Visualizing Usability - Extended Abstract

Filipe Morgado
filipe.morgado@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

May 2018

Abstract

While developing a product and more specifically an interface, measuring User Experience and how users interact with it is imperative, since a poorly usable interface can affect users' willingness to interact with it and cause a decline in the overall user interest and satisfaction. The measurement and analysis of Usability related metrics can sometimes be hindered due to the nature and complexity of this data, resulting in a poor extraction of such information. Therefore, Information Visualization can be used as a mechanism to better study and detect product deficiencies at an early stage and represent a powerful tool to improve the analysis of usability and complement user testing, providing insight for changes and, ultimately, an interface with a suitable and pleasurable experience for users. This work addresses ways of exploring and correlating usability metrics and possibly identifying patterns through the usage of Visualizations, which can potentially help draw "high level" conclusions based on both derived and statistical measures, and thus contributing to better results while designing interfaces. Furthermore, the presented work includes a validation approach through both Usability Testing and Case Studies on the developed solution.
Keywords: User Experience; Information Visualization; Analysis; Usability; User Testing.

1. Introduction

Usability is very important in a way that it helps designers and developers to better understand how their interface presents itself and interacts with the users, as well as ensuring that it is both easy to learn, effective and enjoyable for them, providing an overall positive experience. Disregarding usability as a whole can compromise the overall interaction between an interface and users, since if it is hard to learn and to manage, they will often leave and look for other options. These situations can be managed and improved by observing customers and by involving them in the design process, allowing developers to introduce a more stable and ready final product in the market. Some of this approaches share the intent of retrieving feedback and performance metrics, in order to evaluate how easily users interact with the design, but the problem resides in the difficulty in extracting information and analyzing data with such complexity and nature.

Research on this topic was made and a shortage of work relating Usability and Information Visualization was found. Thus, the assessment of usability in this context represents a new challenge and hopefully a new way to explore these areas. It is known that Information Visualization is used as a way to provide contextual information, but there must be concerns regarding how it is visually encoded: as expressed by Munzner[1], some representations might incur additional cognitive workload to the user, rather than the optimal, which is reducing it. Ideally, a visualization is developed in order to reduce the cognitive load associated with the analysis of the metrics represented, and to provide users with intuitive ways to answer their questions, as well as finding patterns and correlations between the datasets. Therefore, an interactive visualization regarding this type of data can, ultimately, help identify problems during the product's early stages and force developers to address concerning areas more effectively.

1.1. Objectives

The main objective of this work is to **Study interactive ways of assessing Usability measures through Information Visualization**. In order to reach this goal, a number of other tasks were achieved along with the development process, such as defining a conceptual idea for an Information Visualization Dashboard and which questions were the most important for the visualization to answer, as well as how the data used would be mapped. The design of the proposed solution and changes made throughout the several design iterations relied on a *User Centered Design* approach and therefore each option and decision made during its evolution was substantiated accordingly.

1.2. Document Structure

The presented document is structured as follows. Chapter 2 includes a background study and fundamental concepts used. Chapter 3 presents study done on related work and what conclusions the authors derive from a comparative discussion made. The prototype's evolution and approaches adopted regarding its development is described on Chapter 4. Chapter 5 presents two evaluation methods used on the final phase of the solution, and Chapter 6 draws the final conclusions and limitations, as well as future improvements for the presented solution.

2. Background

While designing an interactive product, whether it is a cell-phone, a remote control or even a calculator, one has to pay attention not only to the product's capability of carrying out their tasks, but also have in mind certain aspects, such as if it is easy enough to use, effective and enjoyable from the user's perspective. Therefore, it is critical to take into account where the product is going to be used and who is going to use it, since different populations may have different needs and requirements.

2.1. Usability

One of the major concerns while designing a product is optimizing the user's interaction with the interface provided, so that the activities that are supported can be matched by the users. This process should be iterative, parallel and incremental, involving continuous user feedback in order to develop new features or improving existing ones, as mentioned by Jakob Nielsen [2].

By supporting the development process on user's understanding, there should be some concerns regarding which users will use the product, as well as the requirements and needs the target group has. Not only do users need to perform successfully the tasks they set themselves to, the final solution must also provide an overall quality user experience to whoever uses it. In order to reach these goals, listening to what people need and getting them involved in the design allows developers to better understand their needs and goals and improve the overall usability of the product, as described by Preece et. al [3], but it should be validated through Summative evaluations and using different metrics based on user testing.

In short, Usability is a very important component of interaction design and it is generally regarded as ensuring that the interactive interface is both easy to learn, effective and enjoyable to be used, providing an overall positive experience. The usability improvement is made by optimizing the interactions people have with the interface, making it easier for them to carry out their daily tasks, whether it is at school, work or in their everyday life. From a user's perspective, usability is crucial because it establishes the difference between successfully performing a task (with both accuracy and completeness) and failing, whether it is by committing errors and ending up frustrated, or by simply not being able to reach his goal. Studies performed have shown that people prefer to use systems that provide the highest usability metrics 70% of the time, according to Nielsen [4].

2.2. User Experience and User Centered Design

With the increasing use of new technologies in various application areas by the population, such as education and entertainment, new concerns have been brought. Researchers not only have to worry about improving efficiency and effectiveness, but also have to bear in mind what the system represents and feels like to the users, in other words, the User Experience. With the ever growing importance of the users' satisfaction while developing a product, and since there is not a clear path to reach all of the usability and user experience goals due to subjectivity, information about the users and their tasks has been taken in order to adapt the design. Therefore, it is of great importance to continuously ensure the development of the design by involving real users throughout a *User Centered Design* approach, enabling a better overall understanding of the needs and goals, leading to a more appropriate and usable product.

While *User Centered Design* promotes both expectation management, ensuring that the product is similar to what the user had in mind, and a sense of ownership since they participated in the design, there are some concerns. The reasons that are more consistently cited for not involving users in the development stage is the amount of time it actually takes to organize, interact, manage and control this process. It takes both time and resources to prepare interviews, meetings, workshops and user testing laboratories. There is still a debate about how actively should the users be involved in the design, as some studies have shown that largely depending on user involvement can hinder the process: Wilson et al. [5, 6] presented some useful information on the topic, and concluded that users have to be educated about the design, are unaware of implementation constraints and consume precious time to the developers with meetings arrangement. Despite all these concerns, *User Centered Design* is used with the sole purpose to optimize opportunities for success, providing a better final product to the users and can actually translate into money and life savings.

2.3. Usability Measurement

The favored method to ensure the quality of the design is testing with actual users on a working system, requiring focus on the end-users' needs and adapting the design based on them. Some factors can hinder this approach, such as the users' schedule and the budget of the development team, reduce its proficiency. There are other alternatives, such as user testing on prototypes (both low and high fidelity ones), cognitive modeling or a usability audit conducted by experts using heuristic evaluation. Usability evaluation has been around since the 1980s, where similar techniques have been used for this purpose. In heuristics evaluation, experts are guided by heuristics and simulate what an actual user would do, by stepping through tasks and thus identifying the major problems with the system. Compared to user testing, this approach is far less expensive and does not have the problem of gathering a group of users nor preparing a laboratory. After experts detect a list of problems and conditions, feedback is reported to the developers in order to improve the interaction with users. User testing is used so that the product can be tested and experimented by a group of users from the focus group, and to guaranty that it is usable by the time it is released to the clients. Thus, every piece of information that can be used from these tests is collected: the time taken to complete a task, the number of errors, the type of errors and sometimes the path chosen by the user [3]. To make sense of this information, observational data, recordings, user satisfaction questionnaires and interviews are also used for context.

3. Related Work

While there is plenty of work developed in the area of information visualization, there is a lack of research relating it to usability metrics, and thus the visualizations studied were categorized into three different types: pure Statistical Analysis, Correlations and Performance/Benchmarking, allowing for a comparative discussion and identifying potential flaws and strengths of each type. In order to provide a novel visualization and to better explore usability, the analysis of each type of visualization contributed to the related work research in a sense that it allowed the authors to better understand both the vulnerabilities and advantages of each visualization technique used, and therefore improve the decisions regarding the types of visualizations implemented in the solution.

3.1. Comparative Study

The importance of Information Visualization in statistical analysis has been increasing steadily following Tukey's Exploratory Data Analysis [7]. A comparative study between InfoVis and Statistical Analysis developed by Gelman et. al [8] suggests that there is lack of cooperation between both fields, and that by joining InfoVis' capacity of attracting people's attention with the conveyance of data provided by pure statistics, both areas could improve their visualizations. The brazilian population's voting behavior was analyzed in *CivisAnalysis* [9], where a set of visualizations was implemented in order to represent different political positions and dimensions. Expected InfoVis behavior was used, from filtering by parties, states or deputies, and other data exploration features, while the overview controlled the majority of the interactive behavior of the visualization through a timeline based both on period and party selection.

Answering questions about complex relationships between multi tabular data requires a statistical analysis technique to make them visually discernible. Kamasan et. al [10] introduced a study to visualize statistical tests and to display the distributions of data using color schemes, providing the user a tool to quickly evaluate large sets of data for statistical correlations. Sandra Cano et. al [11] developed an interesting study on visualizing structures, patterns and correlations in the context of Spanish pronunciation. The quality of the input signal is compared with the training done on the vowels and these differences and similarities are highlighted through the use of Chernoff faces [12] and diagonal lines. The assessment of performance related metrics has been commonly obtained through the analysis of simple visualization tools as graphs and bar charts, such as Price's [13] and Gusev's [14] work.

In 2016, Price et. al [13] studied and examined whether the use of an info-vis could lower working memory demand on elder people and if it would consequently affect complex decision making in a positive way, while displaying the results through bar charts. On the other hand, Gusev et. al [14] presented in 2014 an extension of their *EDUCache Simulator*, which consisted in graph based visualization of the cache performance behavior when a program accesses huge data arrays by analyzing real three level cache systems.

After a comparative analysis performed on the existing related work, statistical visualizations provide great tools that allow exploring, correlating information and interact with appealing techniques. Some of the views used over the assessed work include bundle, network, clustering and tag clouds visualizations, but they commonly lack objectiveness when dealing with large datasets and thus have problems with scaling. On the other hand, performance visualizations present contrary characteristics, where its simple techniques, such as graphs, bar charts and time line based visualizations compromise user control functions, while accessing clearness and scalability. Finally, the correlation visualizations showed mixed results, affording lower interactive power compared specially to statistical ones, and presenting scaling vulnerabilities while affording great tools to compare and connect information through necklace, density and heatmaps.

The authors stress that the nature of this thesis' solution proposal is innovative and that there is not much work nor studies relating User Experience and Information Visualization, thus the lack of absolute correspondence between the related work and the presented solution. Overall, the authors acknowledge *CivisAnalysis*, and along with the other visualizations studied, believe they can represent important cornerstones throughout the development of the final solution.

4. Proposed Solution

This Chapter intends to describe the evolution of the present solution, focusing on the development approaches used throughout the process and how changes adopted were supported on formative evaluation steps.

4.1. Development Approaches

The progress of the visualization, which envisions to facilitate the extraction of information regarding usability metrics, is based on a *User Centered Design* approach, where users help finding usability problems with the prototype in early stages, so they can be fixed before the final iteration of the design. Based on this description, a focus group of six elements with previous experience with both Information Visualization and Usability was used, allowing continuous judgment on the quality of the prototype and providing information and potential resolutions for future improvements, through discussion sessions and formative evaluations. A survey was handed to 67 users from the target audience and the authors interviewed 3 Usability experts, allowing a perception of the first conceptual idea and what Information the solution intended to show, through the first set of questions.

After some informal sessions of discussion and brainstorming with members of the focus group, the authors decided that some ideas proposed were not necessarily converging with the main objective of this thesis, making them dispensable and thus discarded from the initial idea. Consequently, these changes originated a revision on the set of Questions previously defined which the authors pretended to answer through the help of the *Information Visualization* solution. Thus, following is described the revised set of twelve questions, which would serve as guidelines throughout the development phase.

- What is the time distribution between all sections for a given task?
- Is the section that takes users the most time also the most error prone?
- Do users make more errors when trying to spend less time on a given section?
- Are areas in the interface with higher cursor movement associated with more errors?
- Is there any individual test which is an outlier in terms of time/#errors?
- Which individual tests failed to complete their tasks?
- Are the most time consuming section areas associated with more errors?
- Where is the exact location of the errors made by an individual test?
- What is the most common error made across all tests for a given task?
- Does *System Usability Scale* rating correlate with time/#errors?
- What is the average time spent/#Errors/*System Usability Scale* Rating for a given group of individual tests?
- Do tests with similar metrics (time and #errors) share similar *System Usability Scale* Ratings?

The authors felt that this set of Questions assembled not only important conclusions regarding the Usability status of a given interface, but also kept a cohesive compilation of insight users pretend to obtain while analyzing such visualization.

4.2. Development Phase

After the first conceptual idea was subject of discussion and brainstorming sessions with members from the focus group, the first low-fidelity prototype emerged. Since this first prototype had the main goal of examining the visual representation and key elements present in content, the limited interactivity associated with this technique suited our purposes, while allowing a much clearer expectation about the upcoming stages of the solution for users. Supported on an *Iterative Design* approach, the first Heuristic Evaluation was performed on the low-fidelity prototype and the violations reported from users allowed the authors to improve the solution and to test a first functional prototype.

The Dashboard Visualization consisted of six visualizations that envisioned to analyze different section areas on an interface, as well as how testing users performed within them. The following figures present early stages of the three different idioms used.

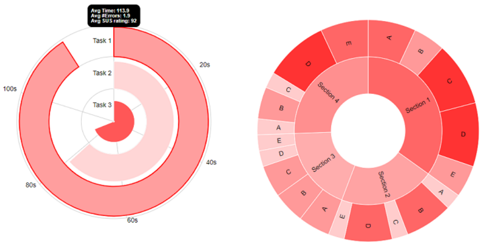

Figure 1: First idiom of the functional prototype.

Fig. 1 presents the first idiom, composed by a Radial Chart (left) and a Sunburst visualization (right), with the main purpose of selecting and beginning the exploration process for a specific task or area section that the user wants to assess. Users can analyze the different tasks and how they compare, and the Sunburst allows users to observe test participants' performance for each area section of the interface.

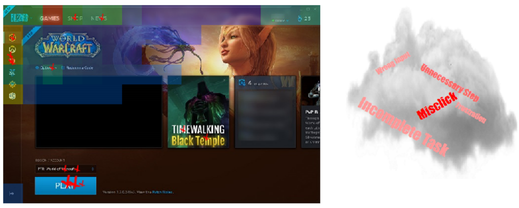

Figure 2: Second idiom of the functional prototype.

Fig. 2 presents the second idiom, composed by a Heatmap visualization and complemented by an Index like visualization as a Word cloud. The objective of this idiom was to provide an intuitive way of identifying concerning section areas on the interface through mouse usage and error locations. Users can explore the different section areas of the interface by cursor movement and analyze errors distribution. The Word Cloud provides a filtering option as well as information regarding a specific type of error.

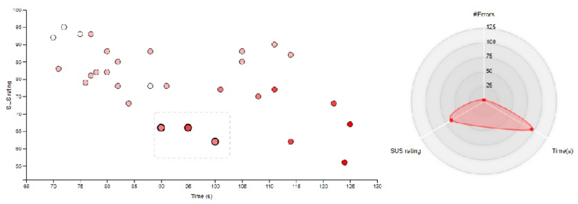

Figure 3: Third idiom of the functional prototype.

Fig. 3 shows a Scatter plot which visualizes and correlates the different Usability metrics assessed (Time, Errors and Satisfaction), along with a Radar Chart visualization that quantifies and compares the individual tests selected by the user in the Scatter plot. Users can explore through the individual tests population and use a selection box to further assess a specific cluster of test instances, with the help of the Radar Chart.

Such idioms were tested in different dispositions, bearing in mind the goals and objectives associated with each idiom described. Along with it, the focus group contributed throughout the development and evolution of the dashboard, as important questions regarding the design were discussed. One of those examples is the Radar Chart and how it would address the representation of three metrics with different connotations associated: while time and number of errors share a negative connotation, higher *System Usability Scale* Rating are generally associated with positive outcomes. As for the representation of individual tests that failed a given task, there was a reflection on using low saturation levels or gray scale to represent incompleteness - even though mixing colors would not be the best idea, the authors believed that with minimal learning it should not be a problem. Concerns regarding help for the user were also considered, as individualizing chart legends against sharing some, or using labels and buttons to describe the current state or to revert some actions.

The functional prototype was subject of a few more formative evaluations, as discussion sessions and Heuristic Evaluations that lead to some changes regarding the interactivity, flexibility and consistency of the prototype, as well as data mapping and layout usage. The second Heuristic evaluation provided a better insight on how users interact and envision the final version of the design, and resulted in important feedback on particular elements that seemed to hinder their experience. Those violations were taken in consideration and the authors proceeded to further improve the design and implement such changes, resulting in one last phase of the prototype.

The layout of the design suffered some changes, resulting from the feedback received in the second Heuristic evaluation. In order to remove unnecessary blank space and to reduce scrolling for the users, the final phase of the design and respective layout is presented in Fig. 4. The layout was adjusted according to the different objectives across each visualization: on the left side, it is possible to see visualizations that select and prompt other visualizations with exploration (Fig. 4, views 1 and 4); next to them are shown visualizations more exploratory related purposes, while keeping visualizations from the same idiom close: Heatmap with Word cloud, and Scatter plot with Radar Chart.

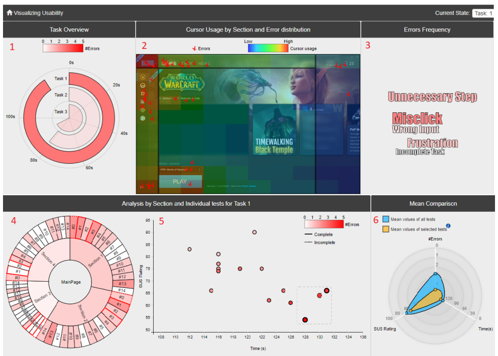

Figure 4: Final version of the functional prototype.

The final version of the functional prototype[1] shown in Fig. 4 had as concern not only providing interactivity between all idioms, but also keeping track of the current state for the users, one of the most recurrent problems within the focus group, due to the amount of data presented at a time. After selecting a specific task for analysis, users can observe the most cursor heavy section areas as well as the most error prone ones. Furthermore, by selecting a given section, information about the individual tests are updated for that context throughout the Dashboard. Analyzing a particular group of individual tests is possible, as well as comparing their values against the mean values of the whole population. Feedback given on the previous versions allowed additional filtering options that were made available for user's comfort while exploring the Dashboard.

5. Evaluation

This chapter intends to describe both evaluation methods utilized on the solution developed in Chapter 4, including results and insight that can be derived and concluded from this analysis.

5.1. Usability Tests

In order to assess the quality of the design and to validate the solution presented, a quantitative study through a Summative Evaluation process was done. Since an *User Centered Design* approach was used throughout this thesis, validating the solution near the target group allows the authors to assess its overall status and how far the solution is from being deployed and used daily.

Regarding the number of test users on Usability studies, Jakob Nielsen [15] defends that using five users almost always gets close to User Testing's maximum benefit-cost ratio, but in order to obtain enough data to develop a quantitative study and aim at statistical work, the authors decided to administer these Usability tests to a larger audience. Therefore, The authors tested with twenty users with advanced experience with *Information Visualization* and the concept of Usability, while not having any previous experience with the design. The higher number of participants, contrary to the numbers used on previous formative evaluations (five on average), allowed the authors to get statistically significant numbers, and to obtain confidence intervals on the performance metrics, as well as reaching correlations and conclusions on the matter.

After a short introduction to the main objective of the Visualization and a period five minutes of training and adaptation to the design, the authors presented test participants with the following set of tasks:

- Identify the Task with the highest average time and the Task with the highest average number of errors.
- Identify the most frequent error made for Task 1. How many occurrences are there?
- Which individual test had the highest number of errors for Task 3, Section 3? How many errors does it have?
- Identify any incomplete individual test for Task 1. How many errors were made by said test across all sections? Which ones(label)?
- Select the two most time-consuming individual tests for Task 1, Section 1. What are their calculated mean values?
- Select every individual test with SUS Rating values $\geq$ 65 and $\leq$ 70 for Task 3, Section 3. How many errors of type "Misclick" where made in this group? Which Test ID made a "Wrong Input" error?

The protocol used was equal for all test participants and both completion time and errors were written down for each of the tasks completed. An error was considered whenever the participant made an incorrect interpretation of a visualization purpose, or whenever he took a wrong path towards the goal. After the test was complete, users were given a *System Usability Scale* satisfaction questionnaire, a quick to administer and cheap tool to use for Usability measurement [16], in order to collect information the user's overall satisfaction while using the interface. Besides Satisfaction, the authors retrieved Effectiveness (represented by completion rates and number of errors) and Efficiency (represented by completion time). The following paragraphs present the results obtained during the usability tests performed and statistical derived work.

Regarding the completion rate of the tasks administered to the test participants, only one user failed to complete Tasks 5 and 6, while the other nineteen users managed to successfully complete every task. As for the time taken by participants to complete each task, the previous Table. 1 presents the results obtained.

[1] The final version of the Visualization can be obtained in: https://fmorgad0.github.io/visualizing-usability/

Table 1: Confidence intervals on time taken to complete each task.

Task no.	Low(s)	High(s)	Margin of error(s)	Confidence level
1	8,683	11,846	1,582	95%
2	10,677	15,826	2,574	95%
3	18,970	27,319	4,175	95%
4	46,159	63,564	8,702	95%
5	47,163	56,541	4,689	95%
6	63,054	80,815	8,880	95%

From this table can be derived that the time needed to complete each task increases simultaneously with each task, with the exception of the fifth, which may be justified due to the disparity in the margins of error encountered between the forth and the fifth tasks.

Table 2: Confidence intervals on number of errors for each task.

Task no.	Low	High	Margin of error	Confidence level
1	0,042	0,458	0,208	95%
2	0,311	0,789	0,239	95%
3	0,080	0,520	0,220	95%
4	0,267	0,833	0,283	95%
5	0,080	0,520	0,220	95%
6	0,393	1,007	0,307	95%

Table 2 presents information regarding the number of errors made by the users for each of the tasks handed. As it can be observed on the presented table, and again with a confidence level of 95%, that tasks 2, 4 and 6 are more error prone for the users, while not necessarily taking users the most time (from observation of Table 1). This situation might be justified by the complexity or difficulty while interacting with the specific visualizations associated with the given tasks.

Table 3: Confidence intervals on *System Usability Scale* score.

Low	High	Margin of error	Confidence level
69,528	77,722	4,097	95%

Satisfaction levels were also subject of evaluation near the test participants. In order to assess Satisfaction, the authors used a *System Usability Scale* satisfaction questionnaire, with a score between 0 and 100. The authors stress the fact that, according to Brooke [16] and Sauro [17], such scores are not to be interpreted as percentages, but in fact as a percentile. From the examination of Table 3, the design used in the Usability tests scores, with a confidence level of 95%, between 69,528 and 77,722. Thus, this result suggests that the design ranges above average score, with some design flaws and improvements needed but with an acceptable Satisfaction near the users.

The authors asked users to rate each task form a difficulty point of view. The objective was for the authors to perceive any particular disparity in the complexity of the tasks handed, as well as obtaining proof regarding the increasing level of difficulty with each task. With five different possible responses, users classified each task performed from 1 (hard) to 5 (easy). As shown in Table 4, and again using a confidence level of 95%, the tasks used in the Usability tests are considered progressively harder for users, with the exception once more of Task 5, which shows similar results to Task 4, while having a greater margin of error. Even though the tasks were presented in this order, the novelty of each one and the new interactions presented throughout the Usability test session displayed an increasing level of complexity for users to complete them. This information is somehow complemented by the increasing time users took to accomplish the same tasks (Table 1).

Table 4: Confidence intervals on difficulty levels for users in each task, ranging from 1 (hard) to 5 (easy).

Task no.	Low	High	Margin of error	Confidence level
1	4,691	5,000	0,209	95%
2	4,433	4,967	0,267	95%
3	4,075	4,625	0,275	95%
4	3,016	3,784	0,384	95%
5	2,842	3,758	0,458	95%
6	2,144	2,856	0,356	95%

After gathering proof from the participants on how the difficulty levels of each task are consequently harder, while also obtaining information on time and number of errors users had, there were some further analysis and questions regarding the results the authors wanted to corroborate. Thus, one of the questions that needed validation was verifying if users that took more time also made more errors.

		TIME	ERRORS
TIME	Pearson Correlation	1	,260**
	Sig. (2-tailed)		,004
	N	120	120
ERRORS	Pearson Correlation	,260**	1
	Sig. (2-tailed)	,004	
	N	120	120

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 5: Pearson's Correlation Coefficient value between time and number of errors.

As perceivable through the analysis of Fig. 5, there is in fact a positive correlation associated with the relationship between time taken and errors made, meaning it is observable a increasing number of errors while the users perform more time consuming tasks. The correlation is significant at the 0.01 level (2-tailed) and thus, with 99% confidence, the authors found a correlation of 0.260 between these two variables. Although a positive correlation was found, the small coefficient suggests a weak positive relationship (.20 to .29) between the time taken to complete a certain task and the number of errors made, leading to a weak dependency.

The authors believe that such results may be influenced by the learning pattern of tests users while completing the first tasks, easing the process once reaching the latter ones.

5.2. Case Studies

Three case studies were administered to people with extended knowledge and experience on both Usability and *Information Visualization* areas, and were made with the objective of obtaining additional insight on the viability of the visualization for practical use, as well as some suggestions regarding future possible functionalities to add. As participants work on related areas, the authors intend to substantiate the value of the developed solution for the target group and which purpose it serves regarding the daily tasks performed on the analysis of Usability metrics.

From the results obtained, the authors could infer some conclusions regarding the state of the solution and what future steps should be taken in order to continuously improve and to fulfill the target group's requirements. Although the participants involved in the Case Studies had positive remarks regarding the integration of the various functions across the Dashboard, there were some concerns with a few visualizations and some choices made for the mapping and representation of data. One participant revealed a difficulty in the identification incomplete tests in the design. Furthermore, the Word Cloud was particularly pinpointed from one participant as being sometimes confusing and potentially hard to differentiate between the different levels of error frequencies, and although the usage of color assisted the issue in hand, the color scale used could be further improved. The color gradient scale was not only mentioned in the previous problem, but throughout all Case Studies, and thus the authors acknowledge that as being one of the most important priorities for further work. The concerns expressed by the Case Study participants were also pinpointed by User Testing participants after their sessions, while some of which even triggered errors while performing the given tasks. This insight corroborates that both evaluation methods used can complement each other and highlight the most important changes for the next version of the functional prototype.

These studies served as well for identifying potential fields of interest for the usage of the presented solution. A Dashboard Visualization with the effort of improving the analysis of Usability Testing related metrics was viewed as a promising tool not only for educational purposes but also to aid users to assess the Usability of products developed in Research contexts. Additionally, the potential shown by the solution allowed participants to provide new possible features to be included, such as suggestion based functionalities by pinpointing the most concerning areas and important types of errors.

6. Conclusions and Future Work

The authors considered the main objective of this thesis achieved, as Information Visualization was applied in the analysis of Usability related metrics and the Visualization Dashboard developed was validated by target group users, who rated the solution with an above average Satisfaction, as shown in Chapter 5.

The results obtained from Usability Testing supported additional statistical studies. By revealing a weak dependency between the time taken to complete a given task and the number of errors made, the authors found that although more complex tasks consumed more time for users, those were not necessarily more error prone for them, which leads to the conclusion that the design provides a stable learning curve for new users. Case Studies represented a significant source of conclusions and insights regarding the state and utility of the presented solution. Participants were chosen based on their experience with both *Information Visualization* and Usability areas, and acknowledged the importance and utility of having such a tool to analyze Usability data, which some suggested being sometimes a complex task. Furthermore, they recognize the importance of the metrics that are being visualized throughout the Dashboard, although some concerns were shown regarding the overload of information shown at a time, which deviates from the main idea of using *Information Visualization* in the first place.

A plurality of utility areas and practical usages for the solution were suggested, and such tool was acknowledged by Case Study participants to be particularly useful in an educational context to help students identify flawed areas in their designs, as well as representing a supportive tool for researchers on the interpretation of Usability Testing results collected. Such areas were not initially thought of, and such testaments enhance, in the authors eyes, the contribution of this study.

6.1. Future Work

Results obtained from Usability Testing revealed some concerns with some of the visualizations regarding their scalability, particularly the Sunburst one, which may hinder the analysis for users while performing exploratory tasks. Additionally, other flaws reported by users expose flaws concerning data mapping, as the color scheme and representation for some metrics (number of errors and task completion). These problems contributed to deter results for some users, who struggled to complete some tasks, and in a particular case, even resulting in failure.

The group of visualization techniques utilized on the resulting Dashboard solution were subject of validation, and the quantitative study performed would benefit from further refinements and statistical analysis, such as comparing results with two different populations: one from the focus group and other from outside of it. Additional studies could be done on future iterations based on the Usability problems found on the present version of the solution, with the example of a comparative analysis regarding the performance of two or more design choices.

Finally, future work on this study includes the distribution of such tool, which helps users assessing the Usability of their design on different areas, such as educational and research, as well as testing with users on different contexts, while using real data.

References

[1] T. Munzner. *Visualization Analysis and Design*. AK Peters Visualization Series. CRC Press, 2015.

[2] Jakob Nielsen. Parallel & iterative design + competitive testing = high usability, January 2011. URL: https://www.nngroup.com/articles/parallel-and-iterative-design [Retrieved March, 2017].

[3] Jenny Preece, Yvonne Rogers, and Helen Sharp. *Beyond Interaction Design: Beyond Human-Computer Interaction*. John Wiley & Sons, Inc., New York, NY, USA, 2001.

[4] Jakob Nielsen. User satisfaction vs. performance metrics, October 2012. URL: https://www.nngroup.com/articles/satisfaction-vs-performance-metrics [Retrieved March, 2017].

[5] Stephanie Wilson, Mathilde Bekker, Hilary Johnson, and Peter Johnson. *Costs and Benefits of User Involvement in Design: Practitioners' Views*, pages 221–240. Springer London, London, 1996.

[6] Stephanie Wilson, Mathilde Bekker, Peter Johnson, and Hilary Johnson. Helping and hindering user involvement — a tale of everyday design. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, CHI '97, pages 178–185, New York, NY, USA, 1997. ACM.

[7] John W. Tukey. *Exploratory Data Analysis*. Addison-Wesley, 1977.

[8] Andrew Gelman and Antony Unwin. Infovis and statistical graphics: Different goals, different looks.

[9] Francisco G. de Borja and Carla M. D. S. Freitas. Civisanalysis: Interactive visualization for exploring roll call data and representatives' voting behaviour. In *SIBGRAPI*, pages 257–264. IEEE Computer Society, 2015.

[10] Kamasan Vaishnavi, Ashwin Kannan, David Cline, and Ronak Etemadpour. A visualization tool for learning statistical analysis in multi tabular datasets. *2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT)*, 00:222–226, 2016.

[11] Sandra P. Cano, Gloria Inés Alvarez, César A. Collazos, and Jaime Muñoz Arteaga. A visualization and human-computer interaction proposal in the context of pronunciation information. *IEEE-RITA*, 11(1):12–17, 2016.

[12] H. Chernoff. The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association*, (68):361–368, 1973.

[13] M. Margaux Price, J. Jessica Crumley-Branyon, R. William Leidheiser, and Richard Pak. Effects of information visualization on older adults' decision-making performance in a medicare plan selection task: A comparative usability study. *JMIR Hum Factors*, 3(1):e16, Jun 2016.

[14] Marjan Gusev, Sasko Ristov, and Dimitrij Mijoski. Enhancing the educache simulator with visualization of cache performance. In *2016 IEEE Global Engineering Education Conference, EDUCON 2016, Abu Dhabi, United Arab Emirates, April 10-13, 2016*, pages 375–382, 2016.

[15] Jakob Nielsen. How many test users in a usability study, June 2012. URL: https://www.nngroup.com/articles/how-many-test-users/ [Retrieved February, 2018].

[16] John Brooke. *"SUS-A quick and dirty usability scale." Usability evaluation in industry*. CRC Press, June 1996. ISBN: 9780748404605.

[17] Jeff Sauro. Measuring usability with the system usability scale (sus), February 2011. URL: https://measuringu.com/sus/ [Retrieved March, 2018].