

User Experience: Beyond Decoration and Usability

*Roxana M. Barbu*¹

¹ *Macadamian Technologies*

Gatineau, Quebec, Canada

ABSTRACT

A foundational method of User Experience (UX) is usability. Practitioners assess the effectiveness and efficiency of their product: can users complete a given task, and can they do so with ease? However, usability is only one facet of a user's experience. This paper draws attention to the subjective, intangible aspects of user experience. It reviews current approaches to assess subjective experiences and provides guidelines on how one may tailor these approaches to needs at the levels of users, teams, and products. Additionally, using digital health apps as a case study, the paper introduces the concept of a shared user-app journey and reviews the main stages the two traverse together. It provides examples of approaches to evaluate subjective experience at various phases along the user-app journey.

Keywords: User Experience, Usability, Subjective Experience, Desirability

INTRODUCTION

The textbook definition of user experience is user-centric and a co-creation of an interaction between a group or a person and a product (Baxter et al. 2015). However,

in early stages, user experience was limited to usability. A notable shift has been from ‘traditional’ usability to the broader notion of user experience. ‘Traditional’ usability was mostly concerned with the system or user interface: whether the task can be completed with ease or not. This ‘narrow’ definition persists; however, a ‘broader’ notion, which incorporates the user’s feelings, is evolving. Though UX has advanced as a field, and its importance in product/service development is undeniable, there is still room for growth.

Most practitioners agree on the value of usability testing. The approach is widely practiced, with clear guidelines. However, the ‘narrow’ usability practice prevails. The focus remains on function and is commonly measured by relatively objective means: whether users can complete tasks (effectiveness) and how long it takes them to complete the tasks (efficiency) (Rashid and Quigley, 2009). However, the International Organization of Standardization’s definition of usability includes not just effectiveness and efficiency, but also users’ “satisfaction in a specified context of use” (ISO). Others speak of the utility, usability, and desirability hierarchy such that, at the very least, a product/service should be functional, then easy to use, then pleasant. This hierarchical approach may be sometimes misleading. An un-pleasant product may deter users before they even get to experience the product’s functionality or ease of use. In such cases, one may say that whether the product is pleasant is more important than whether it is useful or usable. The takeaway is that these two notions of usability, ‘narrow’ - focused on functionality, and ‘broad’ - encompassing subjective user experience measures such as satisfaction, are complementary to each other.

One may argue that traditional UX methods do capture users’ subjective experience. Interviews elicit emotive words, quotes, and facial gestures. Usability testing and observations can record users’ sighs, facial expressions, and body posture. However, these methods capture subjective experience as a secondary goal, through the filter of the researcher, rather than directly from the user. Overall, there is a lack of emphasis on the users’ subjective experience and a general lack of understanding of its value.

This paper concerns itself with assessing users’ subjective experience, the more intangible aspects of user experience that reflect how the user feels. It reviews existing methods, after which it introduces the notion of stages. These stages serve as an initial guideline to start incorporating some of these approaches at different steps along a user’s journey.

USER EXPERIENCE: A MORE COMPREHENSIVE VIEW

Despite the ISO’s broader definition, usability is most often evaluated mainly through the objective measures of efficiency and effectiveness, with subjective experience being overlooked (Rashid and Quigley, 2009). Before discussing how one may assess

subjective experience, it is worth highlighting the myriad of terms used in the literature to describe it: aesthetics (Robins and Holmes, 2008), beauty (Tractinsky et al., 2000; Hassenzahl and Monk, 2010), desirability (Rashid and Quigley, 2009; Benedek and Miner, 2002; De Guzman and Schiller, 2011; Schiller and De Guzman, 2013), satisfaction (Rashid and Quigley, 2009; De Guzman and Schiller, 2011), coolness (Bruun et al., 2016), emotional reactions (Petrie and Harrison, 2009), to name a few. Note, however, that these terms appear only in a handful of publications and are even scarcer in practice. Furthermore, the terms are not always clearly defined. The fact that these terms are ill-defined in the context of user research makes them difficult to operationalize (i.e., capture as constructs of subjective experience). Increased visibility into how to assess subjective experiences helps reinforce the overall understanding of the concepts and improves the prevalence and standardization of assessment approaches that capture how the user may feel.

Though intangible, subjective experiences are not any less valuable to UX research than objective ones. The overall user experience is impacted by subjective feelings. Today, whether it is the appearance of a website, the ambience in a restaurant, the personalized responses from a chatbot that knows your name, all these extend beyond usability and shape our everyday experiences. In a competitive market, two products may have the exact same or negligible differences in functionality. The only differentiating factor between them may be how they make the user feel.

Approaches to Assessing Subjective Experiences

In this paper, the term subjective experience is used as an umbrella term to encompass approaches that focus on how the user feels or perceives something. This contrasts with objective measures of UX which assess performance and learnability. The intent is to provide consensus in the literature, which currently uses multiple terms without a common understanding of the definitions. This section consists of a review of current techniques, most rooted in Microsoft's desirability toolkit, which is the first reviewed.

Microsoft's desirability toolkit consists of a large set of words typically presented on individual cards. Users are presented with stimuli such as visuals of a website and given the cards set. Their task is to sort through the cards and select the ones that they associate with the stimuli at hand (Benedek and Miner, 2002). For example, a user may have selected the words "engaging, worthwhile, appealing". In the next phase, the researchers discuss with the user to understand what informed their answer. Advantages of this approach, according to the authors, include increased engagement on the user's side, that it does not involve a questionnaire, and those users do not have to generate words themselves. Additionally, the authors draw attention to the fact that users may be biased towards positive feedback during a usability testing. Thus, the set consists of at least 40% negative words, leading to users being more willing to share negative feedback. Uses of this approach include assessing one product or

comparing different versions of the same product. Assessing a single product provides insights into how the users feel, whereas comparing different versions of the same product can also inform one's decision as to which version to pursue (e.g., A/B/C testing).

Microsoft's toolkit provides a starting point to capture how the user feels; however, it can be adapted. One may modify or streamline the list of words by adding or removing words, shortening the list, or changing the presentation method (e.g., auditory instead of visual, for example) to match their user and product's needs. When working with a list of words, it is important to note that not all words have the same frequency. This may not have a huge impact in a university setting, or a high literacy and high language proficiency setting. However, different audiences may differ in their literacy skills and language barriers, among others. Therefore, it is imperative to revise any pre-existing tools such as Microsoft's toolkit through the lens of one's users.

De Guzman and colleagues provide an example of such a revision. They use a subset of the words from the Desirability Toolkit (De Guzman and Schiller, 2011). The subset was created by keeping only terms relevant to visual design feedback. Additionally, the researchers manipulated the time the users have for the exercise. Visual appeal can be assessed within 50 ms, suggesting that users need only to glance at a design to form their first impression (Lindgaard, et al., 2006). This decision is pre-reason, as users have not yet processed all the elements on the page. To capture feedback based on the users' analysis of the navigation structure, considering their mental models, and so on, users need more exposure and time. Researchers allowed users sufficient time to review the design so that behavioral-level processing of visual stimuli can occur, as opposed to a visceral response. The distinction between visceral and behaviour-led responses is crucial, and it leads to a discussion of which reactions should practitioners capture, when, and why.

Another adaptation of the desirability toolkit was to start by having participants sort through the 118 cards (Rashid and Quigley, 2009). The adjectives they selected were then converted into nouns. For example, the adjectives "easy to use, hard to use, effortless, simple, difficult" were converted into the "ease of use" noun; similarly, the adjectives "time-saving, fast, slow" were converted into the "promptness" noun. These nouns became the basis of the Likert scale, where users ranked several techniques from 1 to 5. The Microsoft team mentioned the lack of a questionnaire as an advantage whereas Rashid and colleagues adapted the method in the form of a Likert scale. Each method has its advantage. The initial toolkit is a qualitative measure not intended to be generalizable across all product users. The Likert scale provides a standardized way to easily compare across products. Additionally, it may address business constraints where the clients or stakeholders prefer quantitative data. It also provides a means for teams to prioritize. For example, if the users rated a given product equally low on both "ease of use" and "promptness", the team can discuss which of the two should be addressed first.

Teams also differ in whether they want users to select from a predefined set of words, or generate their own, outside of the set. To address this, one may include additional cards that are empty allowing participants to input their own words (De Guzman and Schiller, 2011). There are arguments for both approaches. Not having users generate words themselves minimizes the difficulty of their task. On the other hand, there is value in learning users' genuine reactions without any priming and in their own words.

A different approach known as a completion task, either fragment or sentence completion, is well established in the psycholinguistic/psychology literature, but new to the HCI field. The task, Visual Design Mad Libs (Schiller and De Guzman, 2013) in the authors' terminology, is a fill in the blanks exercise. Users see a visual image after which they have a fixed period such as one minute to complete a one to two sentence statement where parts of the sentence are blank. To assess the visual design of a product logo, the sentence to complete might be: "This is a logo for [company name]. The logo is [fill in the blank] because [fill in the blank]" (Schiller and De Guzman, 2013).

Finally, Bruun and colleagues put forward the COOL questionnaire which contains questions related to Desirability, Rebelliousness, and Usability (Bruun et al., 2016). The questionnaire is a 7-point scale ranging from 'strongly disagree' to 'strongly agree'. The Desirability component consists of the following statements: "This device – is cool; can make me better; is meant for people like me; can make me happy; can make me look good; can make me look in control of things; totally connects with me". The authors suggest that the questionnaire allows teams to identify which of the three factors, Desirability, Rebelliousness, or Usability mostly affect the cool perception for a given product. For example, in the context of a bank app, Usability may play a bigger role, whereas when buying a watch, Rebelliousness may trump Usability. Additionally, the questionnaire is purposely designed to be easily used for a variety of products by replacing the word 'device' with, for example, 'website'.

This section highlighted current methods of assessing subjective user experience. It also illustrated how one may start with an existing approach and tailor it to their users, teams, and products.

Roadmap to Incorporate Subjective Experience Approaches in your Process

Subjective experience is relevant across all kinds of products and services. However, the discussion to follow is written with digital health apps in mind for several reasons. For one, there is an intuitive difference between functionality-driven products such as thermometers, and experience-infused ones such as digital health apps. When looking to buy a thermometer, one may be more interested in its specs rather than aesthetic experience. A digital health app, however, has an additional personalization component to it.

When it comes to digital health apps, the user and the app are going through a journey together, including the following stages: Pre-onboarding, Onboarding, and Daily Usage. These stages can also be thought of as an analogy for the relationship between two people where the equivalent stages would be: First date, Honeymoon, and Long-term commitment. Different approaches to subjective experience work best at different stages.

Pre-onboarding. The pre-onboarding stage refers to the time when the user scrolls through the app store looking for an app to help them with a health goal such as losing weight. That scrolling may be fast paced, with the person not actively engaging with content beyond visual appeal. This first minimal interaction may determine whether they download the app or not. Like on a First date, the person needs to be convinced of a second date, or that downloading the app is worthwhile.

The measure chosen to assess subjective experience during the pre-onboarding stage should account for the minimal interaction. If the assessment approach allows for 2 to 3 minutes per app, the results would not translate to the real context of use when the interaction may be seconds or even milliseconds. At that speed, their attention will be captured by aesthetic elements alone (e.g., colour). The 50 ms decision test is a way to capture these pre-reason user reactions.

Doing in-depth testing of what users see briefly may be interesting, but it would not necessarily reflect reality because users cannot always accurately report on how they have arrived at a certain decision. For example, in a famous study, researchers placed four identical pairs of stockings on a table and asked shoppers to pick the pair of the highest quality. Right-most stockings were preferred over the left-most ones by a factor of almost four to one. When asked to explain their choice, none of the shoppers mentioned the position of the article. When prompted about a possible effect of the position of the article, virtually all shoppers denied it, based on the authors' interpretation, "a worried glance at the interviewer suggesting that they felt either that they had misunderstood the question or were dealing with a madman" (Nisbett, and Wilson, 1977).

This pre-onboarding stage is crucial in a saturated market where the users have plenty of choices. If the decision to open or download the app is made in the split of a second, for example as someone is scrolling through the app store, the testing approach must account for the duration of exposure in the real context. There are however products where this initial search may extend beyond aesthetics alone. When searching for a project management tool, the search may be dictated by functionality.

How users approach a search will depend on other factors such as their motivations, time, device compatibility, digital literacy skills etc. Thus, one needs to consider what the pre-onboarding stage looks like in the context of a given product and user base.

Onboarding. The onboarding stage refers to the first few times the user interacts with and discovers the app. At this stage, if the learning curve is too high, or the user is not clear about the value of the product, there is a high risk of drop-out.

Given the importance of learnability, one may be tempted to focus on 'narrow'

usability - efficiency and effectiveness - to ensure users can smoothly advance through the training. However, the mapping between objective measures such as completion rates and the users' perception of usability is not a one-to-one mapping. Thus, these more objective approaches should be complemented by approaches that capture the user's perception. For example, the customer net effort score assesses the perceived ease of the task. Additionally, measures of beauty or aesthetic experience should not be overlooked either. Perceived usability is related to perceived beauty such that given two products identical in content and usability but differing in aesthetics, users are more likely to perceive the aesthetically more pleasant one as more usable as well (Robins and Holmes, 2008). There is also a relationship between design and credibility suggesting that design has impact beyond decoration (ISO). In this Honeymoon stage of the user and app relationship, in addition to the aesthetic experience, it is critical to assess perceptions of qualities such as: trust, empowerment, reliability etc. Especially within the context of designing a digital health app where the users' perception of these qualities will impact their relationship to the application and therefore its efficacy.

Daily usage. The Daily usage stage refers to the active usage of the app. Depending on the purpose the app serves, this stage could last days to years. At this point in the relationship, the Long-term commitment, both subjective experience and 'narrow' usability come into play. The visual design and tone of the app should provide an engaging experience that delivers the qualities of the relationship that the user is expecting, and keeps the user coming back, whereas efficiency and effectiveness ensure the user can smoothly navigate the app and see their progress.

CONCLUSIONS

This paper presented a review of current approaches complemented by a discussion of how they can be further adapted. The concept of a journey and stages within an app provide a baseline to start thinking of what approaches may be used at different stages in the relationship between the users and the app. These methods lend themselves to comparisons between the "old" and the "new" experience as well as between versions in the context of A/B testing. Moreover, the approach facilitates alignment across stakeholders with respect to which metrics they value the most and informs their prioritization. It also guides them towards a consensus as to what brand sentiment they may want their product to evoke, and how that sentiment should evolve across the user's journey with their product. Subjective measures are harder to test, validate and verify as psychologically real and that may be why people tend to shy away from them. However, products do not exist in a vacuum, and users are becoming more trained at comparing experiences and their expectations are set according to what else is out there. The 'narrow' approach to usability is no longer sufficient to assess the efficacy of a design because the idea of efficacy now needs to be looked at across a journey during which the 'broader' definition of usability plays a critical role.

ACKNOWLEDGMENTS

The author would like to thank Macadamian Technologies for the constant support to bridge academia and industry. I especially thank Jennifer Fraser, Lorraine Chapman, Simran Singh and Cathy Agyemang for their invaluable input. A special thank you goes to Caroline Zenns for her insightful contributions on the importance of visual design beyond being aesthetically pleasant.

REFERENCES

- Baxter, K., Courage, C. and Caine, K. (2015). Understanding your users: a practical guide to user research methods. Morgan Kaufmann.
- Benedek, J. and Miner, T. (2002). Measuring Desirability: New methods for evaluating desirability in a usability lab setting. *Proceedings of Usability Professionals Association, 2003*(8-12), p.57.
- Bruun, A., Raptis, D., Kjeldskov, J. and Skov, M.B. (2016). Measuring the coolness of interactive products: the COOL questionnaire. *Behaviour & Information Technology, 35*(3), pp.233-249.
- De Guzman, E.S. and Schiller, J. (2011, July). How does this look? Desirability methods for evaluating visual design. In *International Conference on Human-Computer Interaction* (pp. 123-127). Springer, Berlin, Heidelberg.
- Hassenzahl, M. and Monk, A. (2010). The inference of perceived usability from beauty. *Human-Computer Interaction, 25*(3), pp.235-260.
- International Organization for Standardization, <https://www.iso.org/home.html>
- Lindgaard, G., Fernandes, G., Dudek, C. and Brown, J. (2006). Attention web designers: You have 50 milliseconds to make a good first impression!. *Behaviour & information technology, 25*(2), pp.115-126.
- Nisbett, R.E. and Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological review, 84*(3), p.231.
- Petrie, H. and Harrison, C. (2009). Measuring users' emotional reactions to websites. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems* (pp. 3847-3852).
- Rashid, U. and Quigley, A. (2009, July). Interaction techniques for binding smartphones: A desirability evaluation. In *International Conference on Human Centered Design* (pp. 120-128). Springer, Berlin, Heidelberg.
- Robins, D. and Holmes, J. (2008). Aesthetics and credibility in web site design. *Information Processing & Management, 44*(1), pp.386-399.
- Schiller, J. and De Guzman, E. (2013, July). Desirability Methods for Evaluating Visual Design. In *International Conference on Human-Computer Interaction* (pp. 154-158). Springer, Berlin, Heidelberg.
- Tractinsky, N., Katz, A.S. and Ikar, D. (2000). What is beautiful is usable. *Interacting with computers, 13*(2), pp.127-145.