UX-Analyzer: Visualizing Interaction Effort for Web Analytics

Juan Cruz Gardey
LIFIA, Fac. de Informática, Univ. Nac. de La Plata
La Plata, Buenos Aires, Argentina
jcgardey@lifia.info.unlp.edu.ar

Julián Grigera
LIFIA, Fac. de Informática, Univ. Nac. de La Plata
CICPBA
CONICET
Argentina
julian.grigera@lifia.info.unlp.edu.ar

Andrés Rodríguez
LIFIA, Fac. de Informática, Univ. Nac. de La Plata
La Plata, Buenos Aires, Argentina
andres.rodriguez@lifia.info.unlp.edu.ar

Alejandra Garrido
LIFIA, Fac. de Informática, Univ. Nac. de La Plata
CONICET
Argentina
garrido@lifia.info.unlp.edu.ar

ABSTRACT

Evaluation of User Experience (UX) in online systems is essential but also costly in terms of both time and money. We are interested in building automatic tools to assist development teams in evaluating the UX even with a limited budget and short iteration cycles.

Analyzing user interaction may provide a good indication of the effort that users are required to spend on certain page elements, which we consider an important factor of UX. In this paper we present UX-Analyzer, a web tool that automatically calculates and displays the user interaction effort of web pages. The interaction effort is a score that represents the extra burden, difficulty or discomfort of a user when interacting with a specific webpage element. This score has been learned from manual UX expert ratings and different measures of interaction logs, called micro-measures. UX-Analyzer further aggregates individual element scores from different users to compose a single score for each analyzed web page. The tool can also show the compared scores among alternative versions of a page.

Thus, UX-Analyzer allows UX experts or other team members to evaluate web pages automatically in a transparent way, since the interaction effort may be calculated with multiple users in their real context of use. Visualizing the interaction effort may provide a good indication towards the level of UX of an online system. It may also be used in the context of an A/B testing approach that instead of revenue or conversions, it compares the UX of alternative versions of a website. That is, this approach enables web analytics based on UX, i.e., measuring, analyzing and improving UX quality in each development cycle.

CCS CONCEPTS

• **Human-centered computing** → **Visualization systems and tools**; **HCI design and evaluation methods**.

SAC '24, April 8–12, 2024, Avila, Spain

ACM ISBN 979-8-4007-0243-3/24/04…$15.00
https://doi.org/10.1145/3605098.3636013

KEYWORDS

User Interaction, User Experience, UX refactoring, A/B testing, Machine Learning

ACM Reference Format:
Juan Cruz Gardey, Julián Grigera, Andrés Rodríguez, and Alejandra Garrido. 2024. UX-Analyzer: Visualizing Interaction Effort for Web Analytics. In *The 39th ACM/SIGAPP Symposium on Applied Computing (SAC '24), April 8–12, 2024, Avila, Spain.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3605098.3636013

1 INTRODUCTION

The standard definition of User Experience (UX) considers the hedonic factors that influence the user's emotions, pleasure and comfort, as well as the instrumental factors related to usability, interaction, etc. [11]. Given the relevance that UX has gained for the success of online systems [1, 17, 24]), large companies invest in frequent UX evaluation through user testing, interviews, surveys and expert inspections [22]. However, these methods are usually too costly for small and medium-sized companies; evaluations involving users are specially challenging to organize in the typical short iterations of a product's life-cycle, when there are not sufficient resources, while experts are not always available for frequent inspections. The result is that, for most online systems, UX is neglected after the initial design phase [13].

Our work aims at supporting companies with limited resources in improving UX during the entire product lifecycle, in the context of agile methods with frequent deliveries. Thus, its imperative to incorporate some automation in UX evaluation. Kohavi and Longbotham suggest that automatic controlled experiments like A/B testing are specially useful in the context of agile software development [12]. In A/B or split tests, the universe of users is randomly exposed to one of different variants of a system. To select the best alternative or variant, it is important to define a single metric, which is called Overall Evaluation Criterion (OEC). Typical metrics used are revenue, conversions, loyalty [12]. However, UX is rarely evaluated in the context of A/B testing [23].

We are specially interested in defining a metric that could be used to compare different designs in the automatic evaluation of UX. With that goal, in this paper we propose using the concept of "interaction effort" [7, 8]. The interaction effort has been defined

as a score that a UX expert assigns to the user's interaction with a particular element or widget of a web page. The important aspect is that it may be predicted from micro-measures that are automatically captured while a user interacts with the page [7]. While the interaction effort has been proposed to evaluate individual web page elements, our hypothesis in this work is that it may also be used to provide a "global picture" of the effort demanded by a complete user interface (UI) design. Thus, we propose aggregating the interaction effort of different users and widgets to compose a global effort score on a web page. Having a single effort score should be useful to easily assess and communicate a factor of the overall UX of a web page, and it also facilitates the comparison of alternative designs.

In this paper we present UX-Analyzer, a tool for calculating and visualizing the user interaction effort on alternative versions of a specific web page. The tool captures interaction logs from user sessions, which are fed into a set of prediction models to get the interaction effort on each widget contained on the target page. Then, the resulting scores are aggregated to compose the global effort of the UI under analysis. UX-Analyzer is intended for UX experts, who can use the tool to evaluate different designs with multiple users in online experiments like A/B testing or even with a few users to validate some design changes in relation with the interaction effort. The main benefit of this metric is that it is transparent to the users and easy to calculate, as the tool only requires embedding a script in the target page.

We also describe a case study of the tool in which we analyze the results provided by UX-Analyzer for five selected websites. Different examples are given as to what can be learned from these results, and we finally identify possible design changes that can help to reduce the global interaction effort of the target websites.

2 RELATED WORK

User experience is a key component for the success of an interactive product. The ISO standard defines UX as "user's perceptions and responses that result from the use and/or anticipated use of a system, product or service" [11]. According to this definition, UX is a broad concept that includes the instrumental factors related with user's performance while interacting with a system (effectiveness, efficiency), as well as the hedonic aspects that are subjective to the users like enjoyment, comfort and pleasure. In other words, UX is a consequence of the user internal state (expectations, needs, etc.), system's characteristics (complexity, functionality) and the context in which the interaction occurs [9].

Because of the importance that UX has gained, it is essential to include its evaluation and improvement in software development. However, besides the resources needed, this is especially challenging in agile teams that work in short development cycles focused on delivering new functionalities. For this reason, we are motivated to provide some automated solution to small and medium-size development teams, especially agile teams working under time pressure and scarce resources, to assess the UX of their products. In this regard, there are works that propose automating the assessment of different aspects of the UX. Many of these approaches are concerned with evaluating a UI without testing it with real users. For instance, Dou et al. and Michailidou et al. developed prediction models to get a measure of a website aesthetics [5, 18], while Li et al. tried to estimate the user performance of selecting items from vertical menus [16]. In contrast to these works, our aim is to account for user interaction to evaluate a UI, since the user interaction provides useful feedback that cannot be obtained through the static analysis of the target UI. The closest work to our approach is the one of Speicher et al. who developed a tool with machine learning models to predict on web pages seven usability aspects (confusion, distraction, readability, etc.) from user interaction logs [23]. However, the main limitation of this work is that the interaction logs used as features for the prediction are highly coupled to the structure of the target page, so new models have to be developed every time a different page has to be evaluated. To overcome this limitation, the interaction effort metric has been developed to judge the user interaction on the different widgets that are standard components present in every UI [7]. This metric was inspired in the Cognitive Load (CL) concept which is used to describe the user mental effort demanded by a system to interact with it [3]. Although part of this CL is intrinsic to the task at hand, a sub-optimal design can also cause an increment of the cognitive load [4]. In this way, the interaction effort provides an indicator of the interaction cost which is related with the physical and mental effort made by users to achieve their goals with a system [2].

Since the interaction effort evaluates individual widgets on a target UI, it would be beneficial to aggregate the effort of multiple widgets to compose a global effort score for a specific web page. Having a single score to assess the interaction effort of web pages allows to easily compare alternative designs in the context of online experiments such as A/B testing. Although A/B testing is a suitable approach to evaluate UX with a great mass of users, most of existing A/B testing tools are focused on measuring the revenue of a company with conversion rates, which are not a direct indicator of UX [19]. Instead, with the interaction effort the evaluation of alternative designs can be centered on UX, without having to consider the revenues.

Regarding the methods to get a measure of the UX in a single score, one of the best known works is that of Sauro and Kindlund [21], which combines the three usability factors (efficiency, effectiveness and satisfaction) in a single score. Although the method does not strictly state which measures to use to estimate each factor, the original work uses task-centered measures that cannot be easily calculated in a real context of use. Another established method for quantifying the UX involves the use of questionnaires [10]. With this method, participants must be recruited, exposed to the system under evaluation, after which they choose the suitable value for different statements within a value range. Some well known questionnaires are the User Experience Questionnaire (UEQ) [14], the Standardized User Experience Percentile Rank Questionnaire (SUPR-Q) [20] and UMUX-Lite [15]. The benefit of questionnaires is that they may reach a high level of accuracy in measuring the subjective attitude of the user towards the evaluated system. On the downside, they are costly because they require recruiting participants, and the results can be biased since the users know that they are under evaluation.

3 INTERACTION EFFORT FOR WEB PAGES

Interaction effort is a subjective score that judges a specific user interaction with a target UI widget [7]. UX experts observe a widget interaction and assign a score that ranges from 1 (effortless) to 4 (demanding), according to their subjective analysis. Given that the score has be manually assigned from a UX expert, different prediction models were developed to automatically calculate the effort score using micro-measures captured from user interaction logs. These micro-measures were specifically developed for this prediction task and they capture different user interaction aspects such as the time spent on a specific widget, mouse movements, keystrokes, among others.

The idea of analyzing individual UI widgets was proposed with the aim of evaluating small portions of a UI, motivated by the concept of UX refactorings, which are concrete UI transformations intended to solve UX smells, hints of poor user interaction [6]. Figure 1 shows two alternative refactorings for the UX Smell *Unformatted Input*, which signals that a free text input is used to require data that has a specific format. These refactorings offer different solutions for the same UX issue, so they should be evaluated to assess which one is better (in terms of UX) for a particular situation. Therefore, the interaction effort was first developed to compare the effectiveness of alternative refactorings. The metric is calculated on the specific widgets that are involved in the transformations performed by the refactorings.

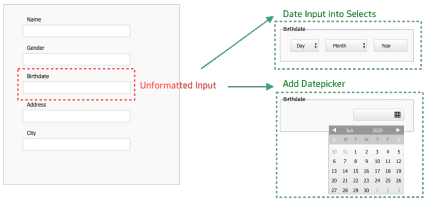

Figure 1: Alternative refactorings for the UX smell *Unformatted Input* [7].

Besides comparing alternative refactorings, having a fine-grained measure of a UI is also useful to precisely determine where users are struggling with it. But on the other hand, since the interaction effort judges a single user interaction with one widget, it could be hard to get an overall measure of how the target UI works when there are many widgets under analysis and multiple users interacting with them. In this way, as a complement to the interaction effort based on individual widgets, we propose aggregating all the interaction effort values to have a single score for assessing the user effort of a complete UI.

Figure 2 shows how the global interaction effort is calculated. Besides the global score, the interaction effort of each user with each widget can also be aggregated either by user or by widget. These aggregations can provide additional feedback about the user experience of a UI, for instance to identify specific users that make high effort or to detect certain widgets that are problematic for many users.

The tool UX-Analyzer described in the next section shows the global interaction effort of a web page, as well as the aggregations by user and by widget to provide a full picture of the user interaction effort.

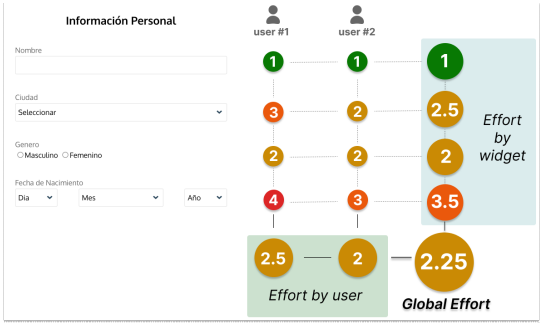

Figure 2: Interaction effort aggregated by users and by widgets.

4 UX-ANALYZER

UX-Analyzer is a tool to estimate and monitor the user interaction effort of a web application. The tool was created mainly for UX experts, who can analyze the resulting interaction effort and propose design changes in response to what they observe. Nevertheless, UX-Analyzer can also be used by others team members such managers or product owners, who may be interested in controlling the UX of a system.

UX-Anayzer is web application in which the users can sign up and create evaluations to observe different pages of a target application. An evaluation contains one or more versions of the observed pages on which the interaction effort is calculated. The tool also includes a JavaScript code that must be pasted in the target application to capture the user sessions of a specific version and send them to the server. Each user session contains the interaction logs that are fed into the prediction models to estimate the interaction effort of each widget in the analyzed page. These scores are then used to calculate the global effort of the version.

4.1 Evaluations

An evaluation groups a set of versions of the target application. Since the web applications can have multiple pages with a wide range of functionalities, the purpose of the evaluation is to separate the different user interfaces that need to be analyzed. The evaluation size will depend on what the user wants to evaluate. For instance, an evaluation could assess general aspects such as the checkout process on a e-commerce, as well as more detailed concerns like a specific set of fields in a form.

Figure 3 illustrates the evaluations list, the view shown after logging in UX-Analyzer. Each evaluation is identified by a name that describes the design aspect that is evaluating.

4.2 Versions

The application versions are design variants that belong to an evaluation. A version can also be defined as a set of user sessions collected on specific pages that are used to estimate the interaction effort. When a new version is created, UX-Analyzer generates a

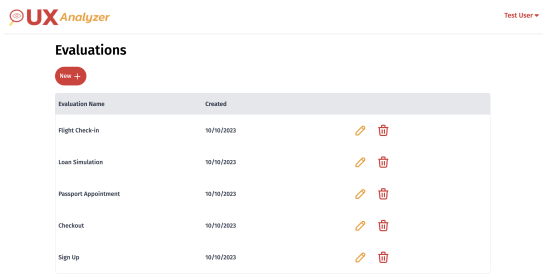

Figure 3: Interaction effort aggregated by users and by widgets.

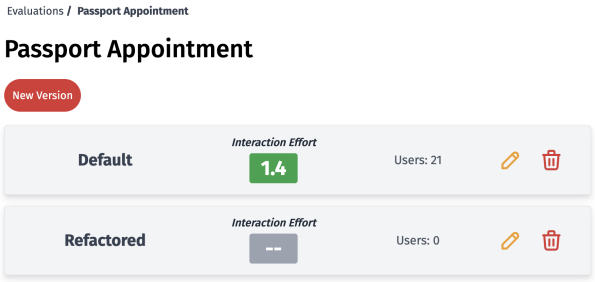

Figure 5: Versions list of an evaluation.

code snippet to include in the target application for capturing the user interactions with the widgets. This snippet contains a specific version token that is used by the backend of the tool to determine the target version of each user session received.

To create a new version, the user must provide a name and the URLs of the pages that will be included in it. Although the global interaction effort was thought for individual web pages, in some cases it may be necessary to know the effort score for a group of related pages. In this way, a version can contain one or more URLs, which then will be used to filter the user interaction logs on the selected pages when a new user session arrives to the server. Figure 4 shows a version recently created.

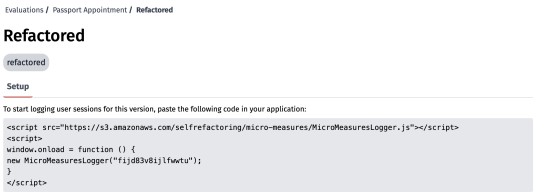

Figure 4: A new version created. The tool shows the code snippet that must be pasted in the target application.

The main view of the evaluation displays a list with all the generated versions (Figure 5). There, the user can observe for each one the global effort score and the amount of user sessions captured so far. This score is updated as new sessions are sent to the server. The user can also click on the versions to visualize the details of each one, like the user sessions and the widgets included in the target pages.

4.3 User sessions

The code snippet generated by UX-Analyzer when creating a version is used to capture the user interaction logs. These logs are basically the micro-measures that result from the user interaction with each widget under analysis. The interaction logging starts when a user navigates to a URL included in the version, and it ends when the page is abandoned. For each widget, besides the micro-measures the script also logs the URL and its XPath, which then together allow to identify that widget instance. When the logging ends, the session duration is calculated and all the information is sent to the backend. There, UX-Analyzer processes the logs and store them in the version identified by the received token.

Figure 6 illustrates how the captured user sessions look in UX-Analyzer. The most relevant attribute of each user session is the interaction effort that results from aggregating the predicted effort on all the widgets the user interacted with. This aggregated score is the *effort by user* shown in Figure 2. In addition to the effort score, the sessions also show their date and duration.

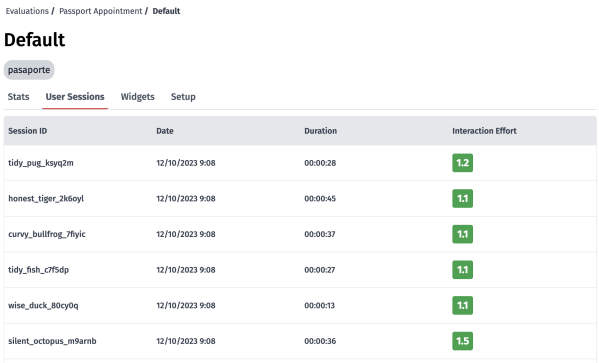

Figure 6: User sessions of a specific version.

This report is useful to monitor the users that interact with a particular version. By analyzing the effort score of each session and its duration, it is possible to detect particular cases in which a user experiments interaction problems.

4.4 Widgets

In the context of a version, the tool also allows to visualize the widgets list included in the target pages (Figure 7). Given that each widget is identified by its XPath and the URL, UX-Analyzer obtains this list grouping the interaction logs by these attributes. Each widget contains the averaged interaction effort (*effort by widget* in Figure 2), and a label along with its type (text input, link, etc.) to recognize it. Moreover, the widgets have an associated weight that determines their influence in the global effort score of the underlying version. According the needs, the UX-Analyzer user can adjust the weights to give more importance to specific widgets and these changes are reflected in the global score. In this way, a UX expert can center the analysis on the widgets that are crucial for

Table 1: Interaction effort score of the five analyzed websites.

Web	Effort
Flight check-in	1.34
Passport appointment	1.25
E-shop checkout	1.25
Sign-up in a ticketing e-shop	1.26
Loan simulation	1.09

the success of the target website, and leave those less important widgets in a second place.

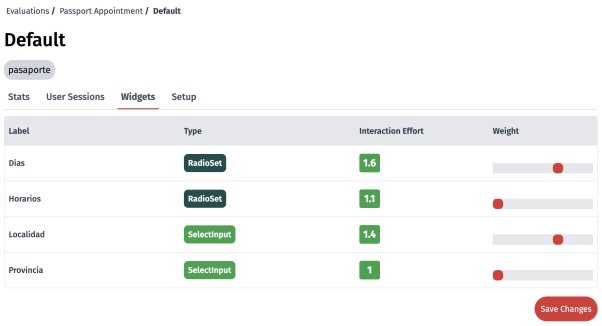

Figure 7: List of widgets that users interacted with using a specific version.

Like the effort of user sessions, the interaction effort aggregated by widget is also important to understand the value of the version global effort. In particular, this aggregation allows to detect the UI elements that users are struggling with.

5 CASE STUDY

In order to assess the applicability of the tool, we used UX-Analyzer to calculate the user interaction effort of five real websites. In these websites, 23 recruited users completed a specific task and the resulting user sessions were saved in the tool for the subsequent processing. The selected websites included common tasks performed on a daily basis like completing a flight check-in, finishing a purchase order on an e-shop and booking an appointment to get the passport. Regarding the participants, they were of different ages (MIN=22, MAX=49, SD=9.1) and the had different occupations.

Table 1 lists the five websites with the interaction effort calculated with UX-Analyzer. For each website under analysis we created a separated evaluation with a single version, and each user test performed by a participant on each website was mapped to a different user session (see Figure 3).

Regarding the first website of Table 1, the one with the greatest global effort, the score of each user session ranges from 1.2 to 1.5, so the interaction effort for each participant in this website is similar. However, analyzing the effort score aggregated by each interactive widget, we observed that those widgets for entering dates (date selects) were the ones that demanded the greatest effort, with a score of 1.9. Concerning the other widgets types (text inputs and radio buttons), the average effort score was 1.1. In order to assess the influence of the date select widgets in the global effort score of the version, we changed the settings to give a higher weight to the selects (3 times the weight of the other widgets) and the global effort increased to 1. That indicates that date selects play an important role in the interaction effort of this website. In this regard, one way to lower the interaction effort would be to evaluate alternative widgets for entering a date like date pickers and masked text inputs.

With respect to the three websites that have similar effort scores close to 1.2, the website to get the passport appointment is the one that has more variability in the interaction effort aggregated by widgets. The ones with the greatest effort were the select box for entering the city (1.4) and the radio buttons set to choose the appointment date (1.6), while the remaining widgets have a score of 1. In this way, we can increase the global effort score up to 1.5 if more importance is assigned to the city and appointment date widgets. A key aspect of these widgets is that they have many selectable options, so the high effort may be due to the fact that a user has to read each option to choose the desired one. Therefore, using other widget types to avoid having to choose from multiple options may be an opportunity for improving the overall interaction effort.

Concerning the website with the lowest interaction effort score, all the widgets have an average score of 1.1, so we could not identify any problematic element.

6 DISCUSSION

In the previous section we described a case study of UX-Analyzer to visualize the user interaction effort score of different websites, and we gave examples of the type of conclusions that we expect to draw from the tool to improve the score. By collecting interaction data from real users, we could observe in UX-Analyzer the global interaction effort of five selected websites and to identify the source of that overall score.

Concerning the reports provided by UX-Analyzer, it gives the possibility to observe the interaction effort aggregated by each interactive widget of the website under analysis. This information is useful to identify the individual elements that can cause problems to the users, and to analyze design alternatives for them. Moreover, the tool also allows to adjust the importance of each widget and inspect in real time how that is reflected in the overall effort score. Besides giving the user the chance to freely modify the widget weights, we are also studying alternatives to adjust them automatically using the data collected for the user session.

Another report available in the tool is the average interaction effort that results from each user session logged. Although this reports aims at proving information of each particular user, during the previous use case we realized that more context is needed besides the session time to understand and draw conclusions from interaction effort at the user session level. In this regard, we plan to include more details about each user session which includes a screen recording of it, so the user of UX-Analyzer can observe specific cases in which a user of the analyzed website is struggling with the user interface.

One important aspect to mention is that the analysis performed in the previous case study bay be limited due to the fact that effort scores obtained in the previous evaluation were very similar, being

all of them close to 1. We believe that this is related with the design of the task performed, which although they are real and known for the users, they were reduced to be done on a single web page for technical reasons, so in some cases they were carried out of their real context. A concrete example of this context is that the user has to buy a ticket before being able to check-in for a flight. Probably, performing the task in its real context adds more complexity and this is reflected in a greater interaction effort, which in turn could give more insights about the problems that users may experience using the websites. Besides that, including more users and the opinion of UX experts in the evaluation will be essential to validate the feasibility of UX-Analyzer for the user experience improvement.

7 CONCLUSIONS AND FUTURE WORK

In this work we showed UX-Analyzer, a web tool to evaluate web pages with respect to the user interaction effort. The tool provides different perspectives of the user interaction effort, giving the chance to visualize a single effort score for each analyzed version, as well as the effort of each user session and the average effort demanded by each specific widget. We showed a case study of the tool in which we analyzed the resulting effort scores and we mentioned possible improvements for the websites under analysis.

We are planning to evaluate the tool in real context of use, with multiple users and UX experts to assess the feasibility of UX-Analyzer to identify issues experienced by the end users and possible solutions for them. Moreover, we also plan to run evaluations to determine the tool applicability for comparing design alternatives of the same UI. This will allow us to validate whether a single interaction effort score can be used as a metric to compare the performance of design variations, for instance in an A/B testing approach.

Regarding UX-Analyzer, we foresee different improvements for the tool. The most important one is that UX-Analyzer not only aims to show interaction effort, but we plan to incorporate new metrics to evaluate other aspects of UX. In this way, the tool could be used to inspect and monitor different metrics that help UX experts or other interested team member to keep the UX under control.

ACKNOWLEDGMENTS

The authors wish to acknowledge the support from the Argentinian National Agency for Scientific and Technical Promotion (ANPCyT), grant number PICT-2019-02485.

REFERENCES

[1] Omar Badran and Shafig Al-Haddad. 2018. The impact of software user experience on customer satisfaction. *Journal of Management Information and Decision Sciences* 21, 1 (2018), 1–20.

[2] Raluca Budiu. 2013. Interaction Cost. https://www.nngroup.com/articles/interaction-cost-definition/. Date accessed: 12/28/2023.

[3] Fang Chen, Natalie Ruiz, Eric Choi, Julien Epps, M. Asif Khawaja, Ronnie Taib, Bo Yin, and Yang Wang. 2012. Multimodal behavior and interaction as indicators of cognitive load. *ACM Transactions on Interactive Intelligent Systems* 2, 4 (2012), 1–36. https://doi.org/10.1145/2395123.2395127

[4] Fang Chen, Jianlong Zhou, Yang Wang, Kun Yu, Syed Z. Arshad, Ahmad Khawaji, and Dan Conway. 2016. Theoretical Aspects of Multimodal Cognitive Load Measures. In *Robust Multimodal Cognitive Load Measurement*. Springer International Publishing, Cham, 33–71. https://doi.org/10.1007/978-3-319-31700-7_3

[5] Qi Dou, Xianjun Sam Zheng, Tongfang Sun, and Pheng-Ann Heng. 2019. Webthetics: Quantifying webpage aesthetics with deep learning. *International Journal of Human-Computer Studies* 124 (2019), 56–66. https://doi.org/10.1016/j.ijhcs.2018.11.006

[6] Juan Cruz Gardey and Alejandra Garrido. 2020. User Experience Evaluation through Automatic A/B Testing. In *Proceedings of the 25th International Conference on Intelligent User Interfaces Companion* (Cagliari, Italy) *(IUI '20)*. Association for Computing Machinery, New York, NY, USA, 25–26. https://doi.org/10.1145/3379336.3381514

[7] Juan Cruz Gardey, Julián Grigera, Andrés Rodríguez, Gustavo Rossi, and Alejandra Garrido. 2022. Predicting interaction effort in web interface widgets. *International Journal of Human-Computer Studies* 168 (2022), 102919. https://doi.org/10.1016/j.ijhcs.2022.102919

[8] Julián Grigera, Juan Cruz Gardey, Andres Rodriguez, Alejandra Garrido, and Gustavo Rossi. 2019. One metric for all: Calculating interaction effort of individual widgets. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–6.

[9] Marc Hassenzahl and Noam Tractinsky. 2006. User experience-a research agenda. *Behaviour & information technology* 25, 2 (2006), 91–97.

[10] Andreas Hinderks, Dominique Winter, Martin Schrepp, and Jörg Thomaschewski. 2019. Applicability of user experience and usability questionnaires. *Journal of Universal Computer Science, 25 (13), 1717-1735.* 25, 13 (2019), 1717–1735.

[11] ISO. 2019. *ISO 9241-210:2019 - Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems.* ISO/TC 159/SC 4. https://www.iso.org/standard/77520.html

[12] Ron Kohavi and Roger Longbotham. 2017. Online Controlled Experiments and A/B Testing. *Encyclopedia of machine learning and data mining* 7, 8 (2017), 922–929.

[13] Marta Kristin Larusdottir, Lene Nielsen, Anders Bruun, Lars Bo Larsen, Peter Axel Nielsen, and John Stouby Persson. 2018. UX in Agile before and during development. In *Proceedings of the 10th Nordic Conference on Human-Computer Interaction*. 984–987.

[14] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and usability engineering group*. Springer, Springer Berlin Heidelberg, Berlin, Heidelberg, 63–76.

[15] James R Lewis, Brian S Utesch, and Deborah E Maher. 2013. UMUX-LITE: when there's no time for the SUS. In *Proceedings of the SIGCHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 2099–2102.

[16] Yang Li, Samy Bengio, and Gilles Bailly. 2018. Predicting human performance in vertical menu selection using deep learning. In *Conference on Human Factors in Computing Systems - Proceedings*. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3173574.3173603

[17] Laura Luther, Victor Tiberius, and Alexander Brem. 2020. User Experience (UX) in business, management, and psychology: A bibliometric mapping of the current state of research. *Multimodal Technologies and Interaction* 4, 2 (2020), 18.

[18] Eleni Michailidou, Sukru Eraslan, Yeliz Yesilada, and Simon Harper. 2021. Automated prediction of visual complexity of web pages: Tools and evaluations. *International Journal of Human-Computer Studies* 145 (2021), 102523. https://doi.org/10.1016/j.ijhcs.2020.102523

[19] Jakob Nielsen. 2005. Putting A/B Testing in Its Place. https://www.nngroup.com/articles/putting-ab-testing-in-its-place/.

[20] Jeff Sauro. 2015. SUPR-Q: A comprehensive measure of the quality of the website user experience. *Journal of usability studies* 10, 2 (2015), 68–86.

[21] Jeff Sauro and Erika Kindlund. 2005. A Method to Standardize Usability Metrics into a Single Score. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Portland, Oregon, USA) *(CHI '05)*. Association for Computing Machinery, New York, NY, USA, 401–409. https://doi.org/10.1145/1054972.1055028

[22] Jeff Sauro and James R Lewis. 2016. *Quantifying the user experience: Practical statistics for user research*. Morgan Kaufmann.

[23] Maximilian Speicher, Andreas Both, and Martin Gaedke. 2014. Ensuring Web Interface Quality through Usability-Based Split Testing. In *Web Engineering*. Springer International Publishing, Cham, 93–110. https://doi.org/10.1007/978-3-319-08245-5_6

[24] Norhanisha Yusof, Nor Laily Hashim, and Azham Hussain. 2022. A Conceptual User Experience Evaluation Model on Online Systems. *International Journal of Advanced Computer Science and Applications* 13, 1 (2022), 1–11.