Memorandum

To: Dr. Lyn Gattis • From: Carolyn Clark • Date: March 2, 2020 |

Usability Test Methods Report

In this report, I will explore usability test methods and the potential application of these methods for our class's work on the Brave Butterfly project. My research for this report focuses on usability test methods for the website and the overall visual design of the project. I will discuss relevant usability test methods as they relate to content or visuals we have designed and/or written or plan to design and/or write for this client project. I will elaborate on three methods: think-aloud protocols, A/B testing, and testing visual design. In a section for each usability test method, I will present information about each testing method as well as the common advantages and disadvantages cited in research. After reviewing usability test methods, I will also discuss the ways to apply them to the Brave Butterfly project.

Think-Aloud Protocols

The think-aloud protocol or the concurrent think-aloud (CTA) protocol is a usability test method often used to evaluate a participant's experience interacting with a website. In "Assessing Concurrent Think-Aloud Protocol as a Usability Test Method: A Technical Communication Approach," Cooke (2010) defines CTA as a process where users verbalize their thoughts while performing tasks. The approach gives users a goal or task to perform as they interact with the website and communicate their reactions. The objective when employing the CTA method in usability testing is to gain insight into a user's behavior which would otherwise be difficult to obtain with observation alone (Cooke, 2010). This approach allows researchers to identify issues users discuss during CTA protocol like expressions of frustration or aspects of document/website the user enjoyed. Alhadreti and Mayhew (2017) state, "the traditional CTA method provides 'real-time' information during the participant's interaction with a system, which can make it easier to identify the areas of a system that cause problems for the user." The CTA protocol grants researchers the opportunity to assess users' immediate reactions and impressions of the website.

Usability testing with the CTA method prompts genuine, in-the-moment reactions from users. These reactions may be coded into different categories to better understand the similarities and differences between participants' responses. Cooke (2010) explains and defines the content categories a CTA usability test addresses:

> The content categories included reading, procedure, observation, explanation, and other….Reading was defined as participants' reading of words, phrases, or sentences directly off the screen. Procedure included participants' descriptions of their current or future actions. Observation was defined as participants' remarks about the webpage or their behavior. Explanation included motivations for participants' behavior. The "other" category included content that did not fit into the other four categories.

Categorizing user comments is helpful for recognizing patterns and common problems users face with the interface or the design of the website. While the CTA protocol provides authentic feedback from users as they interact with a website, there are some issues to acknowledge when using CTA for usability testing.

When considering the CTA method for a usability test, our class will need to acknowledge the limitations and/or problems with this usability test design. Alhadreti and Mayhew (2017) discuss the following problems as disadvantages of using the CTA protocol:

- The completeness of the gathered data.
- Participants may find the process of concurrent verbalization feels unnatural or uncomfortable as people do not commonly think-aloud while working (as cited in Nielsen, 1993).
- Talking aloud or concurrent verbalization while performing a task may interfere with or alter a participant's thought process and/or their performance of a task.

These problems will need to be acknowledged in our report on usability outcomes if our class chooses to use this method. Cooke's (2010) research suggests think-aloud comments from participants are incomplete because "verbalizations alone do not provide a complete picture of their experience." Application of the CTA method may require supplemental usability testing or collection of another type of data. Users are unlikely to purposefully omit their on-screen behavior from their think-aloud accounts, but verbalization may lack behavioral information; for example, if a participant lapses into silence or uses verbal fillers, their eye movements may reveal whether or not they are still actively engaged in performing the task (Cooke, 2010). Accounting for non-verbal cues and body language is helpful to identify more information the user may unintentionally exclude from their verbalization.

Another type of think-aloud protocol, retrospective think-aloud (RTA), could be used to potentially help solve the issues discussed. The U.S. Dept. of Health and Human Services (2014) states "in Retrospective Think Aloud (RTA), the moderator asks participants to retrace their steps when the session is complete. Often participants watch a video replay of their actions, which may or may not contain eye-gaze patterns." The RTA method may take out the potential interference or alteration with a participants' thinking or task performing. Alhadreti and Mayhew (2018) discuss a drawback of RTA "relates to the method's reliance on human memory, which is fallible: with the best of intentions, participants might forget specific things that occurred during a task." While the RTA method does not interfere with the performance of tasks, it may not be as reliable as in-the-moment feedback. McDonald, Edwards, and Zhao (2012) found that practitioners adopted the RTA method because it did not interfere with participants task performance and some researchers felt the method was more natural or spontaneous. The RTA method may provide a more comfortable experience for participants in a usability study.

Applications

I believe the traditional CTA method would provide insights into how the redesigned Brave Butterfly website functions for users and areas for improvement within the site. This method does appear to require supplemental data, like capturing users' eye movements; however, since our class will not have the equipment to track participants' eye movements, I would recommend screen recording as an alternative which allows us to pick up mouse movements to identify where users may be hesitating/confused or easily navigating with faster mouse movements. Filming participants (via webcam or another filming device) participants with their knowledge and consent would potentially provide another way to track eye movements without expensive software or lab settings. Designing a usability test with CTA protocol for the Brave Butterfly website may entail asking participants to find information about the scholarship opportunities, news/events, the four wings, and/or how to donate from the site. These tasks are relevant to site visitors who using the site for the first time or returning users. I believe the CTA and RTA methods could be used in combination to provide real-time reactions as well as more reflective thoughts after participants interact with the website.

A/B Testing

Usability testing with the A/B method is a way to compare different features of visual design (e.g., colors used in the design, organization of text and graphics, etc.) or written content (e.g., different tones, word choice, or etc.). "A/B testing, also known as bucket testing, split testing, or controlled experiment, is a standard way to evaluate user engagement or satisfaction from a new service, feature, or product" (Gui, Xu, Bhasin, & Han, 2015). Usability testing with A/B methods provides researchers with users' preferences and attitudes towards certain features on a website or document. Fichter & Wisniewski (2017) state A/B testing can test a small change like altering the placement of a headline or the color of a search button or a more drastic change like testing different organizational scheme or different copy. The A/B testing method would compare user input for different webpage designs to better understand which design performs better with usability or readability. "For measurable A/B tests, you first need to determine what you're trying to optimize—say the library card sign-up page—and then test some tweaks to the page" (Fichter & Wisniewski, 2017). Testing different versions of the same webpage supplies user commentary about which version they would prefer to use and/or the ways they prefer view or interact with the information on a particular webpage.

Implementing A/B testing effectively requires researchers to test two different versions of a webpage with a clear understanding of what is being tested. In Farmen's (2019) article, "A/B Testing: Optimizing the UX," he states "all successful a/b tests start with a hypothesis." Researchers will need to develop a hypothesis regarding the performance of different featutes prior to testing different versions with users. "To conduct an A/B test that will allow you to conclude, you will need a hypothesis, a controlled test [Test A], and an altered test [Test B]" (Farmen, 2019). Comparing Test A with Test B allows designers to compare which change, design, or writing style has the most positive impact according to participants. In "A/B Testing: Optimizing the UX," Farmen (2019) briefly discusses a case study testing the impact of an FAQ page on a non-profit website and informs readers certain websites require more information than others; with non-profit sites, it is difficult to know what is trustworthy, so providing more information can be better (Farmen, 2019).

Applications

With the Brave Butterfly project, our team would have the opportunity to determine which visual designs or written copy of the website or print/electronic documents would benefit from comparison testing with users. A/B testing will also be helpful throughout the project as we review designs, colors, and/or content with the class. This testing method would be especially beneficial for testing different color schemes, organization of information/content on a page, and the tone of the writing. Following questions may inspire ideas for designing an usability test with the A/B method:

- Which color scheme more effectively conveys the Brave Butterfly's voice/purpose?
- Which color or color scheme appeals to you?
- How would you describe the webpage's organization?
- How would you describe the tone of written copy?

I believe A/B testing will allow our team to compare the more effective options for design and content with the Brave Butterfly project. In our usability testing, it will provide our team with data-driven decisions from participants' responses. The A/B testing method will also create opportunities for our team to explore more options with the creative deliverables (website copy and visual design elements); our team can compare and choose between different design or writing versions based on what we believe will be effective for the Brave Butterfly. Using services like Google Analytics after the website is live will provide our client, Tami, with the option to perform A/B testing in the future for added webpages or content.

Testing Visual Design

There are multiple methods for performing usability testing with visual aspects of a website or document. Whitenton (2018) discusses two methods to gather data on participants' first impressions: the five-second test and the first-click test. These methods would help to garner information about a participant's first reactions to a website. With a five-second test, practitioners show the stimulus for five seconds or another short period of time to accurately capture a user's gut reaction and form an impression which reflects the visual style (Whitenton, 2018). This method would allow researchers to gain insight into a participant immediate thoughts about a web design. For a first-click test, participants are given a specific instruction, and before they are exposed to the design, a researcher stops them after they navigated to where they would complete the task to gather their impressions (Whitenton, 2018). This test may fit into an already designed CTA usability test, so after sharing their initial thoughts, a participant could continue with the traditional CTA method. Whitenton (2018) also discusses assessing user reactions through open-ended or structured questioning:

- Open-ended preference explanation: Ask users to explain *why* they like a design
- Open word choice: Ask users to list 3 to 5 words that describe the design
- Closed word choice (desirability testing): Provide users with a list of terms and ask them to pick the words which best describe the design
- Numerical ratings: Collect numerical ratings about how much the design exhibits specific brand qualities

These approaches to questioning participants in a usability study could provide more specific information or more relevant results for making changes to a website or document.

Applications

Supplementing our usability test design with approaches to test visual design can help our team formulate a more cohesive and usable brand for the Brave Butterfly. I believe the approaches Whitenton discusses can be implemented into other types of usability testing. First impression testing could be implement at the start of a CTA test or an A/B test, and open-ended or structured questions could be asked at the end of the usability test as a part of a RTA protocol. Testing visual design components will help our team to shape the visual identity of the Brave Butterfly.

Conclusion

In this report, I looked at CTA and RTA protocols, A/B testing, and approaches to testing visual designs for usability test methods relevant to our class's work on the Brave Butterfly project. I believe these methods can be used in combination or separately to build an effective usability test for the Brave Butterfly's website and other documents. I will work with our team to design usability test which provides insights into the website's usability, accessibility, and user engagement. I look forward to discussing usability test methods as team to design a usability test which provides insights into the ways we can improve the documents we designed for better user experiences.

References

Alhadreti, O., & Mayhew, P. (2017). To Intervene or Not to Intervene: An Investigation of Three Think-Aloud Protocols in Usability Testing. *Journal of Usability Studies*, *12*(3), 111–132. Retrieved from http://uxpajournal.org/intervene-think-aloud-protocols-usability-testing/

Alhadreti, O., & Mayhew, P. (2018). Are Two Pairs of Eyes Better Than One? A Comparison of Concurrent Think-Aloud and Co-Participation Methods in Usability Testing. *Journal of Usability Studies*, *13*(4), 177–195. Retrieved from http://uxpajournal.org/concurrent-think-aloud-co-participation-methods-usability/

Cooke, L. (2010, September 1). Assessing Concurrent Think-Aloud Protocol as a Usability Test Method: A Technical Communication Approach. IEEE Transactions on Professional Communication., 53(3), 202-215. doi:10.1109/TPC.2010.2052859

Farmen, N. (2019, September 3). A/B Testing: Optimizing The UX. Retrieved from https://usabilitygeek.com/a-b-testing-optimizing-the-ux/

Fichter, D., & Wisniewski, J. (2017). Boost Your Website Performance With A/B Testing. Online Searcher, 41(2), 66–68.

Gui, H., Xu, Y., Bhasin, A., & Han, J. (2015). Network A/B Testing: From Sampling to Estimation. *Proceedings of the 24th International Conference on World Wide Web (WWW 15)*, 399–409. doi: 10.1145/2736277.2741081

McDonald S., Edwards H.M., and Zhao T. (2012, March). Exploring Think-Alouds in Usability Testing: An International Survey. *IEEE Transactions on Professional Communication*, 55(1), 2-19, doi: 10.1109/TPC.2011.2182569

U.S. Dept. of Health and Human Services. (2014, May 15). Running a Usability Test. Retrieved from https://www.usability.gov/how-to-and-tools/methods/running-usability-tests.html

Whitenton, K. (2018, June 17). How to Test Visual Design. Retrieved from https://www.nngroup.com/articles/testing-visual-design/