Annotation in Digital Humanities

Federico Caria and Brigitte Mathiak

Abstract In previous studies on user behavior with Digital Scholarly Editions (DSE), we found that annotating the text is a key technique for working with the text. In this follow-up study, we invited volunteers to perform open research tasks on a DSE by Lope de Vega, providing the annotation tool hypothes.is to support their tasks. What we found is that none of the participants used the tool extensively, yet, it was clear that annotation of text was a major part of their workflow. During a focus discussion after the experiment, the participants gave examples of how they would compensate for a lack of appropriate tools, often at the cost of considerable extra work and overhead. More research is needed to discover why the users were not accepting the tools provided. So, we propose a human-centered structured longitudinal approach to design an annotation tool that would actually be usable.

Keywords Annotation · Digital editions · User study · Usability · Research behavior

1 Introduction

During the three-year period 2014–2017, we explored the usefulness of digital editions in a typical research context, where users interact with digital collections of textual artifacts for research purposes. Our method has been adapted to the evaluation of these environments using best practices in user-centered design [1]. It is based on the capture of mixed data through several in lab testing sessions with a limited number of testers from various universities and humanities departments.

F. Caria · B. Mathiak (✉)
Cologne University, 50923 Albertus-Magnus-Platz, Cologne, Germany
e-mail: mathiak@gmail.com

F. Caria
e-mail: federico.caria@live.it

1.1 Functionality for Digital Editions

In 2000, John Unsworth introduced a list of seven 'scholarly primitives' which are 'self-understood' functions forming the basis for 'higher-level scholarly projects, arguments, statements and interpretations' [2]. His list summarized activities that were 'basic to scholarship across eras and across media', and went on to say that an analysis of these scholarly primitives might result in a clearer sense of how computing tools could support the scholarly endeavour. He also pointed out that comparison and annotation were the most important, yet poorly supported in digital scholarship projects. In valuable user studies on Digital Scholarly Editions (DSEs) we had similar results: the key techniques for working with the text are search, comparison and annotation. However, of these three, only search and comparison have a sound technological basis.

Although Mosaic, the first graphical Web browser, was intended to permit annotation, good tools for marking up what you read or hear, or view online have been slow to develop, observes Schacht [3]. Today, after seventeen years of digital scholarship, annotation tools are available in abundance, yet we have to wonder why most people still prefer to use papers and pens for annotating [4]. Although there are sophisticated computer-based annotation systems available, there is no prevalent service, rather people tend to misuse those, using bookmarks, writing emails, posting forums or creating wiki pages [5]. Moreover, most annotation tools are aimed at implementing collaborative reading or public annotation [6], while personal annotation (what we believe is the necessity expressed by our testers), has played a relatively minor role in the history of computer-based annotation systems.

Annotation in digital editions. In order to better understand how annotation can be integrated into digital editions, we asked volunteers to perform open tasks on a digital edition, which had installed an easy-to-use web annotation tool, hypothes.is. Twelve volunteers were invited to the CAYCIT Buenos Aires lab to sit in front of their laptop and perform open research tasks on a digital edition of Lope de Vega, using the annotation tool to support their task performance.

2 Related Literature

2.1 Annotation Is a High-Reward Technique for Professional Readers

Cognitive science highlights the role of annotation as an information-processing aid. Previous literature suggests that the strenuous attentiveness required by close-reading particularly challenges the information-processing ability of the individual [7], who will find that annotation improves the appropriation of a given text. In this context, annotation is better described by the notion of external cognition, a phrase that refers to ways that people augment their normal cognitive processes with external

aids, such as external notes, visualizations, and work spaces. "External cognition is human or cognitive information processing that combines internal cognition with perception and manipulation of external representations of information."[1] Previous literature suggests a relationship between the ability of note-taking and academic achievement, and one's ability to hold and manipulate propositional information in working memory.

Ubiquitous annotation in the history of the book. All those who have some familiarity with the history of the book know that annotation is fundamental for scholarship. Jardine's [8] study of the reading habits of Gabriel Harvey sheds light on the early press era, observing that critical reading was already "synonym of skillful annotation". According to all the advice books, reading for study or to acquire information or knowledge was supposed to include note-taking [9]. "Otiosa, vana, nugatoria est lectio, cui nulla miscetur scriptio" says a famous manual on note-taking dating back to 1638 [10]. Annotations are a key component of manuscripts as well, as anybody who has had the opportunity to examine them can testify. Notably, the thirteenth century gave birth to two particularly intelligent book designs that accommodated precisely annotation. Both types were connected to the emerging universities, which makes sense as this has been a note-taking environment par excellence—both then and now [11].The massive use of annotation makes us consider a direct link between text, annotation and information processing regardless of the media. In essence, it seems that a professional reader—whose reading activity is always goal oriented—produces annotations by necessity. This seems to reinforce the link with annotation as an information-processing aid.

Effectiveness of different materialities. As far as we are concerned, there are interesting studies, which go into the details of the physical movements and the materials that might determine a more or less successful learning performance. Many of these studies arise from the introduction of tablets in everyday life, especially in education, where stakeholders have felt, more than others, the need of evaluating the impact of new technologies, writing and reading habits on the student's performance. In this context, we are particularly interested in those who have analyzed the effectiveness of note-taking of typing compared to handwriting, discovering that the latter is more effective, due to the greater slowness and deliberation with which the act of external cognition takes place [12]. Curiously, the lesser ease of handwriting forces the annotator to economize and summarize, which means extracting main ideas from environmental information, summarizing, paraphrasing—contrary to typing on a keyboard, where people tend to write verbatim due to the quickness of the operation. The result of the former is a much better appropriation of the content.

---

[1]Glossary of Sensemaking Terms, Parc, 2007 (MET).

2.2   Annotation from Print to Digital

We have seen that there is a deep link between annotation, professional reading, academic achievement, and knowledge construction and transmission, to which Vannevar Bush, in theorizing the Memex, alludes. Marshall [8] reminds us of the annotative principle that characterizes the Web, from Bush to Ted Nelson.

Annotative principle behind the Web. Bush's Memo machine focuses on annotation through trail-blazing [9]; Xanadu takes an approach in which new hypertext seamlessly assimilates portions of older writings [13], etc. The research of the pioneers are all motivated by the objective of elaborating data structures closer to the functioning of the brain. Bush, in particular, discusses the potential for greater efficiency of association (one with annotation) compared to current indexing methods, emphasizing once again, the deeper link between the act of annotation and information processing that we have discussed briefly.

Digital annotation tools, collaborative versus personal annotation. Regarding the link between annotation and computing, we found two main research strands, the first one is closer to our topic, and it is found in the work of Marshall [6, 14, 15], the second is related to the W3C and the focus on semantic annotation. Throughout history, many tools have been developed, so that a systematic review of each system would require a separate paper. It is enough to say that their' functions are as varied as annotations can be, e.g. annotations as mnemonic, interlinking, highlighting important parts, commenting for understanding, etc. [5] Another important thing to note is that annotation tools for collaborative reading and writing outnumber those for personal annotation. The difference between personal and public annotation is described by Marshall, who in the late Nineties devoted several papers to the issue. The author analysed 410 marked-up textbooks from a university bookstore. The examination of these textbooks resulted in the distinction of several properties, describing several dimensions of annotations: formal or informal, explicit or tacit, permanent or transient, and published or private.

Private/personal annotation. While we might think of collaborative or public annotations as subsuming personal or individual annotations, especially from an architectural or system design perspective, the practice that leads to their creation is quite different, as observes Marshall. We shall focus on annotation as a personal device, one that plays into reading as a visible trace of human attention on the page, to gain an insight into the typical work of the scholar, who might not necessarily want to share annotations with others. These markings are not always intelligible by third parties, telegraphic, incomplete, and tacit. A highlighted sentence, a cryptic marginal "No!", an unexplained link, a reading history, or a bookmark, all pose interpretive difficulties for anyone other than the original annotator (and the passing of time sometimes erodes that privilege) [15]. These semi-intelligible signs are the traces (Kirsch's communication device) [16], of the inner dialogue of a reader (with the self) better known as close-reading, which very much typifies the type of scholarship that we are dealing with.

3 Design of the Experiment

In the following sections, we will give an overview of the general experimental design, the setting, the editions we studied, and why we selected them, and finally, the demographics of our focus group participants.

3.1 Hybrid Focus Group

In a focus group, a small group of test subjects discusses their experience with the product and shares their opinion, beliefs, and attitudes, while the moderator keeps the discussion on track. In our hybrid approach, we also give the participants questionnaires before and after the tests, to collect both qualitative and quantitative data. The experiment's design is inspired by Nielsen's discount usability [17] and guerrilla techniques for what concerns the recruiting of a limited number of participants, and the low time/cost of realisation; it can be potentially replicated more or less in any research context. Tasks are designed to let the users explore the media, retrieve content, compare records and interrelate information. The task scenarios are meant to reproduce a goal-oriented context of interaction and are left open.

3.2 Setting

The experimental setting consists of a usability lab, personal computers equipped with an open source tool to screen-record the performance of the tasks, an audio recorder to capture the final debriefing, paper and pen. We gathered 12 participants in a usability lab, asked them to perform research tasks in a given amount of time and to give their feedback. An example of a task was to retrieve various kinds of information and compare the records. The feedback collected was of different kinds: (a) a usability questionnaire, filled for each edition after completing a series of tasks; and (b) an audio-taped discussion the focus group. These data were coded together with demographics, the screen capture of the performance of the tasks, and the participants' answers to the tasks, to provide insight into design, usability issues and behavioral information.

3.3 Digital Edition and Annotation Tool

We tested one digital edition (1) *La Dama Boba,* Marco Presotto (ed.) 20xx, which had the Web annotation tool (2) hypothes.is installed. After surveying the online catalogues of digital editions curated by Sahle (v 3.0, 2008–2011) and Franzini
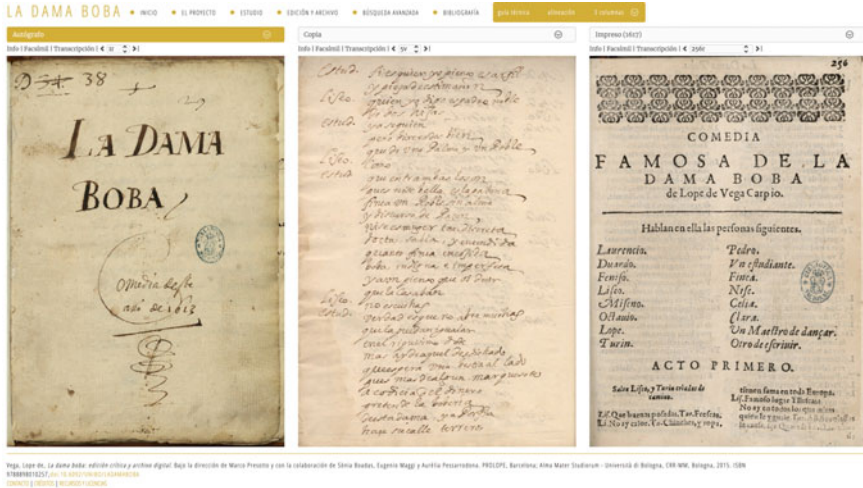
Fig. 1 Main interface of La Dama Boba

(2015), we used a resource we believed was more appropriate to the background of our MA and Ph.D. testers from the University of Buenos Aires.

La Dama Boba http://damaboba.unibo.it. A project developed by Dott. Marco Presotto from Bologna University. The edition reconstructs the text, analyzing the tradition, that according to the editor represents one of the most interesting and significant ecdotic problems of Golden Age theater literature.The editor uses the main witnesses to suggest a reconstruction of the text based on the analysis of the textual transmission of particular specimens. The functions of the site allow you to compare the different witnesses, align the texts, admire the original writing, do research by keyword (Fig. 1).

Hypothes hypothes.is. Open-source annotation tool developed by a non-profit organization funded by Knight, Mellon, Shuttleworth, Sloan, Helmsley, and Omidyar Foundations. The tool enables sentence-level note taking or critique on news, blogs, scientific articles, books, terms of service, ballot initiatives, legislation and more. We found this tool to be one of the easiest available to use. Above all, installing hyphes.is is straightforward, which given the experimental setting, was necessary (Fig. 2).

3.4 Participants

Our testers ranged from 20 to 50 years old and were all enrolled in humanities programs, the majority in the second year of their Ph.D., and two of them were teachers. Their computer skills were medium-high to high, as judged from their self-evaluations on a scale from 1 to 7 (Table 1).

Fig. 2 A screenshot of Hypothes.is annotation tool interface on la Dama Boba

Table 1 Demographics

ID	Department	Level	Web sills
p1	Philology	Ph.D.	6
p2	Philology	Ph.D.	6
p3	Literature	MA	7
p4	Literature	Ph.D.	6
p5	History	Ph.D.	5
p6	History	Ph.D.	5
p7	Literature	MA	4
p8	Literature	Ph.D.	6
p9	History	Ph.D.	5
p10	Philology	MA	6

3.5 The Tasks

The tasks were left open and were designed to push testers to put together several pieces of information, located in different parts of the site. This involved opening and closing various windows, and going through more than one screen—with the result that the information was difficult to record. On the sheets that we normally provided to testers before starting the task exercise, we asked them to respond at the end, that is, to use the time for retrieving the needed information only, since additional time was provided to write down the answers. An example of task is:

Example. There is a significant number of interventions in the manuscript—as in any Lope's autograph—that must be ascribed to the author. These are words or phrases crossed out and replaced by others or reorganized on the flow. Lope's

tendency to what critics defined *intervention in itinere* is well known, also in the case of La Dama Boba, it can be said that the work, or a good part of it, was created directly on the paper that has been preserved for us. Identify at least two examples of this phenomenon in the text. Indicate only alternative examples to those described by the editor. Then, compare at least two of the interventions identified by you with the copy and the printed text and indicate if, and where, there are changes when passing from one witness to another.

4   Findings

This time, the data collected were related to the edition and the annotation tool. The data on the use of the edition are secondary to this paper, but we decided to collect them anyway in order to give feedback to the editor, who was very kind to answer all our questions about the project.

4.1   Satisfaction Questionnaire

Participants were asked to fill in a satisfaction questionnaire for each DSE after completing the tasks. The instrument is adapted from WAMMI (http://www.wammi. com), and Koohang [18], focusing on the need to capture participants' immediate felt experience right after performing the tasks. The main results of the study, as far as we are concerned, are that (a) 100% of our testers finds annotation useful (b) 100% of our testers annotate in their research activity; (c) 50% of testers were not satisfied with how the tool allowed them to access and manipulate their notes (Table 2).

4.2   Suggested Features

In the questionnaire is a section where participants were asked to comment on the functions they thought were the best and to suggest improvements. We went to look at their responses in order to see if any of those who felt unhappy with the annotation tool wrote more specific complains, but we didn't find relevant information on annotation.

4.3   Screen Capture

We went to see the screen capture in order to understand how the testers, at what time in the workflow, had used the annotation tool. We saw that most of them, after having installed and tried it, had just opened it during the reading of the essay, often

Table 2 Some data from the satisfaction questionnaire

Questions	Agree	Disagree
I'd like my workflow to be more supported	xxxxxxxxxx	xx
This website is a precious resource	xxxxxxxxxxxx	
I can quickly find what I want on this website	xxxxx	xxxxxxx
Annotation is useful	xxxxxxxxxxx	x
I rarely take notes when I work with texts		xxxxxxxxxxxx
Synoptic comparison is what I do in my job with texts	xxxxxx	xxxxxx
This website has some annoying features	xxxxxxxx	xxxx
I am happy about how the system gives access to my notes	x	xxxxxxxxxxx

not more than twice. 3/12 only tested it once after installing it, 7/12 opened it during the essay and took up three annotations, only two tried out the instrument during all the tasks, mostly in underlining.

Highlight, comment, tag. The tool gives the ability to mark portions of texts in three ways: highlight, comment and tag. Most users highlighted, comments and tags were very little used.

4.4 Final Discussion

The final group discussions of the panel were audio-taped, transcribed and coded. Their aim was to expand issues relating to note-taking habits, based on a set of open questions, where participants were asked to draw on their experience. The discussion was structured according to the following lineup and then consequently coded.

Confidence with digital or web-based research. We always ask for more information regarding the use of print resources and the web, in order to warm the group up and introduce it to the central topics. This time, the major trends were: (1) advanced use of the Web, at least where possible; (2) predominant use of the print books, libraries, and/or traditional paper resources; (3) lack of digital resources related to authors or themes studied by the participants; (4) general preference for pdf, (or journal's articles to download) when it comes to using a digital resource (Fig. 3).

Confidence with tasks/usefulness of Lope de Vega. In this way we tried to understand if the tasks were suitable for the sample's background, engaging in a
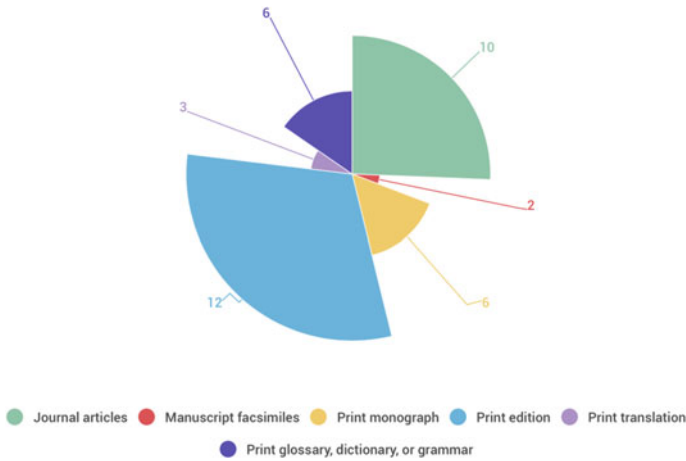
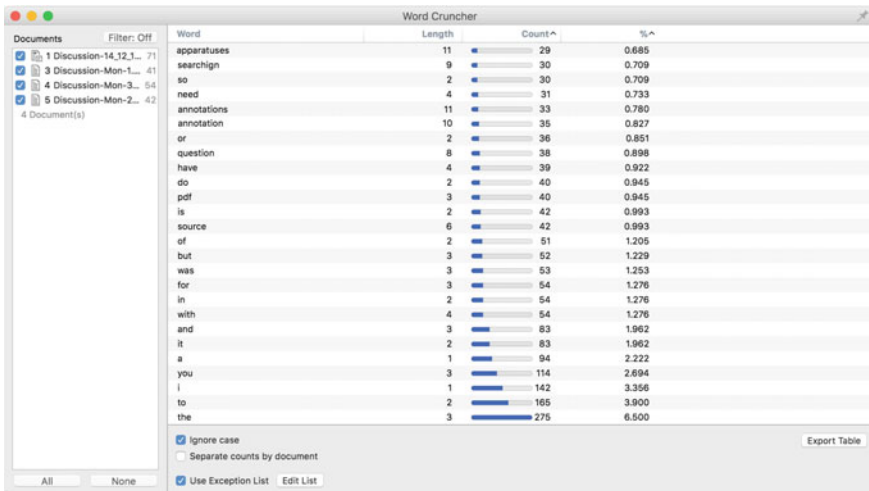**Fig. 3** A chart showing the most used resources by our testers

Fig. 4 A screenshot from Atlas.ti showing the most frequent terms in the final discussion

quick discussion about the edition and the satisfaction with the exercises. Testers were all experts—we have no BA student, this time—belonging to departments of literature, philology and history. They said they were confident with text comparison tasks, although we expected the opposite. Analysis of the discussion shows that the two words most frequently pronounced by the participants (excluding the semantic field of the annotation, which we come to later), are search, and apparatus (Fig. 4).

4.5 Annotation

"What about hypothes.is? did you use it?" we asked at this point. We were about to discuss the most problematic result of the experiment, as most of the users did not used it at all. "I opened it, but I didn't take any notes" was the most common answer. Thus we tried to investigate further, asking if there were particular reasons or usability barriers that discouraged them from using it, but we did not get any useful insight. In general, our testers were not able to articulate detailed explanations of why they did not use the tool, let alone able to say why it did not satisfy them. We have not received sensible answers on this subject that go beyond various comments, like not being accustomed to taking notes online, etc. "I highlighted a bit, then I was not able to find my highlights" answered one of the participants (P3). "I did not find it useful here", (P10) said another. Another has claimed to have used it to look at the underlining of others, and then passed on to the exercises (P11), etc.

Annotation in user workflow. Those who used hypthes.is more frequently did it in the exercises that involved the reading of the introductory essay (P5): in general the essay was much more annotated that the texts, although most of the exercises required to navigating the texts themselves, using the search engine, and expanding on information derived from the essay. Highlights were the most common type of annotation in the essay.

Annotation techniques on print. Testers declared marking extensively, as we discovered in our previous study. Also in this case we recorded comments regarding the felt experience of a better learning performance. "I always note," said one, "my books are full of notes," said another. "Without writing down I seem to read for nothing", said another, etc. Highlighting, underlining, paraphrasing, summarizing, or using simple keywords whose function is to link to other resources are the most frequent types of marks made by our testers. Among the stratagems to visually organize the notes is the use of different colors.

Annotation techniques on digital. The majority of users said they do not annotate when working on digital. A small number of testers, precisely 4, said they can not do without annotating, and in the absence of tools use a notebook, where they manually report the references from screen to paper, often using screenshots organized in a folder. Curiously, two of these four, before the session began, sat in front of the laptop with a notepad and a pen next to it.

5   Observations

The study confirms that users need annotation tools, which seems to further validate including annotation in the edition's functionality. Regarding the fact that users did not use the tool we provided, nor were able to explain why they did so, there can be two explanations. On the one hand, there is the fact of not being accustomed to note-taking on the digital, as more than one has reported, on the other there is the possibility

that the tool's usefulness was not immediately clear, as was suggested by some testers who, for example, hinted at the partial usefulness that it would have for them to access or search notes via keyword. This is definitely too little to derive conclusions. More research is needed as to why the users were not accepting the tools provided. So, we propose a structured longitudinal approach based on human-centered design to design and annotation tool that would actually be usable for scholars.

References

1. Cooper, A. et al.: About Face 3. The essential of interaction design. Wiley Publishing Inc., Indianapolis (2007).
2. Unsworth, J.: Scholarly Primitives: What Methods do Humanities Researchers Have in Common, and How Might our tools reflect this? part of a symposium on "Humanities Computing: formal methods, experimental practice" sponsored by King's College, London (2000), http://www.people.virginia.edu/~jmu2m/Kings.5-00/primitives.html, last accessed 03/10/2017.
3. Schacht, P. Annotation, MLA Modern Language Association Journal. https://digitalpedagogy.mla.hcommons.org/keywords/annotation/.
4. Bargeron D. & Moscovich T.: Reflowing digital ink annotations, CHI '03. In Proceedings of the SIGC Conference on Human factors in Computing Systems (eds G. Cockton & P. Korhonen), pp. 385–393. ACM Press, New York, NY, 2003.
5. Hoff, C., Wehling, U., Rothkugel, S.: From paper-and-pen annotations to artefact-based mobile learning, in Journal of Computer Assisted Learning (2009), 25, 219–237.
6. Marshall, C. C.: The Future of Annotation in a Digital (Paper) World). in Proc 35th GSLIS Clinic, Success and Failures of Digital Libraries, University of Illinois Urbana-Champaign 98.
7. Kiewra, K. A., The Relationship between Information Processing Ability and Note-Taking, in Contemporary Educational Psychology, 13, (1988) 33–44.
8. Jardine, L., Grafton, A.: "Studied for Action": How Gabriel Harvey Read His Livy, in Past and Present, Volume 0, Issue 129 (Nov., 1990), 30–78.
9. Bruce, Blair, A. 2003. Reading strategies for coping with information overload, ca.1550–1700, in Journal of the History of Ideas 64, no. 1: 11–28.
10. Drexel, J. S., Aurifodina artium et scientiarum omnium; excerpendi sollertia, omnibus litterarum amantibus monstrata, Antwerp, 1638.
11. Kwakkel, E.: Medieval Books, blog. https://medievalbooks.nl/author/erikkwakkel/.
12. Mueller, P., Oppenheimer, D.: The pen is mightier than the keyboard. Psychological Science 6(25) pp. 1159–1168 (2014).
13. Nelson, T. Literary Machines. South Bend, Indiana: The Distributors. Edition 87.1. Also 6th Edition, 1984.
14. Marshall, C. C.: Towards an Ecology of Hypertext Annotation. 98.
15. Marshall, C. C.: Annotation: from paper books to the digital library. in Proc DL 97, Philadelphia, PA (July 23–26, 1997), pp. 131–140.
16. Prié, Y.: Lecture Active et Annotation de documents. Slide presentation. http://pagesperso.lina.univ-nantes.fr/~prie-y/ens/04-05/MasterRecherche/lecture-active.pdf.
17. Nielsen, Jakob. "Guerrilla HCI: Using Discount Usability Engineering to Penetrate the Intimidation Barrier", 1994. [Accessed at http://www.nngroup.com/articles/guerrilla-hci/].
18. Koohang, Alex. "Expanding the Concept of Usability." Informing Science Journal, vol. 7, 2004, pp. 129–141. [Accessed at https://pdfs.semanticscholar.org/524f/7dc44ca12f79074bd3064ca3887bcb938218.pdf].