidence that the usability audit process based on heuristic evaluation is understandable. The fact that the validation was carried out by the first author of the paper constitutes a risk to validity. In order to mitigate this, the study and validation were carried out using a rigorous process and published with a high level of detail to allow the process to be reproducible.

6. CONCLUSIONS AND FURTHER WORK

This paper presents Usevalia, a web-based tool with which to support usability audits based on heuristics evaluation. The innovative aspect of Usevalia in comparison to other tools that support heuristic evaluation is that it allows the management of multi-user usability audits that customize both the heuristic dataset and the scoring scheme to be used during evaluation. Usevalia also supports three types of evaluation (basic, standard, task-based) and allows the generation of audit reports adapted to each type of evaluation.

Usevalia eases auditors' work during the evaluation of a usability audit, along with the analysis and compilation of results for the generation of a final report. Time is drastically reduced, since most of the steps have been automated, such as the creation of reports or the calculation of audit results.

In the future, the tool could provide a wider variety of audit reports, e.g. filtering a checklist report with the guidelines that have not been fulfilled or those that contain suggestions for improvement. Moreover, the management of the heuristic dataset directly from Usevalia would make the process of guideline definition easier, especially for users who do not usually work with CSV files. Another improvement concerns the format of the guideline catalogs and consists of the possibility of migrating to ReqIF (Requirements Interchange Format) (Object Management Group, 2016), which is basically an XML format used for requirements interchange that could be adopted in Usevalia.

Finally, a major future research topic is the integration of agile software development principles into a rigorous, inspection-based usability audit process enabled by automated support. Non-functional issues, and in particular, usability concerns are often neglected in agile or lean methods (Fitzgerald & Stol, 2017), and usability is not, therefore, something that can be taken for granted when applying agile in software projects (Salah et al., 2014; Silva da Silva et al., 2012). Reconciling agility and usability, which are often seen as opposing concerns, requires the definition of a new continuous audit usability process, sprint by sprint. As long as the time for audits is limited within each sprint, a kind of estimation and priorization algorithm could indicate the guidelines that might be audited in the current sprint. This process should be enabled by new automated support that could be an evolution of Usevalia.

ACKNOWLEDGMENTS

This research is part of the OASSIS-UMU (PID2021-122554OB-C32) project (supported by the Spanish Ministry of Science and Innovation), the BIZDEVOPSGLOBAL-UMU (RTI2018-098309-B-C33) project (supported by the Spanish Ministry of Science, Innovation, and Universities), and the Network of Excellence in Software Quality and Sustainability (TIN2017-90689-REDT). These projects are also founded by the European Regional Development Fund (ERDF).

REFERENCES

Alhadreti, O. (2022). A Comparison of Synchronous and Asynchronous Remote Usability Testing Methods. *International Journal of Human–Computer Interaction*, *38*(3), 289–297. https://doi.org/10.1080/10447318.2021.1938391

Ardito, C., Lanzilotti, R., Buono, P., & Piccinno, A. (2006). A tool to support usability inspection. *Proceedings of the Workshop on Advanced Visual Interfaces*, *2006*, 278–281. https://doi.org/10.1145/1133265.1133322

Arisholm, E., Briand, L. C., Hove, S. E., & Labiche, Y. (2006). The impact of UML documentation on software maintenance: An experimental evaluation. *IEEE Transactions on Software Engineering*, *32*(6), 365–381. https://doi.org/10.1109/TSE.2006.59

Athauda, R., Skinner, G., & Regan, B. (2008). A methodology to minimise excessively permissive security configurations. *Proceedings of the 8th WSEAS International Conference on APPLIED COMPUTER SCIENCE (ACS'08) A*, 187–192.

Bănărescu, A. (2015). Detecting and Preventing Fraud with Data Analytics. *Procedia Economics and Finance*, *32*(15), 1827–1836. https://doi.org/10.1016/s2212-5671(15)01485-9

Barnum, C. (2021). Usability Testing Essentials: ready, set...test! In *Usability Testing Essentials*. Morgan Kaufmann.

Basili, V. R., & Dieter Rombach, H. (1988). The TAME Project: Towards Improvement-Oriented Software Environments. *IEEE Transactions on Software Engineering*, *14*(6), 758–773. https://doi.org/10.1109/32.6156

Beecham, S., Carroll, N., & Noll, J. (2012). A decision support system for global team management: Expert evaluation. *Proceedings - 2012 IEEE 7th International Conference on Global Software Engineering Workshops, ICGSEW 2012*, 12–17. https://doi.org/10.1109/ICGSEW.2012.14

Bird, D. K. (2009). The use of questionnaires for acquiring information on public perception of natural hazards and risk mitigation - A review of current knowledge and practice. *Natural Hazards and Earth System Science*, *9*(4), 1307–1325. https://doi.org/10.5194/nhess-9-1307-2009

Blazek, P., Kuca, K., Krenek, J., Krejcar, O., & Nepovimova, E. (2019). Comparative evaluation of open source laboratory information and management systems. *2018 IEEE Conference on Open Systems, ICOS 2018*, 19–24. https://doi.org/10.1109/ICOS.2018.8632810

Brooke, J. (1996). SUS - A quick and dirty usability scale. *Usability Evaluation in Industry*, *189*(194), 4–7.

CareerFoundry. (2016). *The Trillion Dollar UX Problem : A Comprehensive Guide to the ROI of UX*.

Carver, J., Jaccheri, L., Morasca, S., & Shull, F. (2003). Issues in using students in empirical studies in software engineering education. *Proceedings - International Software Metrics Symposium*, 239–249. https://doi.org/10.1109/METRIC.2003.1232471

Chamba-Eras, L., Jacome-Galarza, L., Guaman-Quinche, R., Coronel-Romero, E., & Labanda-Jaramillo, M. (2017). Analysis of usability of universities Web portals using the Prometheus tool - SIRIUS. *2017 4th International Conference on EDemocracy and EGovernment, ICEDEG 2017*, 195–199. https://doi.org/10.1109/ICEDEG.2017.7962533

Chau, P. Y. K., & Hu, P. J. H. (2001). Information technology acceptance by individual professionals: A model comparison approach. *Decision Sciences*, *32*(4), 699–719.

https://doi.org/10.1111/j.1540-5915.2001.tb00978.x

Constantinides, E., & Fountain, S. J. (2008). Web 2.0: Conceptual foundations and marketing issues. *Journal of Direct, Data and Digital Marketing Practice*, *9*(3), 231–244. https://doi.org/10.1057/palgrave.dddmp.4350098

Cruz Zapata, B., Fernández-Alemán, J. L., Toval, A., & Idri, A. (2018). Reusable Software Usability Specifications for mHealth Applications. *Journal of Medical Systems*, *42*(3), 1–9. https://doi.org/10.1007/s10916-018-0902-0

Cysneiros, L. M., & Yu, E. (2004). Non-Functional Requirements Elicitation. *Perspectives on Software Requirements*, 115–138. https://doi.org/10.1007/978-1-4615-0465-8_6

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly: Management Information Systems*, *13*(3), 319–339. https://doi.org/10.2307/249008

Davis, F. D. ., Bagozzi, R. P. ., & Warshaw, P. R. . (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*, *35*(8), 982–1003.

Deshpande, Y., Murugesan, S., Ginige, A., Hansen, S., Schwabe, D., Gaedke, M., & White, B. (2002). Web engineering. *Journal of Web Engineering*, *1*(1), 003–017. https://doi.org/10.4018/978-1-59904-492-7.ch001

Dix, A., Finlay, J., D.Abowd, G., & Beale, R. (2004). Human-Computer Interaction - Design rules. In *Pearson Education limited*.

Dumas, J., & Sorce, J. (1995). Expert reviews: how many experts is enough? *Proceedings of the Human Factors and Ergonomics Society*, *1*, 228–232. https://doi.org/10.1177/154193129503900402

Dutta, B., Peng, M. H., & Sun, S. L. (2018). Modeling the adoption of personal health record (PHR) among individual: the effect of health-care technology self-efficacy and gender concern. *Libyan Journal of Medicine*, *13*(1). https://doi.org/10.1080/19932820.2018.1500349

Dzulfikar, M. F., Purwandari, B., Sensuse, D. I., Lusa, J. S., Solichah, I., Prima, P., & Wilarso, I. (2018). Personalization features on business-to-consumer e-commerce: Review and future directions. *2018 4th International Conference on Information Management, ICIM 2018*, 220–224. https://doi.org/10.1109/INFOMAN.2018.8392839

Fazio, R. H., Chen, J. mei, McDonel, E. C., & Sherman, S. J. (1982). Attitude accessibility, attitude-behavior consistency, and the strength of the object-evaluation association. *Journal of Experimental Social Psychology*, *18*(4), 339–357. https://doi.org/10.1016/0022-1031(82)90058-0

Fazio, R. H., & Zanna, M. P. (1981). Direct experience and attitude-behavior consistency. *Advances in Experimental Social Psychology*, *14*(C), 161–202. https://doi.org/10.1016/S0065-2601(08)60372-X

Fitzgerald, B., & Stol, K. J. (2017). Continuous software engineering: A roadmap and agenda. *Journal of Systems and Software*, *123*, 176–189. https://doi.org/10.1016/j.jss.2015.06.063

Gavrila, S. I., & Barkley, J. F. (1998). Formal specification for role based access control user/role and role/role relationship management. *Proceedings of the ACM Workshop on Role-Based Access Control*, 81–90. https://doi.org/10.1145/286884.286902

Güncan, D., & Onay Durdu, P. (2021). A user-centered behavioral software development model. *Journal of Software: Evolution and Process*, *33*(2), 1–26. https://doi.org/10.1002/smr.2274

Hollingsed, T., & Novick, D. G. (2007). Usability inspection methods after 15 years of research and practice. *SIGDOC'07: Proceedings of the 25th ACM International*

Conference on Design of Communication, 249–255. https://doi.org/10.1145/1297144.1297200

Hom, J. (1998). *The Usability Methods Toolbox Handbook*. http://www.idemployee.id.tue.nl/g.w.m.rauterberg/lecturenotes/usabilitymethodstoolboxhandbook.pdf

Hwang, W., & Salvendy, G. (2010). Number of people required for usability evaluation: The 10±2 rule. *Communications of the ACM*, *53*(5), 130–133. https://doi.org/10.1145/1735223.1735255

Intechnic. (2016). *User Experience Audit: Sample Report*. http://cdn2.hubspot.net/hubfs/439788/Content_Offers/UX_Design/User_Experience_Audit_-_Sample.pdf?t=1487693668325

Iqbal, M., Noman, M., Talpu, S. R., Manzoor, A., & Abid, M. M. (2020). An empirical study of popular Content Management System-Wordpress vs Drupal vs Joomla. *ICTACT Journal on Management Studies*, *06*(02), 1214–1219. https://doi.org/10.21917/ijms.2020.0168

ISO 9241. (2018). *Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts*. https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en

Ivory, M. Y., & Hearst, M. A. (2001). The state of the art in automating usability evaluation of user interfaces. *ACM Computing Surveys*, *33*(4), 470–516. https://doi.org/10.1145/503112.503114

Jacoby, J., & Matell, M. S. (1971). Three-Point Likert Scales Are Good Enough. *Journal of Marketing Research*, *8*(4), 495. https://doi.org/10.2307/3150242

Ježovita, A., Tušek, B., & Žager, L. (2018). The state of analytical procedures in the internal auditing as a corporate governance mechanism. *Management (Croatia)*, *23*(2), 15–46. https://doi.org/10.30924/mjcmi/2018.23.2.15

Kendall, M. ., & Smith, B. B. (1939). The problem of m rankings. *The Annals of Mathematical Statistics*, *10*(3), 275–287.

Kendall, M., & Gibbons, D. (1990). *Rank correlation methods* (5th ed.). E. Arnold ; Oxford University Press.

Krug, S. (2014). *Don't make me think, revisited: a common sense approach to Web usability*. Pearson Education. https://doi.org/10.5860/choice.51-6218

Liu, I. F., Chen, M. C., Sun, Y. S., Wible, D., & Kuo, C. H. (2010). Extending the TAM model to explore the factors that affect Intention to Use an Online Learning Community. *Computers and Education*, *54*(2), 600–610. https://doi.org/10.1016/j.compedu.2009.09.009

Mahajan, D., Blakeney, C., & Zong, Z. (2019). Improving the energy efficiency of relational and NoSQL databases via query optimizations. *Sustainable Computing: Informatics and Systems*, *22*, 120–133. https://doi.org/10.1016/j.suscom.2019.01.017

Martin-Michiellot, S., & Mendelsohn, P. (2000). Cognitive load while learning with a graphical computer interface. *Journal of Computer Assisted Learning*, *16*(4), 284–293. https://doi.org/10.1046/j.1365-2729.2000.00141.x

Matera, M., Costabile, M. F., Garzotto, F., & Paolini, P. (2002). SUE inspection: An effective method for systematic usability evaluation of hypermedia. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans.*, *32*(1), 93–103. https://doi.org/10.1109/3468.995532

Molich, R., Ede, M. R., Kaasgaard, K., & Karyukin, B. (2004). Comparative usability evaluation. *Behaviour and Information Technology*, *23*(1), 65–74. https://doi.org/10.1080/0144929032000173951

Morrissey, K. (2014). A Review of ¨ Universal Methods of Design: 100 Ways to Research

Complex Problems, Develop Innovative Ideas, and Design Effective Solutions¨. *Visitor Studies*, *17*(2), 222–224. https://doi.org/10.1080/10645578.2014.945354

Moslem, S., Ghorbanzadeh, O., Blaschke, T., & Duleba, S. (2019). Analysing Stakeholder Consensus for a Sustainable Transport Development Decision by the Fuzzy AHP and Interval AHP. *Sustainability*, *11*(12), 3271. https://doi.org/10.3390/su11123271

Nielsen, J. (1994). Enhancing the explanatory power of usability heuristics. *Conference on Human Factors in Computing Systems - Proceedings*, 152–158. https://doi.org/10.1145/191666.191729

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. *Conference on Human Factors in Computing Systems - Proceedings*, *April*, 249–256. https://doi.org/10.1145/97243.97281

Object Management Group. (2016). *Requirements Interchange Format (ReqIF) Version 1.0.1*. *July*, 104.

Olson, K. (2010). An examination of questionnaire evaluation by expert reviewers. *Field Methods*, *22*(4), 295–318. https://doi.org/10.1177/1525822X10379795

Richardson, B., Campbell-Yeo, M., & Smit, M. (2021). Mobile Application User Experience Checklist: A Tool to Assess Attention to Core UX Principles. *International Journal of Human–Computer Interaction*, *37*(13), 1283–1290. https://doi.org/10.1080/10447318.2021.1876361

Rossi, G., Pastor, O., Schwabe, D., & Olsina, L. (2008). *Web Engineering: Modelling and Implementing Web Applications*. Springer.

Salah, D., Paige, R. F., & Cairns, P. (2014). A systematic literature review for Agile development processes and user centred design integration. *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, 1–10. https://doi.org/10.1145/2601248.2601276

Salman, I., Misirli, A. T., & Juristo, N. (2015). Are students representatives of professionals in software engineering experiments? *Proceedings - International Conference on Software Engineering*, *1*, 666–676. https://doi.org/10.1109/ICSE.2015.82

Sauro, J., & Lewis, J. R. (2016). *Quantifying the User Experience: Practical Statistics for User Research*. Morgan Kaufmann.

Schmidt, R. C. (1997). Managing Delphi surveys using nonparametric statistical techniques. *Decision Sciences*, *28*(3), 763–774. https://doi.org/10.1111/j.1540-5915.1997.tb01330.x

Silva da Silva, T., Selbach Silveira, M., Maurer, F., & Hellmann, T. (2012). User Experience Design and Agile Development: From Theory to Practice. *Journal of Software Engineering and Applications*, *05*(10), 743–751. https://doi.org/10.4236/jsea.2012.510087

Sivaji, A., Soo, S. T., & Abdullah, M. R. (2011). Enhancing the effectiveness of usability evaluation by automated heuristic evaluation system. *Proceedings - 3rd International Conference on Computational Intelligence, Communication Systems and Networks, CICSyN 2011*, *August*, 48–53. https://doi.org/10.1109/CICSyN.2011.23

Smidt, L., Ahmi, A., Steenkamp, L., van der Nest, D. P., & Lubbe, D. (2019). A Maturity-level Assessment of Generalised Audit Software: Internal Audit Functions in Australia. *Australian Accounting Review*, *29*(3), 516–531. https://doi.org/10.1111/auar.12252

Smoker, W. R. K., Lehmann, T. R., Hingtgen, W. L., Spratt, K. F., Gentry, L. R., & Weinstein, J. N. (1988). Reduction of the gastrointestinal side effects of metrizamide myelography with oral dexamethasone. *American Journal of Neuroradiology*, *9*(1),

179–185.
Soileau, J., Soileau, L., & Sumners, G. (2015). The evolution of analytics and internal audit. *Edpacs*, *51*(2), 10–17. https://doi.org/10.1080/07366981.2015.1012441
Suarez Torrente, M. D. C., Conde Clemente, P., Martínez, A. B., & Juan, A. A. (2016). Improving web user satisfaction by ensuring usability criteria compliance: The case of an economically depressed region of Europe. *Online Information Review*, *40*(2), 187–203. https://doi.org/10.1108/OIR-04-2015-0134
Tang, F., Norman, C. S., & Vendrzyk, V. P. (2017). Exploring perceptions of data analytics in the internal audit function. *Behaviour and Information Technology*, *36*(11), 1125–1136. https://doi.org/10.1080/0144929X.2017.1355014
Tharindu, P. K., & Koggalage, R. (2021). Usability of E-Commerce Websites : State of the Art and Future Directions e-Commerce. *International Research Journal of Modernization in Engineering Technology and Science*, *3*(February), 314–319. https://doi.org/10.5281/zenodo.4529411
Tory, M., & Moller, T. (2008). Evaluating Visualizations : Do Expert Reviews Work ? *IEEE Computer Graphics and Applications*, *25*(5), 8–11. https://doi.org/10.1109/MCG.2005.102
Vakode, B., & Chaudhari, D. (2013). Study of content management systems Joomla and Drupal. *IJRET: International Journal of Research in Engineering and Technology*, *02*(12), 569–573.
Venkatesh, V., & Davis, F. D. (1996). A model of the antecedents of perceived ease of use: Development and test. *Decision Sciences*, *27*(3), 451–481. https://doi.org/10.1111/j.1540-5915.1996.tb00860.x
W3C Working Group Note. (2017). *Developer Guide for Evaluation and Report Language (EARL) 1.0*. https://www.w3.org/TR/EARL10-Guide/#changes
WebCriteria. (2002). *Heuristic Evalualtion - A System Checklist*. Usability Analysis & Design. file:///C:/Users/34631/Downloads/Heuristic Evaluation Checklist.pdf
Zanna, M., & Rempel, J. (2008). Attitudes: A new look at an old concept. *Attitudes: Their Structure, Function, and Consequences*.
Zazelenchuk, T. (2008). *How to Install and Use the Usability DataLogger (v5. 0)*. Agedi.Ae. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:How+to+Install+and+Use+the+Usability+DataLogger+(v5.0)#0

Raimel Sobrino-Duque received his M.Sc. degree on New Technologies in Computer Science at the University of Murcia (Spain) in 2021, and a B.Sc. degree in Software Engineering from the University of Informatics Science (Havana, Cuba) in 2017. His research interests include usability and secure and sustainable business processes. Contact him at raimel.sobrino@um.es

Dr. Juan Manuel Carrillo de Gea is an Associate Professor at the University of Murcia. He received the Ph.D. degree in Software Engineering from the University of Murcia in 2016. His current research interests include software engineering, sustainability, medical informatics, and eLearning. Contact him at jmcdg1@um.es

Juan José López-Jiménez is currently a Ph.D. student at the University of Murcia. He received the Technical Engineering degree in System Informatics, the M.Sc. degree in Computer Science from the University of Murcia, and the B.Sc. degree in Computer

Science from the University of Castilla-La Mancha (Spain). Contact him at juanjoselopez@um.es

Dr. Joaquín Nicolás Ros is an Associate Professor in the Department of Informatics and Systems at the University of Murcia. He received B.Sc. and Ph.D. degrees in Computer Science. Currently, his main research interests are in agile, continuous requirements engineering and audit, usability and sustainable processes. Contact him at jnr@um.es

Dr. José Luis Fernández-Alemán is a Professor of Computer Science at the University of Murcia. His research interests include DevOps, testing and computer-based learning. Fernández-Alemán received a PhD in computer science from the University of Murcia. Contact him at aleman@um.es