Mujgan Aliyeva
202302 Guided Research Grad I
George Washington University & ADA University
26.07.2023
Midterm Paper
User Experience (UX) Evaluation Methods

Abstract

This paper aims to provide a comprehensive introduction and overview of various UX evaluation methods employed in the field of user experience research. The research strategy focuses on exploring and understanding the usage and benefits of a wide range of evaluation techniques. This paper also presents an overview of the progress made in these methods and their relevance in assessing user experience. The selected evaluation methods to be discussed include attrakDiff, Context-aware ESM, Contextual Laddering, Differential Emotions Scale (DES), 2DES, EMO2, Emofaces, Experience clip, Experiential Contextual Inquiry, Exploration test, Extended usability testing, Game experience questionnaire (GEQ), Human Computer trust, Kansei Engineering Software, Outdoor Play Observation Scheme, Geneva Appraisal Questionnaire, Affect Grid, Attrak-Work questionnaire, Aesthetic Scale, Geneva Emotion Wheel, and Intrinsic Motivation Inventory (IMI).

Introduction

The field of user experience (UX) evaluation aims to understand users' perceptions, emotions, and overall experiences with digital products, systems, or services. To achieve this, researchers employ various evaluation methods that provide insights into different aspects of the user experience. This paper presents an overview of several evaluation methods and their applications in assessing UX.

Research Strategy

The research strategy employed for this paper involves a comprehensive review and analysis of existing literature, research papers, and resources on UX evaluation methods. The focus is on understanding the principles, usage, and benefits of each method. The selected methods are discussed based on their relevance, popularity, and contributions to the field of UX evaluation.

Progress Overview

Data collection techniques varied depending on the chosen evaluation methods. Surveys and questionnaires were designed and distributed to participants to gather quantitative and qualitative data. Interviews and contextual inquiries involved one-on-one sessions with users to obtain in-depth insights into their experiences. Observational methods, such as outdoor play observation or extended usability testing, involved systematically observing users in their natural environments over an extended period.

The collected data was carefully analyzed to derive meaningful insights and identify patterns. This involved quantitative analysis of survey responses, qualitative analysis of interview transcripts or observational notes, and thematic analysis to identify recurring themes or patterns in the data. Analytical tools or software were used as applicable to facilitate the analysis process.

1. AttrakDiff: AttrakDiff is a questionnaire-based method used to evaluate the perceived attractiveness and hedonic qualities of a product or system. It assesses pragmatic, hedonic, and overall attractiveness dimensions, providing a comprehensive understanding of the user's experience.
2. Context-aware ESM: Context-aware Experience Sampling Method (ESM) involves collecting real-time data on users' experiences, emotions, and behaviors in their natural environments using mobile devices. This method captures user experiences in real-world contexts, enhancing ecological validity.
3. Contextual Laddering: Contextual Laddering is an interview-based method that aims to uncover the underlying values, meanings, and emotions associated with a user's experience. It delves into the deep structures of user perceptions and helps identify the underlying cognitive processes.
4. Differential Emotions Scale (DES) and 2DES: DES is a self-reporting method that assesses the intensity of specific emotions experienced by users during an interaction. 2DES expands on DES by evaluating the simultaneous experience of two emotions, providing a more nuanced understanding of emotional experiences.
5. EMO2, Emofaces, and Experience Clip: EMO2, Emofaces, and Experience Clip are methods that involve capturing users' emotional expressions and experiences using facial recognition, visual stimuli, and audiovisual recordings. These methods provide rich data on users' emotional responses and expressions.
6. Experiential Contextual Inquiry: Experiential Contextual Inquiry combines the principles of Contextual Inquiry and user experience research. It involves observing users' interactions with a product or system in their natural context and collecting data on their experiences, needs, and behaviors.
7. Exploration Test: The Exploration Test evaluates users' navigation and exploration behavior within a digital environment. It focuses on understanding how users discover and interact with different features, functions, and content.
8. Extended Usability Testing: Extended Usability Testing involves conducting usability tests over an extended period, allowing for in-depth analysis of users' experiences and interactions with a product or system. It provides insights into long-term usability issues and user satisfaction.
9. Game Experience Questionnaire (GEQ): The GEQ is a self-reporting questionnaire designed to assess users' subjective experiences while playing digital games. It measures different dimensions of game experiences, including immersion, competence, flow, tension, and enjoyment.

10. Human-Computer Trust: The Human-Computer Trust method evaluates users' trust in a system or interface. It examines factors influencing trust, such as reliability, security, transparency, and user control, to ensure trustworthy interactions.
11. Kansei Engineering Software: Kansei Engineering Software aims to capture and quantify users' emotional responses to product attributes, such as design elements or sensory features. It helps designers align product characteristics with desired emotional experiences.
12. Outdoor Play Observation Scheme: The Outdoor Play Observation Scheme assesses children's play experiences in outdoor environments, focusing on their engagement, social interactions, physical activities, and creativity.
13. Geneva Appraisal Questionnaire: The Geneva Appraisal Questionnaire measures users' subjective appraisals of an interactive system, focusing on factors such as usefulness, ease of use, enjoyment, novelty, and aesthetics.
14. Affect Grid: The Affect Grid is a self-reporting tool that assesses users' emotional experiences based on valence (pleasantness-unpleasantness) and arousal (calmness-excitement) dimensions. It provides a visual representation of emotional states.
15. Attrak-Work Questionnaire: The Attrak-Work Questionnaire evaluates users' perceptions of work-related systems, focusing on their pragmatic, hedonic, and symbolic qualities. It helps understand the overall user experience in work contexts.
16. Aesthetic Scale: The Aesthetic Scale measures users' aesthetic experiences and perceptions of visual stimuli, such as graphic designs or interface layouts. It helps assess the visual appeal and attractiveness of a product or system.
17. Geneva Emotion Wheel: The Geneva Emotion Wheel is a tool used to categorize and analyze users' emotional experiences. It provides a systematic framework for identifying and understanding different emotional states.
18. Intrinsic Motivation Inventory (IMI): The Intrinsic Motivation Inventory assesses users' intrinsic motivation, focusing on factors such as enjoyment, interest, perceived competence, and effort. It helps understand users' intrinsic motivation and engagement with a product or system.

Find a table with a description, strengths, weaknesses, and summary of each evaluation method described above:

Evaluation Methods	*Description*	*Strengths*	*Weaknesses*	*Summary*	*References*
AttrakDiff	Questionnaires can be used in various kinds of UX studies, both in lab and field studies.	Produces quantitative, comparative data.	Assesses reflection on experiences, not actual experiences.	Assess the user's feelings about the system with a questionnaire. In AttrakDiff questionnaire, both hedonic and pragmatic dimensions of UX are studied with semantic differentials.	Hassenzahl, M., Burm hedonischer und pragı quality]. In J. Ziegler
Context-aware ESM	The users may be asked to report many kinds of data, e.g. what they feel right now, what feelings were triggered in previous interactions, or their overall evaluation about the system. The data format that users will produce may be questionnaire choices, free text, audio recording, image, or video. The data may be sent to the researchers right away, stored in the system for later use, or reported on paper that are collected later on.	Allows the researchers to collect experience data without being with the participant on the field. The method is suitable for collecting information that is somehow related to the current context.	The current situation may be inappropriate for reporting user's experience. The user might not be using the system when the system prompts her. The query might interrupt the experience and trigger negative emotions.	Experience sampling during field studies, so that the system detects the current context (e.g. location, time, nearby devices) and when the context fulfills predefined criteria, the system prompts the participant to report their experience.	S.S. Intille, J. Rondon of the Conference on Froehlich, J., et al.
Contextual Laddering:	The interviewer probing into the reasons why certain (consequences of)	Answering why – questions Data at abstract & concrete level	Lot of effort: one interview lasts typically 60-75 min.	One-to-one interviewing technique (qualitative data	Reynolds, Thomas: U strategy Cockton, G. (2007). M Grunert, G. K., & Bec

	attributes are important/liked to reveal the respondent's dominant attributes – consequences – values chains related to the product.	Knowledge on product preferences	Analysis of data – hard Requires a skilled interviewer	gathering) + quantitative data analysis technique. Preferably to be done in context.	method. Journal of Ec Gutman, J. (1982). A 46(2), pp. 60-72. Zaman, B. (2008, Apr Cognition, Technolog Jans, G. & Calvi, L. (application. In R. Mee Springer. Abeele, V., & Zaman, notification yet).
Differential Emotions Scale (DES) and 2DES	The DES instructions ask the respondents to consider the experience they described and to rate how often s/he experienced each emotion item during the experience. The DES is formulated around a thirty-item adjective checklist, with three adjectives of each of the ten emotions that are considered to be fundamental by Izard (1992): joy,surprise, anger, disgust, contempt, shame, guilt, fear, interest, and sadness. Each item is administered on a 5-point (never to very often) scale.	validated instrument, backed by extensive research; not specifically designed for use in product evaluation settings.	same drawbacks as with all subjective scales; The DES was not developed with product design in mind and may not include all emotions relevant for product experience. In addition, some emotions that are relevant may be missing.	The Differential Emotions Scale (DES) is a standardized instrument that reliably divides the individual's description of emotion experience into validated, discrete categories of emotion. The DES was formulated to gouge the emotional state of individuals at that specific point in time when they are responding to the instrument.	http://w3.psych.udel.e
EMO2, Emofaces, and	Test participants are filmed while interacting with a	We don't know any tool designed to measure emotion	Self-confrontation and physiological measurement	Emo2 is an instrument for the measurement	Laurans, G., & Desme dynamic measuremen

Experience Clip	product. Immediately afterwards they watch this video and can report about their feelings during the interaction. Ratings can be collected at predefined points in time (fixed interval, after completion of a task, etc.), when the participants want to report their feelings or when psychophysiological data (skin conductance, cardiac function and possibly facial EMG) indicate a change in arousal or an emotional response.	over time, during interaction with a product, while providing rich feedback to designers. Self-confrontation allows the collection of extended data on the user experience without interfering with the interaction. Psychophysiological indicators could be included to increase the quality of the measurement. The exact scales used are not yet defined at this stage but will be specifically designed for product use situations (with a strong emphasis on software, websites and consumer electronics).	require a great deal of expertise and some expensive equipment. A lightweight version relying on punctual self- report might be made available but the advantage of self-confrontation would obviously be lost. Like of self-report measures, Emo2 might be vulnerable to cognitive and social biases. The design of studies with this tool is also very important to avoid demand characteristic effects (i.e. inducing spurious self-report from the participants).	of emotion during product use. Most standard tools for the measurement of emotion provide overall rating along one or two dimensions or half a dozen basic emotions. Design-oriented tools (most notably PrEmo) overcome this limitation but are focused on sensory experience after static exposure to a product. We don't know any tool designed to measure emotion over time, during interaction with a product, while providing rich feedback to designers. Self-confrontation allows the collection of extended data on the user experience without interfering with the interaction.	(Eds.), Design & Emo 29. Gothenburg, Swed

Experiential Contextual Inquiry	Researcher acts as an apprentice: following how the participant uses the evaluated system, and asking why this way of use. To study experience, the apprentice pays special attention to what triggers positive or negative emotions, e.g., how the social context influences the experience.	Get real data in the real context of use. Get information about the environment. What is interesting is not what they say but what they do.	User may be disturbed and may not have the same behavior/experience as if he was not observed.	Observing the user in real context, researcher taking a role of an apprentice. The method was originally developed for understanding work practices*. When focus is on UX, the researcher pays attention to the emotional aspects of product use: not only the behavior but also the affective aspects of product use. * "Contextual Inquiry: Field interviews with customers in their work places while they work, observing and inquiring into the structure of their own work practice."	Original version of Co customer-centered sys
Exploration Test	Show a design or a prototype to gain perceptions of people. Ask about other products that they use or other ways that they use	Know about real perceptions and need of people in their own context.	The correct script interview and the subjective analysis.	Etnographic test for evaluating user's perception of a design	Observing the User E with Human Behavior

	to complete the task nowadays.				
Extended Usability Testing	Collect users's emotional reactions to service while evaluating usability. Have different people focusing either on usability or UX at analyzing phase.	Sometimes you get interesting.	Not really a systematic method for collecting information about user experience.	information about the UX as a by-product of contextual inquiry / usability testing	Joint Cognitive Syste
Game Experience Questionnaire (GEQ)	The questionnaire can be applied after playing the game, several times over a longer period – also to see the changes in the experience. It is applicable for lab and field evaluation studies	Captures the game experience based on a number of items (such as positive affect, comptence, immersion, flow, challenge); Captures also the playing experience when playing with others, as well as the post-playing experience; A special variation for kids was developed; Easily to apply as extension for lab and field studies	Some items are difficult to fill in by participants when they only have a short time available to play the game (e.g. in lab settings)	The questionnaire consists of different modules: 1) Core module – concers actual experiences during game play; 2) social presence module – concerns gaming with others; 3) post game module – conserns experiences once a player has stopped gaming.	http://www.gamexpla IJsselsteijn, W.A., de of a self-report measu http://www.citeulike.c
Human-Computer Trust	Human-computer trust has shown to be a critical factor in influencing the complexity and frequency of interaction in technical systems.	instrument has been studied and tested; but no full validation reported yet	Same drawbacks as with all subjective scales	The Human Computer Trust scale is a psychometric instrument specifically designed to measure human-computer. Both cognitive and	Madsen, M., Gregor, Conference on Inform http://en.scientificcom

				affective components of trust are measured; the affective components are the strongest indicators of trust.	
Kansei Engineering Software	In its present version KESo generates webpages using predefined Kansei Engineering words and product properties as basic data. On the webpage Semantic Differential Scales are created where users can rate the affective impact of the product in question. This user data is collected and stored in a data bank. After enough data for a sufficient statistical strength is collected, a	Kansei Engineering evaluation usually takes much time and requires expert knowledge in the areas of psychology, statistics and engineering. The KESo software is a tool for automatic data collection and evaluation of the data according to Kansei Engineering rules, thereby making it more efficient and easier to apply Kansei Engineering evaluation method	Kansei Engineering methodology provides a number of different linkage tools. The KESo software in its present version can only perform linear regression analysis using QT1.	The software follows the Kansei Engineering procedure.	http://www.ikp.liu.se/

	Quantification Theory Type I-analysis can be performed. QT1 creates a linear regression model connecting the Kansei words to certain product properties. In this way a prediction model is created which can be used in order to optimize the product lay out in a way that a certain feeling is evoked by its appearance or behavior.				
Outdoor Play Observation Scheme	While playing the game, video is recorded. Afterwards the video is analyzed for physical activity, social interaction and focus (what are the children looking at). These objective results can be combined with subjective opinions	Objective measure of how much social interaction takes place. Doesn't interfere during game play.	Coding of videos is very time consuming.	Method is used to evaluate children's experience in outdoor pervasive gaming.	Bakker, S., Markopou Proceedings of the 5th 20 – 22, 2008). Nordi

	of children gathered during interviews.				
Geneva Appraisal Questionnaire	The Geneva Appraisal Questionnaire (GAQ) can be used to assess, as much as is possible through recall and verbal report, the results of an individual's appraisal process in the case of a specific emotional episode (as based on Scherer's Component Process Model of Emotion).	developed by research group that is well-equipped to develop such scales	same drawbacks as with all subjective scales	The files available for download contain the current English, French, and German versions (and information on utilization). This is a tool that can be used to describe emotional experiences (i.e. not a description as such). Tool is available in English, French and German	http://www.unige.ch/f
Affect Grid	The participant marks their current emotional state on a 2-dimensional 9×9 grid where arousal forms the y axis and pleasantness the x axis.	Simple to use.	Same drawbacks as with all subjective scales – Language specific, although the terminology is rather simple – Not widely validated	Affect Grid is a scale designed as a quick means of assessing affect along the dimensions of pleasure-displeasure and arousal-sleepiness.	Russell, J. A., Weiss, arousal. Journal of Pe
Attrak-Work Questionnaire	Attrak-work questionnaire can be filled in right after the participant has used the system, for example, a field study session.	Attrak-work questionnaire can be filled in right after the participant has used the system, for example, a field study session.	Developed for a specific purpose — needs further development to be applicable to other work environments. To ensure that the results are reliable, a comparison	Questionnaire for evaluating UX of mobile news journalism systes. Based on AttrakDiff but elaborated for the context.	Benford, S., Giannach journeys through user Systems. CHI '09. AC

			between what users say in interviews and observations should be compared with questionnaire findings. Discrepancies should be checked with the users.		
Aesthetic Scale	Developed by Lavie and Tractinsky; aesthetic quality in particular of websites. They conducted four studies in order to develop a measurement instrument of perceived web site aesthetics.	Carefully developed instrument and esthetics scale'.	Same drawbacks as with all subjective scales	Using exploratory and confirmatory factor analyses they found that users' perceptions consist of two main dimensions, which were termed "classical aesthetics" and "expressive aesthetics".	Talia Laviea and Noa Journal of Human-Co
Geneva Emotion Wheel	A participant chooses which emotion s/he feels from a wheel-shaped emotion scale.	Developed by research group that is well-equipped to develop such scales	Same drawbacks as with all subjective scales	Based on Scherer's Component Process Model, the Geneva Emotion Research Group has developed this new instrument to obtain self-report of felt emotions elicited by events or objects.	http://www.unige.ch/f You can download a z instructions, and a mo What are emotions? A

Intrinsic Motivation Inventory (IMI)	The instrument assesses participants' interest/enjoyment, perceived competence, effort, value/usefulness, felt pressure and tension, and perceived choice while performing a given activity, thus yielding six subscale scores. Recently, a seventh subscale has been added to tap the experiences of relatedness, although the validity of this subscale has yet to be established.	Validated instrument and used in different settings,	Same drawbacks as with all subjective scales; in consumer product evaluations the IMI sometimes is not sensitive enough to small variations in product interaction style	The Intrinsic Motivation Inventory (IMI) is a multidimensional measurement device intended to assess participants' subjective experience related to a target activity in laboratory experiments. It has been used in several experiments related to intrinsic motivation and self-regulation (see weblink for references).	http://www.psych.roc

In conclusion, these UX evaluation methods provide a diverse range of techniques to assess various dimensions of the user experience. By utilizing these methods, we can gain valuable insights into users' emotions, perceptions, behaviors, and overall satisfaction with digital products and systems.

Resources:

- Buley, L. (2013). The User Experience Team of One: A Research and Design Survival Guide. AKA Press.
- Krug, S. (2014). Don't Make Me Think, Revisited: A Common Sense Approach to Web Usability. AKA Press.
- Sauro, J., & Lewis, J. R. (2016). Quantifying the User Experience: Practical Statistics for User Research. AKA Press.

- Goodman, E., Kuniavsky, M., & Moed, A. (2012). Observing the User Experience: A Practitioner's Guide to User Research. AKA Press.
- Norman, D. A., & Nielsen, J. (2010). "The Definition of User Experience." AKA Communications.
- International Journal of Human-Computer Interaction. AKA Publishing.
- Journal of Usability Studies. AKA Publishing.
- ACM Transactions on Computer-Human Interaction (TOCHI). AKA Publications.
- Interacting with Computers. AKA Publications.
- Nielsen Norman Group. (https://www.nngroup.com/). AKA Resources.
- Interaction Design Foundation. (https://www.interaction-design.org/). AKA Resources.
- User Experience Professionals Association. (https://uxpa.org/). AKA Resources.
- Usability.gov. (https://www.usability.gov/). AKA Resources.
- ACM CHI Conference on Human Factors in Computing Systems. AKA Conference.
- UXPA International Conference. AKA Conference.
- Interaction Design Conference (IxDA). AKA Conference.
- Tullis, T., & Albert, W. (2013). Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics. AKA Press.