SUSApp: A Mobile App for Measuring and Comparing Questionnaire-Based Usability Assessments

Alba de Castro and José A. Macías
Escuela Politécnica Superior, Universidad Autónoma de Madrid,
Tomás y Valiente 11, 28049 Madrid, España
alba.decastro@estudiante.uam.es, j.macias@uam.es

ABSTRACT
Usability questionnaires are one of the most used methods to measure usability in terms of the user's subjective satisfaction. However, most of the usability questionnaires do not provide a complete environment to store measurements and compare different usability values of application categories and versions over the long term, which makes it difficult to study the usability of a software product or even the usability of different versions of such products over time, hindering the facility to obtain comparisons and thresholds in usability measurements for different product lines. In this paper we present SUSApp, a tool conceived for the analysis of usability through the SUS (System Usability Scale) questionnaire, which is one of the most popular ones. This tool was conceived for mobile platforms, and it is intended to easily analyze usability by storing and recovering past evaluations, and allowing to statistically compare usability measurements among different software products and applications categories. In addition, a user testing is presented. This has provided acceptable usability results concerning SUSApp in an experiment with real users.

Keywords
Usability; Usability Measurement; System Usability Scale; Mobile App.

1. INTRODUCTION
Usability can be considered as a quality feature, and it has to be assured for software applications to experiment a valuable user experience. There are different ways of measuring usability, being the satisfaction questionnaire as one of the most commonly used. Satisfaction questionnaires provide valid psychometric values to obtain the user's subjective satisfaction concerning one or more dimensions of the perceived usability when interacting with a specific software application.

There are already several commercial questionnaires such as SUMI [14], QUIS [12] and WAMMI [17]. However, they are usually not available for free distribution and further comparison among different software products. In general, most of the existing usability questionnaire tools fail at obtaining valuable information as they do not provide a complete environment for storing measurements and comparing different values of usability through custom software categories and versions in the long run.

In this paper, we present SUSApp, a mobile application highlighting the storage, analysis and further comparison of the values obtained from SUS questionnaire [3]. SUS has proved to be a valuable questionnaire with a high psychometric validity to measure perceived usability [15]. This simple questionnaire consists of 10 questions (half of them are positive, while the others are negative) that participants rate using a 5-point Likert scale (where 1 means strongly disagree and 5 means strongly agree). Also, it provides a method for combining the 10 scores

obtained on a wider scale between 0 and 100. The results obtained has to be interpreted by gathering the combination of the 10 questions as a whole, and not separately. The outcome can be interpreted as a percentage representing the final perceived usability value [13].

In addition to solely interpreting the questionnaire value, which comprises the broader usage of SUS, with SUSApp we extend the scope by providing analytical facilities, using a database to store and support summative evaluations of usability for different software artifacts such as prototypes, services, software products and so on. This enables to compare an artifact with others already evaluated in terms of perceived usability, providing graphical charts and statistical information for further analysis. This way, with SUSApp it is possible to compare the usability of different version of a same software product, or even compare the usability of different product lines and software categories, obtaining statistical information of the assessments. SUSApp has been also evaluated using a user test to carry out a formative evaluation and so obtaining the usability of the application itself. To do so, an experiment with real users has been carried out, reporting positive results concerning the usability of the app.

This paper is organized as follows: Section 2 presents related work and a comparative analysis of different approaches. Section 3 introduces and describes our proposal. Section 4 presents the evaluation conducted with real users, as well as the results obtained. Finally, section 5 presents conclusions and future work.

2. RELATED WORK
In this section, we present similar tools and approaches, highlighting strengths and weaknesses, and even taking into account precursor works [4, 5]. In fact, we have considered the information reported by this study to establish and improve some the initial requirements of SUSApp.

Table 1 shows the most related approaches. For each one, main characteristics, strengths and weaknesses are briefly described to study similarities and common points.

Table 1. Analysis of most related approaches

Approach	SUMI [14]
Characteristics	Used for evaluating the quality of a software from the end user' point of view. It consists of 50 points that the user must check "OK", "I do not know" or "Disagree".
Strengths	- Allows comparisons with previous versions of the same product, allowing to set goals for future development. - Enjoys international recognition. - Useful for subjective measurement of the user's satisfaction or anxiety. - Available in several languages.

Weaknesses	- Insufficient data in the gathering of objective aspects. - Proprietary usage (business license). - Questions can be difficult to answer with the given answers. - Does not provide enough results including descriptive and graphical analysis.
Tool name	**WAMMI** [17]
Characteristics	Assessment tool for web sites consisting of 20 items in a 5-point Likert scale. It is based on the questionnaire filled in by users, providing a measure for perceived usefulness and ease of use. It comprises 3 usage modes: prediction (before the release, providing guidance on the reaction of visitors), monitoring (to study certain behavior) and reference (to know the opinion of the audience with respect to other websites).
Strengths	- Enjoys international recognition. - Includes three possible modes of use. - Available in several languages. - More objective, since it includes open-ended questions. - Evaluates satisfaction and usefulness.
Weaknesses	- Proprietary usage (business license). - It is intended only for websites. - Does not provide enough results including descriptive and graphical analysis.
Tool name	**QUIS** [12]
Characteristics	9-point Likert scale questionnaire comprising 27 questions evaluated through adjectives. It is divided into five sections.
Strengths	- 0-9 Likert scale evaluated by specific adjectives. - Enjoys international recognition. - Divided into sections (better structured).
Weaknesses	- High number of questions. - Proprietary usage (business license). - Does not provide enough results including descriptive and graphical analysis.
Tool name	**USE** [10, 16]
Characteristics	It is one of the most comprehensive in evaluating usability using 4 different dimensions. It consists of 30 items in a 7-point Likert scale ranged from strongly agree to strongly disagree. It also has the possibility of adapting the questionnaire to particular needs.
Strengths	- Measures usefulness, satisfaction, ease of use and ease of learning. - Not a proprietary software. - Divided into sections (better structure). - Ability to adapt the questionnaire.
Weaknesses	- High number of questions. - Does not include any analysis supporting tool. So, it is not possible to compare results with assessments made by other users. - Data have to be further analyzed to obtain valid conclusions. - Does not include any graphical analysis.

Tool name	**CSUQ** [9]
Characteristics	7-point Likert questionnaire comprising 19 questions. The questions are divided into 3 sections (not explicitly shown), the first one, consisting of 8 questions, aims to measure the usefulness of the system, and it also collects data about the overall user satisfaction, efficiency and ease of learning. The second section is focused on the quality of information (7 issues), and the third on the quality of the interface (4 questions).
Strengths	- Not a proprietary software. - Divided into sections (better structure).
Weaknesses	- High number of questions. - Does not include any analysis supporting tool. So, it is not possible to compare results with assessments made by other users. - Does not provide enough results including descriptive and graphical analysis. - The division into sections is not shown.

After the comparative analysis, we concluded that SUSApp, unlike some of the system previously compared, is not a proprietary application. In addition, SUSApp provides a framework to evaluate and compare different software products and artifacts, not only web applications, providing also graphical and statistical facilities for further analysis, which is unusual in other questionnaire-based approaches. Besides, since SUSApp is based on SUS, an easier 5-point Likert scale questionnaire of 10 questions is provided, which is quicker to answer in compassion with other approaches having a high number of questions and Likert points. Our approach combines the limited number of questions with the accuracy of its results. Finally, as will be describe down below, SUSApp comprises a mobile application, being more portable and easier to interact anytime and anywhere.

3. OUR APPROACH

SUSApp is a tool for mobile platforms (Android OS smartphones and tablets), multi-language (Spanish and English), and designed to be intuitive and easy to use. It allows to obtain measurements, based on the original SUS method, in a visual way. On the other hand, SUSApp shows quantitative results and charts, using historical data from previous evaluations to compare the usability of different software products, enabling also different statistical views. SUSApp provides the evaluator with different options to consult historical evaluations, as well as facilities to add comments to each evaluation if desired.

3.1 Architecture

SUSApp has been developed for Android devices, using Android Studio, the official IDE for Android app development, and IntelliJ IDEA. Also, different libraries, such as Android Chart for graphics rendering, have been used.

As shown in Figure 1, the architecture of SUSApp consists of a front-end (clients) and a back-end (server). Clients send requests to the server, which returns responses in JSON format being processed and displayed by the clients. The back-end includes a database management system to store the evaluation data.
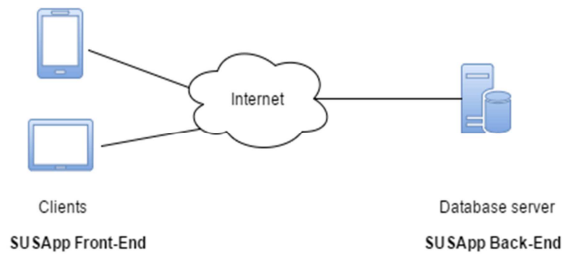
Figure 1. Client/server architecture

Main implementation technologies include: PHP, XML, MySQL and PhpMyAdmin for database management. The application has been designed under a MVC (Model-View-Controller) design pattern, providing an application model highly reusable. The view shows the data obtained from the database located in the back-end, where the information is encapsulated using JSON objects.
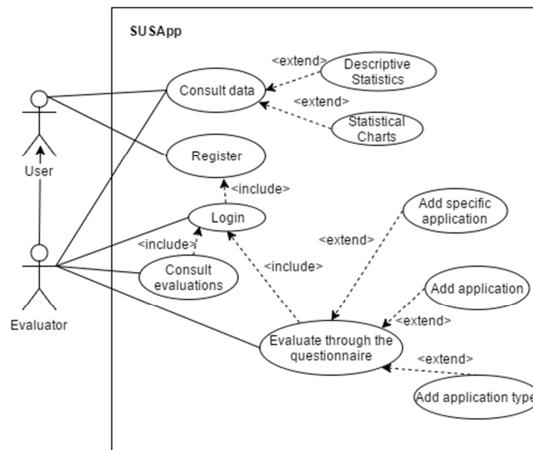

Figure 2. Use case diagram of SUSApp

The main functionalities of the application can be classified into three different groups: evaluation, data visualization and management of past evaluations. The application includes registration and login facilities to access the application and control previous evaluations carried out by evaluators. So, every evaluator has her/his own account in the system, which allows her/him to control sessions and evaluations. Figure 2 shows the general use case diagram of the application, representing how the evaluator operates the system. A user (non-registered evaluator) can query data from the evaluations made by all the evaluators. This data can be visualized using descriptive information or graphical charts. When an evaluator has logged into the system, s/he has access to additional features in the application – i.e., s/he can carry out an evaluation through the SUS questionnaire or consult previous evaluations.

3.2 Information Levels

In order to visualize evaluation data and carry out evaluations, SUSApp structures the information following the hierarchy described in Figure 3, where different application levels are considered. In fact, there are three existing levels: application category, application family and the specific application together with its version and OS (Operating System). Figure 3 shows this hierarchy, with specific examples for each level.
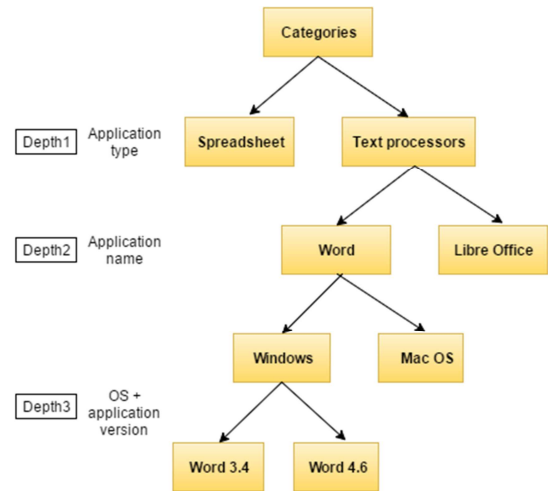

Figure 3. An instance of information levels

When performing an evaluation, the evaluator has to navigate throughout the proposed hierarchy, selecting a particular application to evaluate. For instance, following the scheme appearing in Figure 3, in the first level we can choose a specific applications category such as "Text processors." Then, we can choose a concrete application name between "Word" and "Office". Finally, and once "Word" has been selected for instance, we can choose the OS and version of the software product; in this case: Word 3.6 for Windows. So, the path for this element is: Text Processors -> Word -> Windows -> Word 3.6.

The evaluator can add, at any level of the navigation, a new item comprising a new application, category name, and so on if the existing information (previously created by other evaluators) do not fit the new evaluation requirements. This way, it is also possible to evaluate other software artifacts, such as prototypes, models, services, and so on, by creating a specific category for it. To do so, the evaluator can press on the "Add" button, typing the desired new information that will be stored in the database, and so expanding the navigation tree. This functionality is shown in the use case diagram of Figure 2, where the evaluator, who is evaluating through the questionnaire, can add an application type, an application or a specific application (with OS and version) if desired.

Similarly, when the user (a non-registered evaluator) wants to obtain information from usability evaluations, s/he can choose the level of the hierarchy to view and compare. For instance, the user may want to compare different application types, names or specific versions and OS if desired. To do so, the user simply navigates through the hierarchy and selects (using a checkbox) the items that s/he wants to compare and visualize, choosing among a graphical representation of the data or a textual-descriptive one. As for descriptive representations, numerical values such as the mean, standard deviation, minimum and maximum scores, median, as well as the 95% confidence interval, are presented in a textual way. By contrast, if the user chooses to view graphical charts, this information is represented using two kinds of charts: bar and box plots. These charts are easy to understand and give a formal and comparative view of the usability achieved over the time, as it will described in detail down below.

3.3 App Walkthroughs

In order to explain how SUSApp works and have an idea of the user interface, different walkthroughs are described for the most important use cases of our application detailed in Figure 2.

It is worth mentioning that the graphics and statistical results shown in this section have been previously generated from existing data included in the database, which become from evaluations carried out by experts.

3.3.1 Evaluation

Once the evaluator has registered and logged in, s/he can evaluate through the SUS questionnaire or consult previous evaluations (see Figure 4).
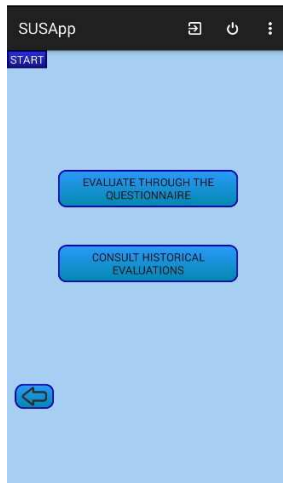

Figure 4. Application menu screen

When the evaluator chooses "Evaluate through the questionnaire", the next three screens (see Figures 5.a, 5.b and 5.c) allow to navigate through the different tree levels of information discussed in the previous section. First, the evaluator has to select an application type. At any level, the evaluator may want to create a new item by clicking on the "Add" button. As shown in Figure 5.a., the evaluator has selected "Text Processors" as application category. Once the category has been selected, next level indicates the applications family (Figure 5.b); in this case the evaluator chooses "Microsoft Word". Finally, once the application is chosen, the evaluator selects the OS and application version, which was in this case, as shown in Figure 5.c, "Word 2015 2.0 for Windows 8".

Once the evaluator has selected the application to evaluate, the next two screens shows the 10 questions of the SUS questionnaire to fill in (see Figures 6.a and 6.b). It is worth mentioning that the questionnaire's questions will be displayed in one or two windows depending on the size of the device. The evaluator can move back and forward anytime by pressing the arrows or the navigation menu on the top. In the case of going back, selected options are saved. In the last SUS screen, the evaluator can add an evaluation comment by pressing on the "Add comment" button. If this button is pressed, a new screen with a textbox appears where, once the comment is inserted and the evaluator has pressed on the button "Ok", the app returns back to the second screen of the questionnaire (Figure 6.b). Finally, if the evaluator has answered all questions and pressed on the button "Save", the evaluation is recorded and s/he is redirected back to the main menu shown in

Figure 4. If the evaluator leaves any question unanswered, s/he is conveniently warned.
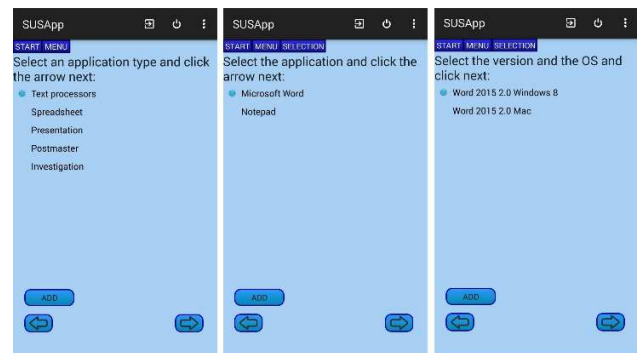
Figures 5.a, 5.b, 5.c. Hierarchical navigation screens
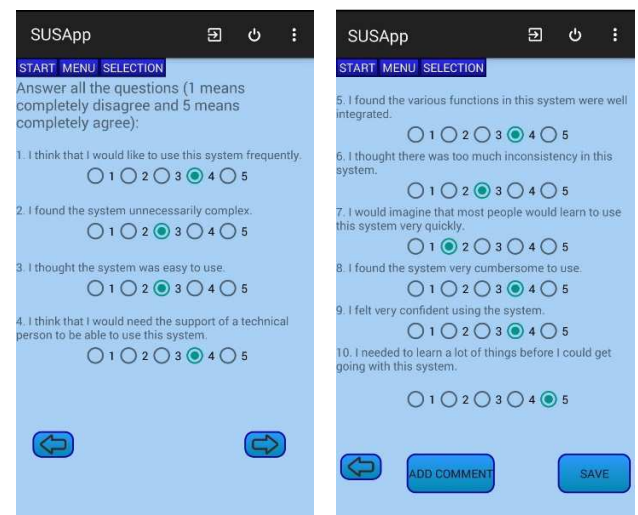
Figures 6.a, 6.b. Screens for the SUS questionnaire to fill in

3.3.2 Visualizing Graphical Scores

Any non-registered user can consult evaluation scores by using the menu depicted in Figure 7, choosing between "Descriptive Statistics" or "Statistical charts". In this case, the user has selected "Statistical charts" to show graphical results. Next, a new screen appears to select the application category, name or version as previously explained in Section 3.2 (see Figure 8).

To follow up, the user has chosen "Application version and OS" to compare different text processors. The available choices (existing from previous evaluations) can be seen in Figure 9. More specifically, two different version of Microsoft Word text processor have been selected: Word 2015 2.0 for both Windows 8 and Mac, as the user wanted to analyze and compare the usability of a same product for different operating systems.
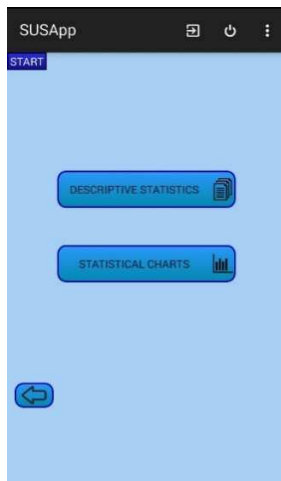
Figure 7. Screens for visualizing statistics data

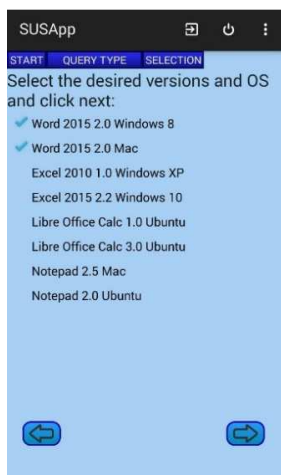**Figure 8. Screen for choosing the data to compare**

Figure 9. Screen with the options available to choose

After choosing both items and pressing the arrow next, two kinds of graphics (a bar graph and a non-parametric box plot, with

values obtained from SUS questionnaire with a 95% confidence interval) are displayed, as shown in Figures 10 and 11.

Figure 10 shows the bar graph comparing the application Word 2015 2.0 for both Windows 8 and Mac operating systems. As we can see, the average score is very similar in both cases (Windows 8 has approximately three points more than Mac). The 95% confidence interval in Windows 8 is smaller than the one corresponding to Mac, so we can finally conclude that, according to the evaluations available in the database, Word for Windows 8 has a slightly higher average value than for the Mac operating system, and its average is more reliable also for such operating system.
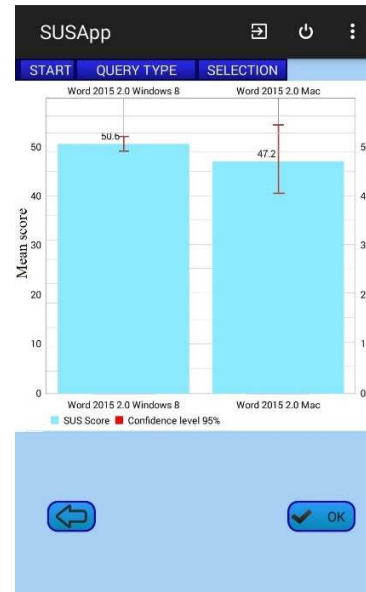

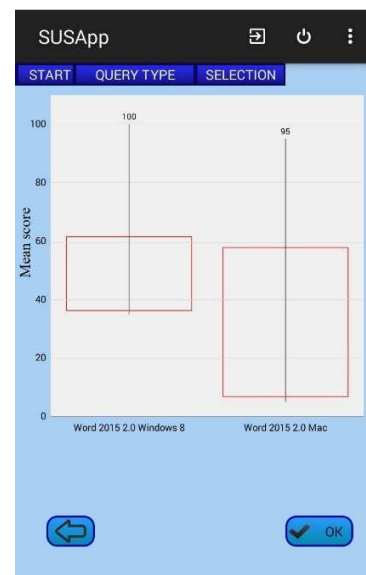**Figure 10. Screen for Bar charts**

Figure 11. Screen for Box plots

As for the box plot shown in Figure 11, we can see information based on minimum, maximum and the following quartiles: Q1 (25%), Q2 or median and Q3 (75%) of the SUS scores obtained.

Word 2015 2.0 Windows 8 has the following values: min=35.1, Q1=36.3, Q2=49.7, Q3=62.6 and max=100. On the other hand, Word 2015 2.0 Mac has the following values: min=5.3, Q1=6.7, Q2=33.5, Q3=56.8 and max=95. As we can see in the case of Windows 8, Q1 and A3 are closer to one another (IQR=26.3), unlike in the Mac OS (where the IQR=50.1). This way, and according to the evaluations available in the database, we can conclude that evaluations for Windows 8 have less dispersion.

3.3.3 Displaying Descriptive Data

As explained before, a non-registered user can choose to display statistical information in text format, instead of visualizing graphics. To do so, the user only has to press on "Descriptive Statistics" (see Figure 7). Once chosen, as in the previous case, the user has to select the level of the information that s/he want to have (see Figure 8). In this case, the user has pressed on "Application Type" and then on "Text Processors". This way, as shown in Figure 12, the descriptive data for the selected application type is displayed. Descriptive data includes: average, standard deviation, maximum and minimum scores, 95% confidence interval and median. This information is useful to analyze the descriptive statistics corresponding to a product family or a specific product. In this case, only one item can be selected, so this option is not suitable for comparisons among software products.
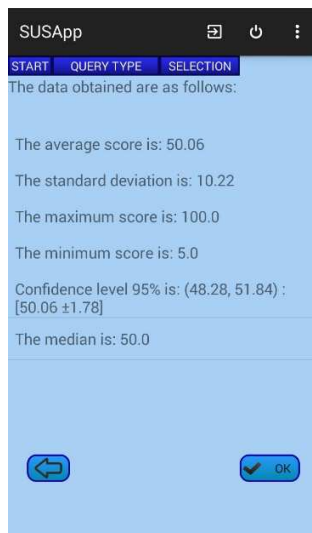

Figure 12. Screen for descriptive statistics

4. EVALUATION

In order to evaluate the usability of SUSApp, we proceeded to design a controlled experiment with real users. Enrolled users shared common characteristics as will be explained down below. It is worth mentioning that, because the selection of participants was not performed randomly and there was not any preselection process, this experiment can be considered as quasi-experimental.

4.1 Evaluation method

As will be described in detail, the context of use consisted of 10 users carrying out 6 representative SUSApp tasks in the same Android device in order to mitigate possible increases in time due to the different characteristics of each terminal.

In order to consider specific metrics, the evaluation was based on the ISO 9241-11 [7], providing values corresponding to effectiveness, efficiency and satisfaction.

The experiment comprised two parts: the first one consisted for each user to accomplish 6 task using SUSApp. We used the Thinking Aloud protocol [8] also measuring, at the same time, metrics such as effectiveness for each task and efficiency calculated as the time to complete each task. The second part consisted for each user to fill in a questionnaire to evaluate the subjective satisfaction. We used the USE questionnaire. This questionnaire provided a good reliability, reporting scores based on four different dimensions of usability: utility, satisfaction, ease of use and ease of learning. The reason for using USE, instead of SUS, was to have a different questionnaire to evaluate the usability, independently of the questionnaire internally used in the app, and so obtaining different dimensions of the usability.

Each user performed the 6 tasks described in section 4.3. The tasks were provided to the user in a sheet, and a short explanation of each one was also provided, but without providing any clue to carry them out.

It is worth noting that, in order to have a more reliable assessment of the usability of the system, the users were not given any previous tutorial on using the application, but only a basic description of it.

4.2 Participants

All users were students of Computer Engineering Degree, having an Android device and being familiar with the use of applications developed for this OS. Specifically, we recruited 10 persons, 5 men and 5 women, aged between 21 and 24 (M=21.9; SD=1.1).

4.3 Defined tasks

For the experiment, 6 different tasks were proposed:

• Registration (T1): the user has to satisfactorily register in the application.

• Visualize graphical scores (T2): the user has to consult the SUS scores by visualizing graphs. To do so, first the level of detail must be "version + OS of the application". The user had the choice to select any number of applications greater than 1, and any of the two graphs (bar or box plot) available.

• Consult descriptive data (T3): the user has consult the SUS score as descriptive data, selecting "Text processors" category.

• Complete a questionnaire (T4): the user has to carry out an evaluation by completing the SUS questionnaire, and also inserting a comment at the end. The evaluated application was chosen by the user from one of the existing ones.

• Consult historical assessments (T5): The user has to consult the evaluation carried out in the T4.

• Add a new type of application (T6): the user has to add, in the category "Word Processors", the "WordPad" application.

The application's database already contained data generated by experts from previous evaluations. This was useful in order for the users to count on existing information to carry out the tasks.

4.4 Results and discussion

It is worth mentioning that users performed all tasks without the help of the evaluator. Therefore, we obtained a 100% effectiveness. On the other hand, to have an idea of the

application's efficiency and better discuss the results, those are shown in Table 2, where the following efficiency values (in seconds) are depicted: mean, minimum and maximum time to perform each task, standard deviation, median and 95% confidence interval. Also, Figure 13 shows the average time to achieve each task, together with the 95% confidence intervals represented as error bars.

As shown in Table 2 and Figure 13, the confidence intervals are acceptable (about half a minute) except for T2 and T4, where the values are higher. Analyzing task T1, which was the first one for users to be achieved, and so the first contact for the users with the application, we have a confidence interval (CI) of 8.7, which is quite reasonable, as it means that we can be 95% sure that a user will take between 19-37 seconds to register in the system, a time considered quite acceptable for the very first task. With respect to tasks T2 and T4, they have a higher mean and deviation. Analyzing the experimental sessions, we conclude that the reason for these values is because T2 implies the selection or two or more applications to visualize, and this greatly varied depending on the selection determined by the users. Besides, T4 has the highest values in terms of mean and deviation, since the completion of the questionnaire clearly took more time than the rest of tasks. Finally, we observed that values for T5 and T6 are acceptable, as the expertise of the user increased as long as s/he interacted with the system.

Table 2. Statistics observed through the experiment

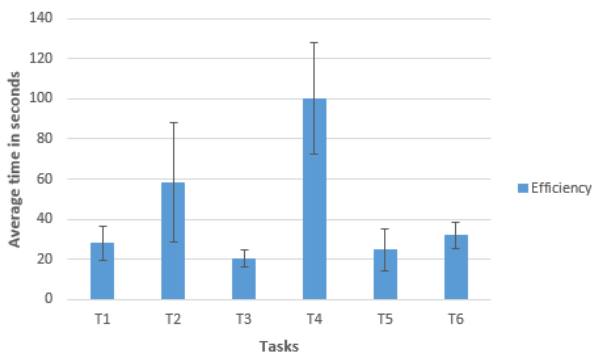|	T1	T2	T3	T4	T5	T6
Mean	28	58.2	20.4	100.2	24.6	31.8
Median	33	55	18	111	20	27
SD	9.9	34	4.8	31,4	11.8	7.3
Min	15	21	16	62	15	26
Max	38	110	28	131	45	45
CI (95%)	8.7	29.8	4.2	27.5	10.3	6.4

Figure 13. Efficiency for each task and CI (95%) as error bars

Finally, Figure 14 shows the results for the USE questionnaire filled in by users after each test session. To perform the analysis, we calculated the average for each question, and then we calculated the averages for each of the dimensions and the corresponding percentages. USE questionnaire comprises 30 questions ranged in a 7-point Likert scale, where 1 means "strongly disagree" and 7 "strongly agree".

As commented before, USE questionnaire reports scores based on four different dimensions of usability: utility, satisfaction, ease of use and ease of learning.
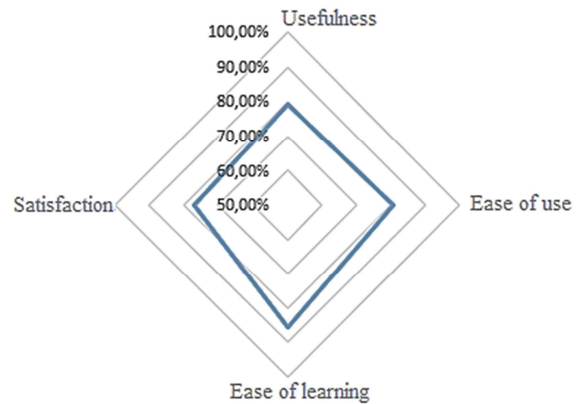

Figure 14. Results for the USE questionnaire

The average usability value, with represents the mean for the four dimensions, was 80.41% (SD=4.07). This value, although acceptable for the first release of the system, it is expected to be improved in the future. Analyzing the four USE dimensions shown in Figure 14, "Satisfaction" had the lower value (76.78%). The dimensions with the highest values were "Ease of learning" (85.71%) and "Ease of Use" (80.95%). In fact, the highest value in the dimension "Ease of Use" was for the question "It is easy to use" (90.50%), and in the "Ease of Learning" dimension the question "I remember how to use it easily" had a high value (90.50%), which means that the application is easy to use and learn. Usefulness also obtained a reasonable value as well (79.16%).

Some qualitative results can be also analyzed from the Thinking Aloud sessions, where most users reported sentences agreeing that the application was very intuitive and easy to use, and it had a simple interface, highlighting also that the navigation bar was useful to quickly navigate through the different screens. By contrast, some negative comments were also identified, related to the fact that the interface did not provide enough help or guiding information for people who are not familiar with box plot representations.

5. CONCLUSION

Usability evaluation is an important issue today. There are different methods and applications to measure usability, which significantly contribute to improve the quality of the software product overall [1, 2, 6, 11]. The user's satisfaction is an important metric to measure usability, since the subjective information reported by users is important to have a perception of the usability as a whole. In fact, questionnaires comprise an important tool to measure perceived satisfaction from users. Although there are different questionnaire and tools to carry out evaluations, we found some deficiencies and drawbacks that avoid having a complete reference framework to analyze usability information from questionnaires and improve designs by validating usability objectives and requirements.

To overcome such drawbacks, we have designed SUSApp, a mobile application based on SUS questionnaire analysis that stores and compares different usability assessments for further

usability analysis. Since usability is a relative concept, it is necessary to compare the value of a usability assessment with others, establishing a threshold to indicate when an application has a reasonable value of usability or not. SUSApp provides statistical representations to have such comparison framework. Since SUSApp is a mobile app, it can be quickly and easily used anytime and anywhere, offering the possibility of achieving statistical comparisons of different artifacts and software products in terms of versions, categories and OS.

The test performed to evaluate SUSApp has reported acceptable usability scores in the four dimensions studied (M=80.41%, SD=4.07). Anyway, we expect to improve the tool in future versions considering also the feedback received from users during the test.

All in all, SUSApp includes good characteristics that can be summarized as follows:

• It is a non-proprietary software.

• It provides storable results that allow further comparative usability analysis and graphical representations.

• It provides historical evaluations management.

• It is based on SUS questionnaire: a 5-point Likert scale of only 10 questions, which results quick and easy to answer. Also, as an improvement, additional comments can be added at the end of the evaluation if desired.

• It is aimed at mobile platforms, giving greater portability and ease of use.

• It is a multi-language app (Spanish and English languages).

As for other weaknesses, SUSApp is focused only on satisfaction, leaving aside other usability attributes. To partially solve this drawback, it has been implemented the possibility for the evaluator to add comments in their evaluations, thus obtaining more information of interest for a further qualitative analysis.

As for future work, SUSApp is expected to be available in Play Store, free to all users. Also, a new improved version will be released, taking into account the results obtained from the evaluations with users. Also, it will be available for iOS in a near future. In addition, new features and evaluations will be accomplished, as well as new data representations and analysis.

6. ACKNOWLEDGMENTS

7. REFERENCES
[1] Baldassarri, S., Macías, J.A., and Urquiza, J. 2014. Trending Breakthroughs in Human-Computer Interaction. *Journal of Universal Computer Science*. 20, 7 (July 2014), 941-1045. DOI= http://dx.doi.org/10.3217/jucs-020-07.

[2] Bangor, A. W. 2000. *Display technology and ambient illumination influences on visual fatigue at VDT workstations*. Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA.

[3] Brooke, J. 1996. *SUS: a "quick and dirty" usability scale*. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A.L. McClellands. Usability Evaluation in Industry. London: Taylor and Francis.

[4] Chin, J.P., Diehl, V.A. and Norman, K.L., 1988. Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI'88). ACM, New York, NY 213-218. DOI= http://doi.acm.org/10.1145/57167.57203.

[5] Davis, F.D., 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*. 13, 3 (Sep. 1989), 319-340. DOI=http://dx.doi.org/10.2307/249008.

[6] Everett, S. P., Byrne, M. D., and Greene, K. K. 2006. Measuring the usability of paper ballots: Efficiency, effectiveness and satisfaction. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*. Santa Monica, CA 2547-2551. DOI=http://dx.doi.org/10.1177/154193120605002407.

[7] ISO 9241-11:1998. *International Standard. Ergonomic requirements for office work with visual display terminals* (VDTs)-Part 11: Guidance on usability. ISO.

[8] Lewis, C. 1982. *Using the "thinking aloud" method in cognitive interface design.* IBM TJ Watson Research Center.

[9] Lewis, J.R. 1995. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *Int. J. Hum.-Comput. Interact*. 7, 1 (January 1995), 57-78. DOI=http://dx.doi.org/10.1080/10447319509526110.

[10] Lund, A. 2001. Measuring usability with the USE Questionnaire. *Usability Interface*. 8, 2, 3-6.

[11] Macías, J.A. Granollers, T. and Latorre, P. 2009. *New Trends on Human-Computer Interaction: Research, Development, New Tools and Methods*. Springer. DOI=http://dx.doi.org/10.1007/978-1-84882-352-5.

[12] QUIS Questionnaire. Retrieved June 21, 2016, from The Laboratory for Automation Psychology and Decision Processes Web Page: http://lap.umd.edu/quis/.

[13] Sauro, J. 2011. Measuring Usability with the System Usability Scale (SUS). Retrieved June 21, 2016 from: http://www.measuringusability.com/sus.php.

[14] SUMI Questionnaire. Retrieved June 21, 2016, from SUMI Web Page: http://sumi.ucc.ie/.

[15] Tullis, T. and Albert W. 2008. *Measuring the User Experience*. Morgan Kaufmann.

[16] USE Questionnaire. Retrieved June 21, 2016, from Gary Perlman's Web Page on User Interface Usability Evaluation with Web-Based Questionnaires: http://garyperlman.com/quest/quest.cgi?form=USE.

[17] WAMMI Questionnaire. Retrieved June 21, 2016, from WAMMI Web Page: http://www.wammi.com.