

HOW STARTUPS SHOULD EVALUATE THEIR WEBSITE'S UX?

Comparison of UX evaluation methods through evaluation of UX of the prototype of a matching platform for the rental housing market in Finland –
Sopia.

Master's Thesis
Lidia Borisova
Aalto University School of Business
International Design Business
Management
Autumn 2019

Author Lidia Borisova

Title of thesis How startups should evaluate their website's UX?

Degree Master's Degree

Degree programme International Design Business Management

Thesis advisor(s) Perttu Kähäri

Year of approval 2019**Number of pages** 102 + 2**Language** English

Abstract

Key concepts of this Master's thesis are user experience (UX), usability, startup and UX evaluation methods. The research question is how startups should evaluate their website's UX. To answer it I conducted a study of five UX evaluation methods: heuristic evaluation (HE), cognitive walkthrough (CW), tree testing, system usability scale (SUS), brainstorming through theoretical and empirical analysis. I collected empirical data in two ways. First, I interviewed seven UX field practitioners on their experiences of different UX evaluation methods. Second, I applied the evaluation methods to the website prototype of a digital startup called Sophia. To be able to consistently compare the UX evaluation methods, I created a theory-based framework that includes a set of generic parameters describing evaluation methods, and the constraints of the startup.

Based on my findings three UX evaluation methods would be useful in the startup context: heuristic evaluation, cognitive walkthrough and brainstorming. Practitioners tend to select flexible, fast and simple evaluation methods. Cognitive walkthrough and brainstorming match these criteria. Cognitive walkthrough when conducted with potential end users, reveals UX mistakes at an early stage of UX design. Brainstorming carried out within the design team afterwards helps to find resolutions for the revealed usability problems. Heuristic evaluation should not be carried out in its traditional definition with usability experts. However, startups should learn 10 heuristics as 10 usability principles to create the ground of good UX.

The key contribution of my study is the framework of *Minimum Viable UX Evaluation Methods for Startups*. The framework represents the list of necessary UX evaluation tools that each startup, despite time, money and human resource constraints, should follow. Each evaluation method, based on my findings, is unavoidable to help the startup to progress with product development.

Keywords: user experience (UX), usability, startup, UX evaluation methods, heuristic evaluation (HE), cognitive walkthrough (CW), tree testing, system usability scale (SUS), brainstorming

TABLE OF CONTENTS

LIST OF TABLES	3
LIST OF FIGURES	4
1. INTRODUCTION	5
1.1. Introduction to the topic and its significance	5
2. LITERATURE REVIEW	8
2.1. Usability and user experience (UX).....	8
2.2. UX design	10
2.3. UX design in startups	14
2.4. UX evaluation methods.....	17
2.5. Theory-based research framework.....	22
3. METHODOLOGY	27
3.1. The basic logic for the methodology.....	27
4. DESCRIPTION OF SOPIA	30
4.1. Description of Sophia's UX design process.....	31
5. SELECTION OF INTERVIEWEES AND THE LOGIC OF INTERVIEW QUESTIONS.....	35
6. STUDY OF UX EVALUATION METHODS	38
6.1. Heuristic evaluation	38
6.1.1. Academic literature analysis of HE.....	38
6.1.2. Empirical part 1– conclusions from the interviews with practitioners.....	41
6.1.3. Empirical part 2 - Heuristic evaluation of Sophia MVP	44
6.1.4. Conclusion	47
6.2. Cognitive Walkthrough.....	50
6.2.1. Academic literature analysis of CW	50
6.2.2. Empirical part 1– conclusions from the interviews with practitioners.....	53
6.2.3. Empirical part 2 - The cognitive walkthrough tests of Sophia prototypes.....	56
6.2.4. Conclusion	58
6.3. The system usability scale (SUS).....	60
6.3.1. Academic literature analysis of SUS.....	60
6.3.2. Empirical part 1– conclusions from the interviews with practitioners.....	62
6.3.3. Empirical part 2 - Sophia case	63
6.3.4. Conclusion	64

6.4.	Tree testing tool	66
6.4.1.	Academic literature analysis of the tree testing tool.	66
6.4.2.	Empirical part 1– conclusions from the interviews with practitioners.....	68
6.4.3.	Empirical part 2 – The Sophia case	69
6.4.4.	Conclusion	70
6.5.	Brainstorming session	72
6.5.1.	Theoretical part	72
6.5.2.	Empirical part 1– conclusions from the interviews with practitioners.....	74
6.5.3.	Empirical part 2 – The Sophia case	76
6.5.4.	Conclusion	76
6.6.	The framework of Minimum Viable UX evaluation methods for digital startups.....	78
7.	DISCUSSION.....	84
7.1.	When should we start UX evaluation?.....	84
7.2.	Factors that affect UX.....	86
8.	CONCLUSION.....	90
8.1.	Research summary	90
8.2.	Practical implications.....	91
8.3.	Limitations of the study	92
8.4.	Suggestions for further research	93
9.	REFERENCES	94
10.	APPENDICE.....	99
10.1.	Heuristic evaluation results.....	99

LIST OF TABLES

Table 1 - The repetitive patterns to describe or compare the usability evaluation methods in six academic articles about usability evaluation methods.....	20
Table 2 – Description of interviewees.....	35
Table 3 – Interview questions.....	37
Table 4 – The interpretation of heuristics.....	39
Table 5 - Heuristic evaluation results for Sophia MVP.....	44
Table 6 - The usage of UX testing tools in UX design practice of interviewees and in the case of Sophia.....	79
Table 7 - The opinion about the effectiveness of UX testing tools from interview respondents and my conclusion in the case of Sophia.....	80
Table 8 - The framework of Minimum Viable UX evaluation methods for digital startups..	82

LIST OF FIGURES

Figure 1 - The UX design process based on the double-diamond model.....	11
Figure 2 - The different stages of the "Develop" phase if a company is creating a digital product.....	12
Figure 3 – Iterative process of product development.....	14
Figure 4 - Theory-based research framework to describe the UX evaluation method.....	23
Figure 5 – Theory-based research framework to describe the UX evaluation method involving startup constraints.....	24;(90)
Figure 6 - Theory-based generic research framework to describe the UX evaluation method involving startup constraints in reference to the stage of UX design process.....	25
Figure 7 – Zooming into the testing phase as a part of the UX design process.....	26
Figure 8 - Research methodology structure.....	29
Figure 9 – Sophia’s development life span.....	32
Figure 10 - The suggested number of evaluators to perform heuristic evaluation.....	41
Figure 11 – The description of heuristic evaluation method using theory-based research framework to describe the UX evaluation method involving startup constraints.....	48
Figure 12 – The description of CW method using theory-based research framework to describe the UX evaluation method involving startup constraints.....	58
Figure 13 – The description of SUS method using theory-based research framework to describe the UX evaluation method involving startup constraints.....	65
Figure 14 – The description of tree testing method using theory-based research framework to describe the UX evaluation method involving startup constraints.....	70
Figure 15 – The description of brainstorming method using theory-based research framework to describe the UX evaluation method involving startup constraints.....	77
Figure 16 - The correlation between the quality of UX and quality of user’s feedback during UX evaluation.....	85
Figure 17 - Factors that impact UX.....	88

1. INTRODUCTION

1.1. Introduction to the topic and its significance

This thesis studied the methods to evaluate the user experience (UX) of the website. The context was the startup environment. The focus was the website of the startup. My analysis of the evaluation methods is relevant in all fields where startups could operate. Nowadays, almost every digital startup has a website as its main product, such as a platform, online shop, service etc., or as a digital representation of the company to the public. My research question in brief is: how startups should evaluate their website's UX? That means that I was investigating how to evaluate the website's UX in the most cost-effective manner in the face of scarcity of resources and time constraints that startups usually experience.

My findings correlate with the principles of running a startup, such as scarcity of resources and time pressure. In other words, I aimed to research how far the startup has to go in UX evaluation of the website and establish the optimal evaluation methods for this research.

When is the right moment to start evaluation? Is it worthwhile to evaluate the UX of the early stage prototype of the website, or the UX of the Minimum Viable Product (MVP) or the UX of the finalized product? How do methods differentiate from each other in terms of effectiveness to reveal usability problems and cost-effectiveness? Which testing methods could form a constant of the evaluation process? Can I define a minimum viable testing process of the website's UX for a startup? This is a list of sub-questions for the study. Sub-questions are both theoretical and practical, related to the main question.

My motivation to take the topic of UX evaluation methods was grounded in my personal professional journey: while working as UX designer for startup Sophia I asked myself the question how could I evaluate the UX of the prototype of our website. I studied early in Aalto University different UX evaluation methods. However, when I was working in Sophia the answer which UX evaluation method to use was not obvious for me because some methods seemed to be too difficult to use as they required special equipment, some – special training, some - too much time to make evaluation. I assumed there were many people in startups facing similar question as I had. Thus, I decided to research UX evaluation tools in the context of startup environment thinking that the results of my findings would be relevant for entrepreneur's developing the websites for their startups.

Working in the case company Sophia that develops a platform for the rental housing market I incorporated researched methods in the UX design process of the case company. We created the prototype of Sophia's website. During prototype development process I simultaneously tested UX evaluation methods effectiveness.

One of the biggest limitations of startups is a lack of human resources. Often during the first years of startup existence, a startup team cannot afford to hire a UX design professional. This can result in a situation when some team members take this new role and perform UX design by themselves. The same case happened to me. In Sophia, I carried out the role of UX designer without previous practical experience in the field because we needed to develop Sophia's website, being Sophia's MVP. In my Master Thesis I aimed to define the minimum viable UX evaluation framework that will help UX design beginners, as was I, to bring value to the startup they work for. By applying the methods that I found to be effective in the constraints of startups, a non-experienced designer could create a website with a good quality UX.

Key concepts of this Master thesis are UX, usability, startup and UX evaluation methods. Usability is the term that describes the quality of interaction of the user with a certain product. In this study the product is a website. In ISO 9241, usability is defined as the effectiveness, efficiency, and satisfaction with which specified users achieve specified goals in particular environments (Bevan, Carter, & Harker, 1998). All three components are defined in the following way:

- effectiveness means the possibility for users to achieve their objectives;
- efficiency represents the way for users to perform in a best manner with least resources of time and efforts to achieve the goals;
- satisfaction could be defined as the feelings that users get during the process, how comfortable the process is for them (Brooke, 2013).

Comparing to the usability UX is a multidimensional concept with multiple definitions made by scholars. In ISO 9241, UX is defined as "person's perceptions and responses resulting from the use and/or anticipated use of a product, system or service" (ISO 9241-210:2010). Thus, usability is attributed to the product, but UX is attributed to the user – what is the user reaction from the experience of using that product. User experience could be defined "as the feelings that the user gets when using a product" (Kraft, 2012).

In my Master's Thesis I focused on startup environment. The startup differs from established company by the following concerns: "being young and immature, having scarce resources, operating with novel technologies in dynamic markets, and being influenced by divergent stakeholders such as investors, customers, partners, and competitors" (Hokkanen et. al, 2016, p.2).

The usability inspection methods and specific UX evaluation methods could check UX. Usability inspection is the umbrella term for a group of methods when evaluators check the usability of a system that could be digital or physical. The main objective for inspection is to diagnose possible usability issues and to produce design-relevant data that can afford to correct the usability of a system (Nielsen & Mack, 1994). The research approach was to analyze four usability evaluation methods: heuristic evaluation, cognitive walkthrough (CW), system usability scale, tree testing, and one UX evaluation method - brainstorming. The research focused on investigating the origins of each method, the purpose of applying the method, the conditions under which to apply it, and the effectiveness of the methods in revealing UX problems.

I conducted the study through theoretical and empirical analysis. I gathered the empirical data of this study through seven interviews with UX field practitioners and through analysis of the results of the application of researched evaluation methods to the prototype of the website of the case company – digital startup Sophia. Among the interviewees were five representatives of startups, one research assistant from university and the director of a UX team in a large manufacturing company. The broad set of respondents allowed me to compare the usage of listed evaluation tools with the practice of UX evaluation in different environments.

My study has an unconventional structure. Instead of separation of theoretical research part from an empirical one, I made the differentiation of the parts of the narrative based on the type of evaluation method. I described every method using the following approach: first, I researched UX testing method based on the literature review examining scholars' papers on the topic. I provided the academic definition of the method and showed the method origins. The next section is the outcome of empirical research through the interviews with usability, UX and UX evaluation practitioners. The third section represents the results of incorporation of researched method in the UX design process of Sophia. This sequence of one theoretical section followed by two empirical ones repeats five times for every methods I was investigating.

The Outcome of my study is the framework of Minimum Viable UX evaluation methods for startups. The framework represents the list of necessary UX evaluation tools that each startup, despite time, money and human resource constraints, should follow. Each evaluation method, based on my findings, is unavoidable to help the startup to progress with product development. Thus, my research question was relevant for entrepreneurs.

Based on my findings represented in the framework of Minimum Viable UX evaluation methods (see the corresponding section above) for startups, three methods from all five that I investigated could be useful in the startup context but in different format: HE, CW and brainstorming. Practitioners tend to get rid of the complexity and special requirements of original usability inspection methods and are navigating towards more flexible, fast and simple tests. CW and brainstorming match this trend. CW allows to get real potential end users on board, as they can provide valuable feedback and reveal flaws in the product prototype at the early stage. Brainstorming carried out after, inside the design team helps to find resolution for revealed usability problems. HE should not be performed in its traditional definition with usability experts. However, entrepreneurs while developing the prototype of the system and later, MVP, should be familiar with 10 heuristics as 10 usability principles to create the ground of good UX.

2. LITERATURE REVIEW

2.1. Usability and user experience (UX)

I early provided the definition of usability from ISO 9241 - “the effectiveness, efficiency, and satisfaction with which specified users achieve specified goals in particular environments” (Bevan, Carter, & Harker, 1998, p.2). Norman Nielsen, publicly acknowledged as a guru of UX design, added to the ISO definition more usability attributes such as learnability, memorability and satisfaction (Khajouei, Esfahani & Jahani, 2017). Agreeing with these definitions, for my personal perception of usability I chose the definition of usability by Nielsen et Mack (1994): usability is a rather vast notion that is associated with “how easy it is for users to learn a system, how efficiently they can use it once they have learned it, and how pleasant it is to use” (Mahatody, Sagar & Kolski, 2010, p.2). To sum up, it means that after the user encounters the system for the first time the process of understanding how it works should be smooth, fast and easy so that the user will not

drop it and will be willing to get back next time when he or she will have the same task to accomplish.

UX is broader concept than usability, appeared as a counter trend to the prevailing task-oriented usability' pattern (Hassenzahl & Tractinsky, 2006). Don Norman, the inventor of the term UX design, told that the UX is incorporated in all facets of the user's interaction with the product from "industrial design, graphics, the user interface, the physical interaction, and the instruction" (Merholz, 2007). If the website has good usability, in other words, it is easy to use, easy to learn and enjoyable, the experience will be positive. Consequently, usability is one of the building blocks of UX.

However, usability is not enough to create a unique user experience. The UX refers to a broad range of experiences. The spectrum starts from "the traditional usability to beauty, hedonic, affective or experiential aspects of technology use" (Hassenzahl & Tractinsky, 2006, p.2). In other words the UX is a flow of emotions that the user feels while anticipating with your product, website, or system (Kraft, 2012).

One of the goals of UX is to reveal an emotional response from the user (Sutcliffe, 2010). In case of such digital experience as a website, the goal of usability is to create a system that is easy to use, whilst the UX is aiming to provide positive user feelings before, during and after the interaction. Consequently, usability refers to the ease of use while user experience is focusing on how users perceive their interactions with that system.

Usability of the system does not include the aesthetics of the system. However, aesthetics make the system appealing and attractive, provoking an emotional response of the user to the system. If the system will effectively help the user to achieve his or her goal, and in addition, will form an emotional linkage, - the chances of the success of such a product or service are high.

The aesthetics form a significant part of the UX design, however, due to the limitations of my study, I did not carry out comprehensive theoretical and empirical research into the aesthetic of the website, but will consider them together with the UX evaluation methods I choose to investigate.

2.2. UX design

UX design is the process of creating a unique UX, consequently building a positive experience for the user of the product. In my study, I was focusing on the design of digital experiences where physical interaction between the user and the product is limited by pushing the buttons on the screen, seeing pictures and listening to the sounds if the website has any. One of the formal definitions of UX is that UX is the development and synchronization of components that influence customers' UX with a particular system, with the intent of affecting customers' opinion and attitude (Unger & Chandler, 2012). Thus, UX design is a methodology that practitioners apply to develop a digital experience.

Among practitioners of UX design, there are many different approaches to a holistic UX design process. This could be explained by the fact that all products and services are different, for different categories of users and different goals, thus, the methodology each time should be adjusted to reach intended goals. The theoretical research of the academic consensus of the UX design process lies beyond the scope of this study. However, in order to show the place of UX evaluation within the UX design process, I had to present and elaborate my personal definition of the UX design process. Being a practitioner and design-thinking teacher myself, referring to my experience and my education, I choose design thinking (DT) as a general theoretical and practical framework for the UX design process.

Design thinking is understood as a relatively new approach for product development that originates from the cognitive process of designers. Currently, design thinking is attributed to a specific methodology that stimulates the creative process, enabling innovation. Design thinking is not only a methodology to boost innovations promoted by designers, but also a set of processes and toolkits which aid to enhance, stimulate and envision every creative journey, performed not only by professional designers but also programmers, UX designers, managers in diverse teams in any kind of organization (Tschimmel, 2012, p.2). Because design thinking is a general methodology that can be applied to any projects whatever the field and scope, I used it as a framework for the UX design process. I analyzed the UX design process through design thinking in theory and I followed the design thinking methodology in practice while performing the UX design for case company – Sophia.

In 2005, the British Design Council made a visualization and description of the modes of thinking that designers use – the Double Diamond model. It is not a unique representation of the design thinking process, but the one that is perceived as popular. The Double Diamond model is a simplistic visualization of the design thinking process: the sequences of divergent and convergent phases that repeats two times. (Tschimmel, 2012, p. 9). “In all creative processes a number of possible ideas are created (‘divergent thinking’) before refining and narrowing down to the best idea (‘convergent thinking’), and this can be represented by a diamond shape. But the Double Diamond indicates that this happens twice – once to confirm the problem definition and once to create the solution” (The British Design Council).

The double Diamond model or the 4 D model presented below in figure #1 is a frame to describe the UX design process.

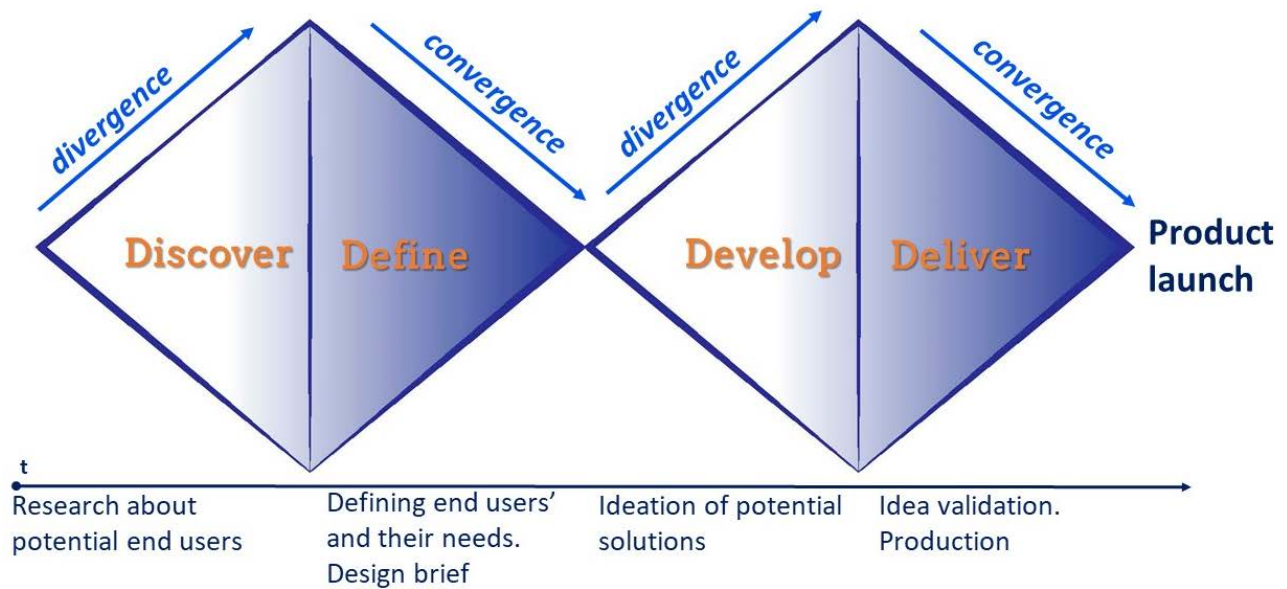


Figure 1 - The UX design process based on the double-diamond model

During the discovery phase, the design team researches the field and collects raw data about potential end users. The phase reflects divergent thinking, because the design team opts for a large amount of information.

Within the define stage the design team analyses all the data gathered previously to create a clear vision of who will be the end users of the product and what are the problems they currently face.

These problems represent business opportunities. The main goal of this stage is to create a design brief with the detailed description of the potential users' profiles and their needs, to frame a design challenge (The British Design Council). During this phase the design team noticeably narrows down the information trying to extract the core from the data gathered previously.

The develop stage could be defined as an ideation process when potential solutions for previously defined problems are created. The third quarter labels the process of iterative development of solutions. The phase represents divergent thinking because the design team prototype and tests different ideas for solutions.

The deliver stage is the last phase of the 4 D model. The stage characterizes by convergent thinking, when the optimal prototype of solution goes final testing, signed-off, manufactured or coded depending on the product type and launched to the market (Tschimmel, 2012).

The double-diamond shows my understanding of the UX design process. Each phase includes a set of different tools that help the design team to progress in the project. The process is flexible and adjustable for different projects from various fields. The flexibility assumes that the design team each time creates a unique set of tools for a specific project. Some tools move across all stages of the process.

To visualize my study scope – the investigation of UX testing methods – I zoomed the “Develop” stage – Figure 2.

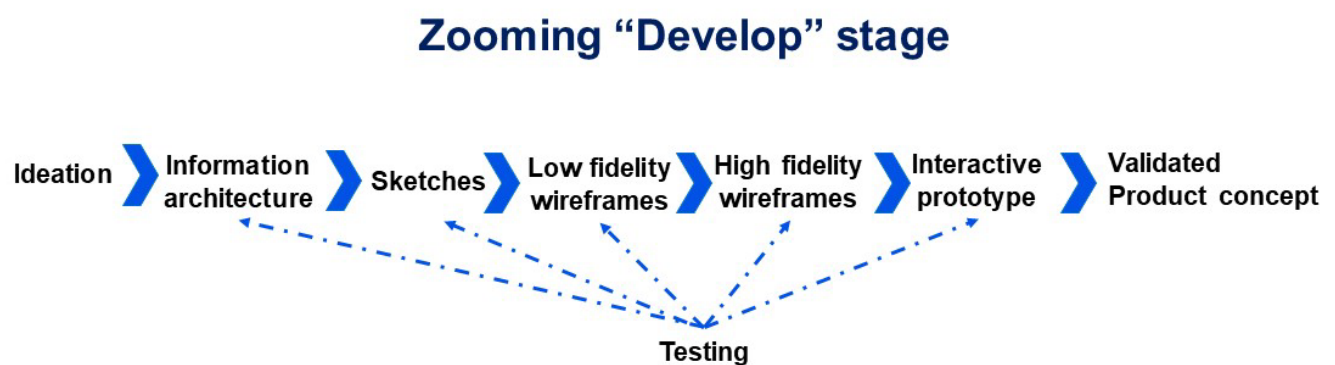


Figure 2 - The different stages of the "Develop" phase if a company is creating a digital product

The picture above shows the sequence of actions that the design team makes when the goal is to create a digital product. If the product or service is not digital, some of the steps will be skipped because they can be attributed only to the digital product. Such steps are information architecture and wireframing. Based on the figure, I show that testing could be carried out at the different stages of "develop" phase. I drew arrows with a dotted line to visualize that testing could or could not be carried out at each step. Design team decides when to complete evaluation. The goal of my study was to define which testing methods are effective at which stages of "Develop" phase.

The core characteristic of design thinking process is that it is iterative. Representatives of human-computer interactions and usability communities agree that an iterative design process is fundamental to create system with high-usability (Terence, 2003). The sequence of sketching, wireframing, prototyping accompanied by tests represents the logic of iterations that enables the creation of a service or a product that will meet the end users' needs.

The result of the "Develop" phase is a validated product concept that a company can use as a brief for the Minimum Viable Product (MVP). Formal definition of MVP refers to the version of a new system which enables a startup team to gather the maximum amount of validated data about potential customers with the least effort (Agile alliance, 2009). The iterative logic of design thinking continues after MVP. The design team uses different testing methods to understand how the market meets the MVP. The next step is to create a Minimum Marketable product (MMP) and beyond. I created figure 3 to represent the iterative nature of the design thinking process.

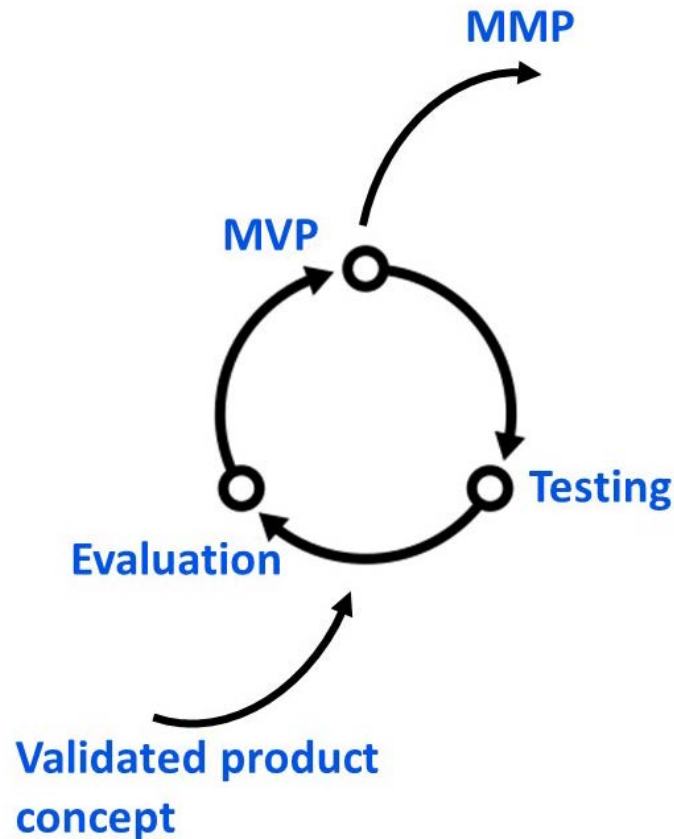


Figure 3 – Iterative process of product development

How many cycles of testing, evaluation and prototyping the design team will make depends on the project goal and constraints such as time pressure and available human and financial resources. The scope of my study was to evaluate UX testing methods within the “Develop” phase and testing methods that are applicable for MVP. However, I did not investigate the method to test a more mature version of product such as MMP.

2.3. UX design in startups

My study was limited to the field of startups and the website’s UX as an object for evaluation. I did not investigate how to evaluate UX of the website in big stable companies that have sufficient human resources to implement accurate UX design for their products. However, I carried out

interviews with UX practitioners from big companies as well as practitioner from academia to compare the startup environment with other environments.

Startup could be defined as a company which intends to design high-tech and innovative solutions. Startup has aggressive plans for scalability (Paternoster et. al., 2014).

The characteristics of startups:

- Competition that results in a time pressure
- Urgent need for growth
- Scarcity of financial resources
- The chaotic, uncertain environment
- Small team, sometimes only the founder

Competition in the market is the prime reason for the urgent need for growth. Startups often plan to enter a new market or disrupt the old one. The first one to do that will have a bigger market share. That results in the race of product development when every startup works hard to be the first to launch its product. “Startups face intense time-pressure from the market and are exposed to tough competition, operating in a chaotic, rapidly evolving and uncertain context” (Paternoster et. al., 2014, p.2).

Compared to big companies, startups regularly experience a scarcity of financial resources. The main challenge is how to be noticed from the crowd when you are young, small and with no money to invest in different activities at full scale. Startups cannot afford to spend a lot of money in the long process of UX design compared to companies with a stable income.

In addition to the scarcity of financial resources, startups have limited human resources, thus a limited team skill set. Majority of startups are launched by solo founder without any employees (Blažica, 2014). This results in a practice common within startups when team members perform multiple roles and quickly learn new skills in order to progress in their startup development.

UX is one of the factors in a startups possible success. UX could be a sustainable competitive advantage in the market. In many cases it is the delighting UX that highlight a successful system from the rivals (Väänänen-Vainio-Mattila, 2008). Without good UX, startup nowadays does not have many chances to attract customers and, consequently, be successful.

The expansion of services and products with high quality UX changed the expectations of modern users towards any new product or system. “The explosion of software-based products into the mass market ...has transformed the user population. Formerly it was a small group of forgiving, technology-loving implementers. Today it is a teeming multitude of impatient, unhappy, nontechnical consumers. Everyone, both inside and outside of the software industry, has heard the users cry in painful frustration” (Cooper, 2004, p.203). Startups cannot afford to end up in the situation of “a crying” user. Since, in that case, the chances for success are low.

The question is how startups should approach the UX design process. Startups usually cannot apply all the tools from the UX design framework visualized in the Double Diamond model presented above. In addition, the crucial point is that a startup cannot dedicate limited resources in developing a product that will not meet customers’ needs in the end. Thus, the culture of experimentation and getting feedback from the customer is becoming a main startup business model: this approach intends to ensure that the final solution will have market potential instead of creating a system first and after trying to find customers for it (Hokkanen et. al, 2016).

Startups have to define crucial steps to take in UX design in order to progress fast. The first important milestone for startups is the development of MVP that reproduces the core functions and is very moderate comparing the vision of the future product. Startups when developing early solution versions while validating the product concept are testing the idea itself and, in addition, specific features and visual design (Hokkanen et. al, 2016).

The idea is to get users’ feedback as soon as possible to avoid wasting scarce resources on developing something that will not meet users’ needs. The iterative nature of UX design helps startups to avoid crucial mistakes and minimize the risk of creating a useless product or service. “The benefits of attending to usability issues through iterative evaluation include improved predictability of the products, greater productivity with fewer user errors, better match with user needs, and savings in development time and cost” (Yen & Bakken, 2009, p.1). This principle came from the Lean startup ideology developed by Eric Ries.

Eric Ries’s approach is commonly acknowledged to be effective, but not commonly used. Still, many startups neglect UX testing to save time and launch the product as soon as possible. In my study I wanted to emphasize that startups need to include UX testing to a certain extent in their UX design process. Without testing, it is impossible to create a good quality UX. However, startups

cannot take into practice full-scale systematic UX evaluation. Thus, there is a need to search for lightweight testing methods suitable for iterative product development process (Väänänen-Vainio-Mattila, 2008).

The reason why evaluation should be included in the UX design process is that the design team that usually consists of designers and programmers while developing the product cannot see it through the eyes of their future users. The labelling that the design team may find clear and precise might be confusing for novice users. The design team might assume that the functions and the placement of certain buttons etc. are obvious. However, for new users, the logic of the system might not be so clear. Thus, to identify the gap between designers' and programmers' point of view on the UX of the system and potential users' point of view, the startup should include UX evaluation in its UX design process. Alan Cooper described the following experience from his practice: "the most valuable contribution of usability testing is made when programmers are forced to sit behind the one-way mirrors to view typical users struggling with their programs. The programmers are shocked and incredulous, shouting sentiments like, "You are testing mental retards" Usability testing is a useful whack on the side of the head for recalcitrant software engineers, showing them that there is indeed a problem" (Cooper, 2004, p.207). This quote shows that people who develop digital products (designers, programmers) could be biased. It is hard to see the mistakes in our own work. However, a startup cannot afford to launch a product with mistakes on the market. Thus, UX evaluation can prevent the startup team from seeing their users "struggling with their programs". In the following chapters I aimed to justify that startups might benefit from applying the framework of Minimum Viable UX evaluation methods to their websites.

2.4. UX evaluation methods

The usability evaluation is the meta term for a set of methods that aim to help product developers and usability experts to find the parts of a system that generate problems for users, slow down the task accomplishment, or does not support users' preferred ways of working — commonly labeled usability problems (Hertzum & Jacobsen, 2010). Usability evaluation or inspection methods are a set of cost-effective approaches of checking user interfaces to reveal usability problems. Usability

evaluation was gaining popularity since about 1990 as a way to check the usability of the user interfaces (Nielsen, 1995).

Evaluation is carried out by usability inspector. Usability inspector could have a special qualification in the field of UX design. However, other people with relevant expertise can also perform the role of usability specialists: programmers, end users, other professionals (Nielsen & Mack, 1994).

Usability evaluation methods differ by many factors. Some of them rely on feedbacks gained from end users of the system. These methods are empirically based. Other are performed by professionals in the field because they require minimum training of evaluators (Heuristic Evaluation). Some methods were developed for usability testing in general (technology agnostic) and only later gained popularity in the IT industry (System Usability Scale). Others, on the contrary, were designed only for digital systems (Heuristic Evaluation).

Usability inspection methods test the usability that is included in the user experience. My research phenomenon is UX evaluation methods. Thus, I researched usability inspection methods and a brainstorming method that is not attributed by scholars to usability evaluation but can be used as a UX evaluation method that I explained later in the corresponding chapter. In the book “Universal methods of design” Martin and Hanington combined usability inspections-methods with a large sample of design methods that constitute the design process in general (Martin & Hanington, 2012). I followed the same logic by investigating how usability inspection methods and brainstorming method can reveal mistakes in the UX of the website. For the purpose of this research, I named the set of methods that I researched - UX evaluation methods.

I analyze the following UX inspection methods:

1. Heuristic evaluation;
2. Cognitive walkthrough (incl. extension as End-User Think-Aloud Protocol);
3. System Usability Scale
4. Tree testing
5. Brainstorming session

My choice of UX evaluation methods for the study was limited. The main foundation for the selection of methods presented above were my studies in Aalto University. All the tools that I researched I learnt during UX training. Thus, facing the need to make UX evaluation of the case company prototype I applied the methods I knew such as all five methods listed above. Further, by scrolling through the academic literature of the topic and titles of articles and books, I found out that some of the methods dominate the scholars' papers: cognitive walkthrough (CW), heuristic evaluation (HE), system usability scale (SUS). Then, I started to research the methods that the UX community often discussed through word of mouth and web sources I noticed that cognitive walkthrough, heuristic evaluation, and thinking-aloud study (TA) are three of the most widely used usability inspection methods (Hertzum & Jacobsen, 2010). Thus, my selection of methods is subjective, limited by the list of the methods I knew before starting to perform the role of UX designer in Sopia. However, I acknowledge that there are many different UX evaluation methods that I did not investigate in my study due to the study's limitation.

The questions I asked myself was how to compare the methods with each other, using which parameters? Nielsen in his fundamental book summarizing usability inspection methods does not provide a unified framework to compare methods. However, he and other contributors of the book analyze each method using similar factors: the qualification and experience of the instructor; input to the method - special preparation before the evaluation in terms of equipment, environment, materials; the number of tests to be done to get reliable results; the involvement of end users; the duration of test; efficiency to reveal usability mistakes and types of problems identified; how the method fits in the design process, the scope of the method and the evolution of the method over time (Nielsen & Mack, 1994).

While analyzing academic articles on the phenomenon of UX, usability, usability evaluation, I found out the repetitive patterns to describe or compare the evaluation methods. To explain these patterns I created the table below. The table is organized in the following way: the first row - refers to the article, the second row refers to the characteristics that authors use to describe methods. I organized the articles in chronological order. I selected articles that describe the methods I was researching. The volume of literature that I used to analyze five evaluation methods is much wider than six articles presented below. However, after reading these articles I concluded that they are based on a common approach to describe evaluation methods.

The table 1 - the repetitive patterns to describe or compare the usability evaluation methods in six academic articles about usability evaluation methods

Article's title and author	The abstract level characteristics of the method(s)
1. Applying cognitive walkthroughs to more complex user interfaces: experiences, issues, and recommendations, by Cathleen Wharton, Janice Bradford, Robin Jeffries, Marita Franzke	Metrics to describe the method: preparation for evaluation, requirements for qualification of test facilitator, the type of data gathered – qualitative or quantitative (Wharton, Je, & Fran, 1992)
2. An Empirical Evaluation of the System Usability Scale, Aaron Bangor, Philip T. Kortum, and James T. Miller	The parameters of the method that authors use to describe the method: how technology agnostic the method is, the involvement of end users in the evaluation process, the type of data gathered – qualitative or quantitative, cost-efficiency and requirements for qualification of test facilitator (Bangor et al., 2008)
3. A Comparison of Usability Evaluation Methods: Heuristic Evaluation versus End-User Think-Aloud Protocol – An Example from a Web-based Communication Tool for Nurse Scheduling” by Po-Yin Yen and Suzanne Bakken	The parameters for comparison are following: involvement of end users in the evaluation process, the type of data gathered, the involvement of experts with special qualification to perform the evaluation, the number of evaluators, the category of end users, preparation for evaluation, duration of the processes (Yen & Bakken, 2009)
4. State of the Art on the Cognitive Walkthrough Method, Its Variants and Evolutions by Thomas Mahatody, Mouldi Sagar, Christophe Kolski	The parameters for comparison are following: cost-efficiency, the category of end users, requirements for qualification of test facilitator, effectiveness to reveal usability mistakes, preparation for evaluation, context and environment during evaluation (Mahatody et al., 2010)
5. Tree testing of hierarchical menu structures for health applications by Thai Le, Shomir Chaudhuri, Jane Chung, Hilaire J. Thompson, George Demiris	Metrics to describe the method: involvement of end users in the evaluation process, the category of end users, the type of data gathered – qualitative or quantitative, duration, special equipment to led evaluation, preparation for evaluation (Le, Chaudhuri, Chung, Thompson, & Demiris, 2014)
6. Comparing Heuristic Walkthrough and User Studies in Evaluating Digital Appliances by Eva-Maria Meier, Patricia Bohm, Christian Wolff	Metrics to describe and compare methods: expert/user study, formative/summative evaluation, quantitative vs. qualitative results, effectiveness and efficiency of the methods (Meier, Böhm, & Wolff, 2017)

I used these patterns to create my research framework that I will describe further.

Nilsen and other researchers (Wharton et al., 1992) of the field described the methods based on three phases: preparation, evaluation and interpretation phases. During preparation phase design team has to carefully assemble everything needed to conduct the evaluation. Some evaluation

methods require special equipment (for example, video and voice recording), some - qualification of the evaluator in the field of usability and UX, other - special environment for evaluation (creative space with snacks, coffee and tea to conduct brainstorming session) etc. From the quality of preparation depends the evaluation itself and the interpretation of results – respectively, the second and the third phases of each method. The evaluation phase refers to the duration of the testing, to the amount of tests, that design team has to conduct, to the profile of the participants of the testing process (usability experts, end users). During the evaluation phase, different types of data are gathered: quantitative and/or qualitative. Data are analyzed further during the interpretation phase (Nielsen & Mack, 1994). Consequently, every testing method could be decomposed on three phases described above.

Another way to describe the UX evaluation methods was presented by Martin and Hanington (2012). The authors combined the methods from different disciplines including usability inspection methods under the umbrella term “design methods”. Their aim was to create a broad overview of different tools. The authors provide a framework to compare the methods on the facets described below.

The first facet is the behavioural/attitudinal parameter that suggests the type of content most appropriately targeted by the method. The second one is the quantitative/qualitative facet that characterizes the form in which that content is typically collected and communicated. The third position to describe the method is how innovative/adapted/ traditional the method is. This facet describes whether the method is original to design, adapted from other disciplines, or used traditionally across disciplines. The forth criteria refers to the primary purpose of the method. How exploratory/generative/evaluative is the method. That affects the phase of design process when method will might bring the most value: early exploration, concept generation, or testing and evaluation. Finally, how participatory/observational/self-reporting/expert review/design process is the method. That facet describes the typical roles of the researcher and participant in evaluation, with design process methods being those conducted by design teams as an integral part of an overall approach (Martin & Hanington, 2012).

From the author's framework, I found relevant for my research the first, second, fourth and fifth facets. I would attribute the behavioral/attitudinal facet to whether end users were involved in the evaluation or not. If end users perform the evaluation, the evaluator observes their behaviour and

based on that makes the judgment about UX issues. However, if experts or design team carry out the test the outcome will be based on the attitude of the evaluator. Consequently, I can relate this facet with the parameter of involvement of end users. The second facet represents the types of data gathered after the testing: quantitative, qualitative or both.

The third parameter is irrelevant for the framework as I have a limited set of methods of which four originated from the field of human-computer interaction (heuristic evaluation, cognitive walkthrough, system usability scale, tree testing) and one method - brainstorming - was developed to stimulate the creative process despite the field of usage.

The fourth facet - exploratory/generative/evaluative frames – is relevant for my research framework because I investigate UX evaluation methods within “Develop” stage of UX process and methods that could evaluate MVP – “Deliver” stage. “Develop” stage is attributed to the concept generation, “Develop” stage to the testing and evaluation.

2.5. Theory-based research framework

From literature analyses, I concluded that scholars did not create a common framework to compare and analyze different UX evaluation methods: “practitioners are far from settled on a uniform UEM (Usability evaluation method), and researchers are far from agreement on a standard means for evaluation and comparing UEMs” (Terence, 2003, p.3).

I did not find one single framework that would count all the parameters that researchers discuss by describing the method or comparing them with each other. Researchers do not have a common vision about the relative merits of the various UEMs (usability inspection methods). Meanwhile new methods keep appearing. The diversity of alternative ways to conduct evaluation and an absence of common perception of the potential and constraints of each method has escalated the need for practitioners and researchers to be capable to define which evaluation tools are more efficient and effective, in what ways and for what goals (Terence, 2003). Thus, my literature analysis helped me to create a theory-based framework that allowed me to logically and consistently analyze the research phenomenon.

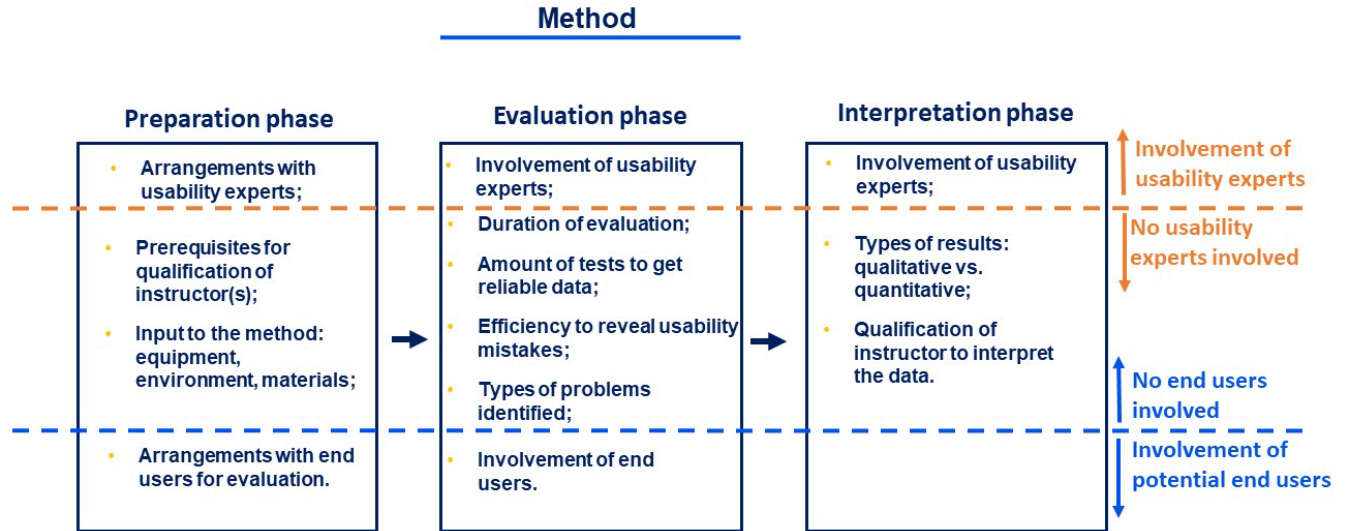


Figure 4 - Theory-based research framework to describe the UX evaluation method

The framework presented above could be used to describe any evaluation method. It includes general parameters of evaluation method that I took from academic literature (Nielsen & Mack, 1994; Wharton, Je & Fran, 1992.; Bangor, Kortum & Miller, 2008; Yen & Bakken, 2009; Martin & Hanington, 2012) and organized in systemic way. Framework consists of three consecutive phases of evaluation process: preparation, evaluation and interpretation phases. In addition, the framework attributes some parameters to participation of usability experts and/or potential end users in the evaluation.

However, as my study was limited to the field of startups and the website's UX as an object for evaluation, I needed to consider in my research framework the general constraints of startups. The characteristics of startups might affect the decision to use or not to use a specific evaluation method. For example, such facets as time pressure and the urgent need for growth could be attributed to the time that the design team needs to apply the evaluation method – the parameter of duration. Another startup characteristic – the scarcity of financial resources – is relevant to the issues of expenses that the startup team might have in order to apply the method: hiring UX experts to perform the evaluation, money spent to get the right group of end users for testing or expenses to set up the right environment to perform the evaluation (materials, equipment etc.).

However, the facet of the chaotic, uncertain environment in which the startup operates cannot be included in the framework because of difficulties in measurement. Such characteristic as “small

team, sometimes only the founder” raises the following question for a startup: does the team need to hire UX experts to perform the evaluation, or is the method easy enough to be learned and applied by novice UX designers and still bring valuable results? That issue is attributed to the prerequisites for the qualification of instructor.

The research framework presented in the previous section is generic – applicable for any kind of evaluation method in any kind of environment. However, my investigations touch only the context of startups. Thus, I visualized how I took into account in my framework the constraints affecting the startup (described in the previous section): the symbol of unicorn refers to the startup constraints. I choose the unicorn icon as it is a commonly accepted symbol of startups. I created my framework by first overlaying two categories of analysis: the set of generic facets describing evaluation methods, and the constraints of the startup.

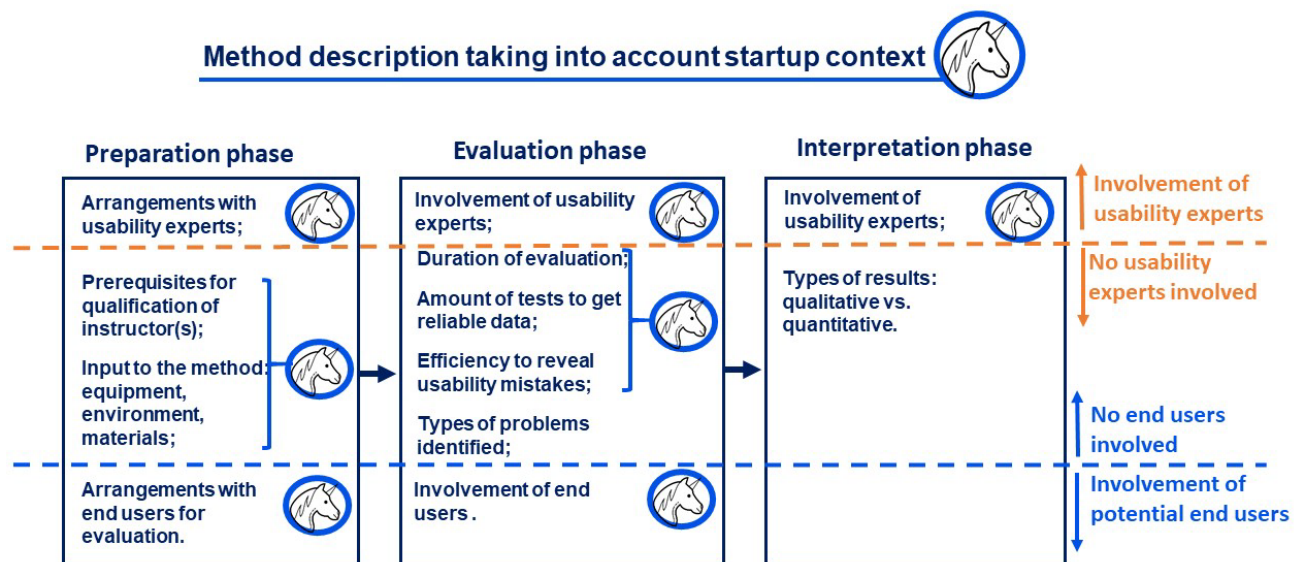


Figure 5 – Theory-based research framework to describe the UX evaluation method involving startup constraints

All evaluation methods that I investigate support the iterative process of UX design: “interactive systems, at least the user interfaces, are usually designed through the iterative process involving design, evaluation and redesign” (Terence, 2003, p. 3). Usability evaluation methods appeared to

help that UX design process by evaluating the usability to reveal usability problems to be revised (Terence, 2003). However, the question of the exact phase (information architecture, sketching, wireframing, prototyping or when MVP is ready) inside the design process is still open. That is why I included the parameter of UX design phase in the research framework. The figure 6 “Theory-based generic research framework to describe the UX evaluation method involving startup constraints in reference to the stage of UX design process” visualizes how to describe the method referring to the specific phase of design process.

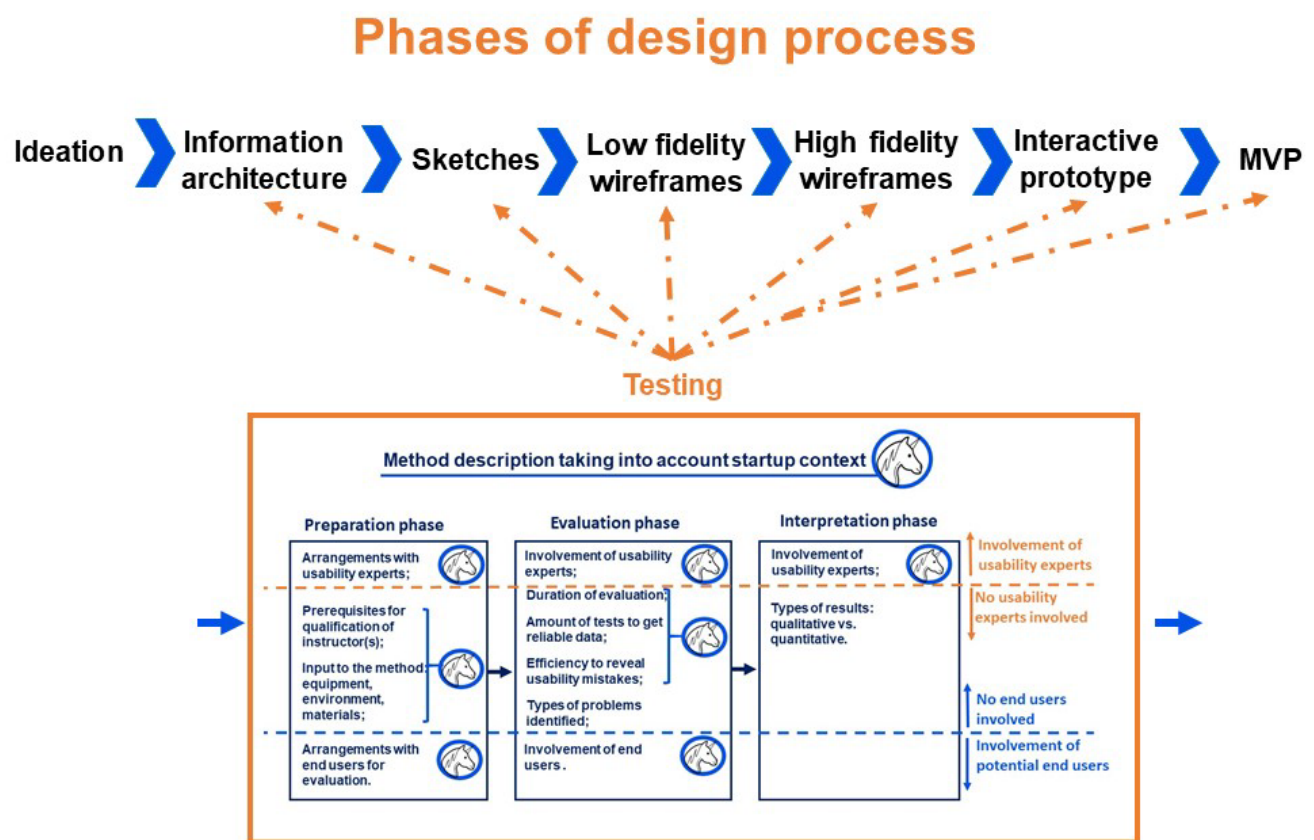


Figure 6 - Theory-based generic research framework to describe the UX evaluation method involving startup constraints in reference to the stage of UX design process

The figure 7 “Zooming into the testing phase as a part of the UX design process” illustrates the bottom of the figure 6. This figure was a basis to describe every UX evaluation method that I

investigated. The figure assumes that by describing any evaluation method I should name the phase of UX design process when the method is applicable.

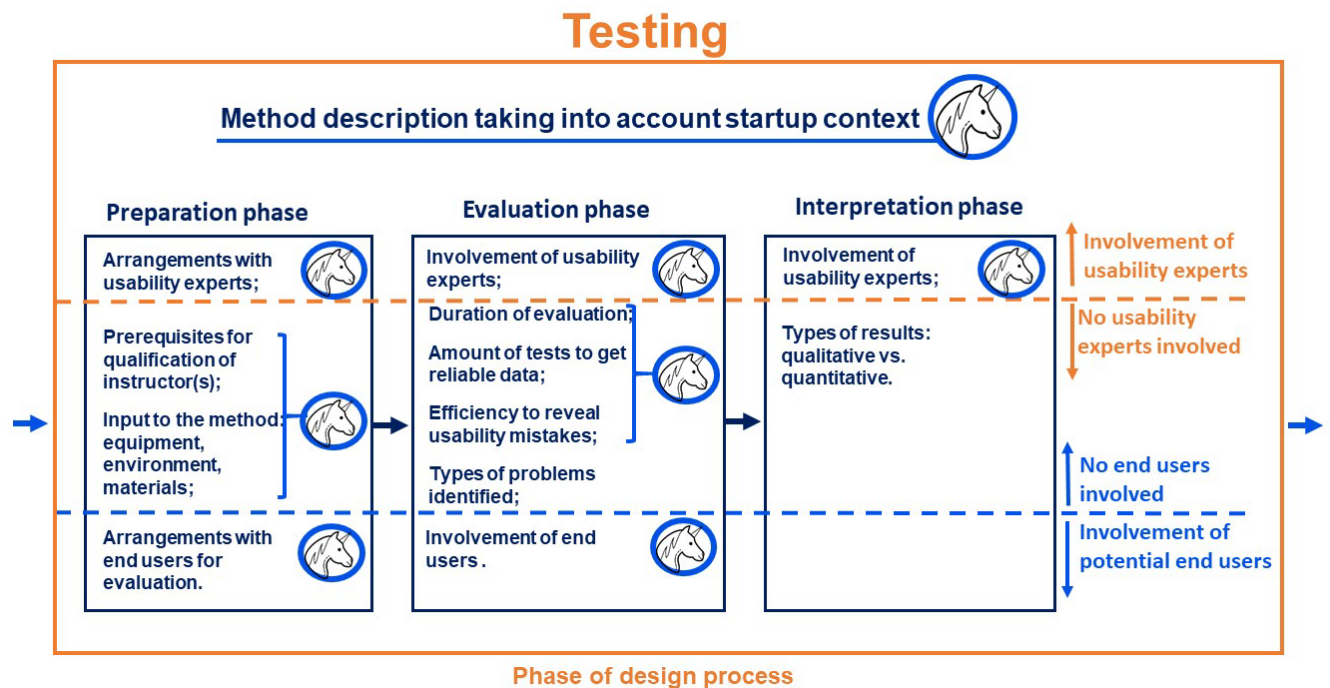


Figure 7 – Zooming into the testing phase as a part of the UX design process

I analyzed and compared every UX evaluation method based on the framework presented above. Every method assumes involvement or absence of end users to test with. Each method could also be described by the need to involve usability experts or for the members of design team to have qualifications in the field of UX. Other generic methods are the duration of evaluation, the number of tests to get reliable data, the types of data gathered for interpretation - quantitative or qualitative. For empirical study I organized the questions for interviews with UX experts to be able to collect the data that covers all the elements described in the framework.

3. METHODOLOGY

3.1. The basic logic for the methodology

My approach towards methodology for the study resulted from my involvement in the case company. I was heavily involved in Sophia as I was working there as a UX designer. This means that I had to evaluate our team's work in UX design that was beneficial for the quality of outcomes that we delivered. As a researcher and practitioner at the same time I was able to reflect the things that I had implemented in Sophia. In addition, I felt empathy towards entrepreneurs trying to create products with a great UX when subject to time, money and skill constraints.

My professional background is in design thinking and service design. I was working as a service designer in different projects in Finland and Russia. I have an education in design thinking and UX design. However, Sophia was the first place where I was working as a UX designer. UX design includes specific tools related to the digital nature of the process such as UX testing tools. Thus, one of the drivers of my study was the will to investigate these tools in depth with the aim of mastering them in practice in different UX design projects. That impacted my choice of research methods.

In the majority of academic papers in the topic of assessment of UX, the authors tend to choose quantitative research methods. However, the general research approach of this study is qualitative because it offers a possibility to examine the phenomenon in depth and with open-ended questions without predetermined answers (Creswell, 2009). It is claimed that qualitative research methods have been effective in social sciences (Berg, 2001). My choice in favour of the qualitative research method was justified by the observation that qualitative research methodology "refers to the meanings, concepts, definitions, characteristics, metaphors, symbols, and descriptions of things" (Berg, 2001, p.11). In my study I do not want to measure the effectiveness of certain UX testing methods, but find the answers as to why some methods are effective and some are not, why some should be used in the startup environment and some should not. These answers lead to the research objectives of this thesis – the conclusion of "how to evaluate UX of the website if you run a startup?". Such questions as "what?" and "how?" are typical for the qualitative research approach. That requires in-depth analysis based on my interpretations and my understanding of the research subject that would not be possible with the use of quantitative research methods.

To collect qualitative data I used the method of case study. “Case studies are a design of inquiry found in many fields, especially evaluation, in which the researcher develops an in-depth analysis of a case” (Creswell, 2009, p.43). The purpose of the case study research is to get the answers on questions such as how or why (Aberdeen, 2013). As case study allows interpret both the process and result of the process of an analyzed phenomenon through observation, reconstruction and study of the cases (Zainal, 2007). The case study methodology suits the purpose of my study to analyze in-depth the researched phenomenon of UX evaluation based on a collected set of data.

My research is based on a single-case study – the evaluation of UX of the web site of case company Sophia. In addition I made multiple interviews to collect qualitative data about the application of analyzed methods is seven cases: five startups, one academia and one established company. Thus, my research methodology of each testing method has three pillars: first, I studied UX testing methods based on the literature review examining scholars’ papers on the topic. I provided the academic definition of the method and investigated the origins of the method. The second pillar of my approach involves the data collected from interviews with usability, UX and UX evaluation practitioners. To gather empirical data I asked interviewees which methods they find useful for a startup to evaluate UX and why. I structured my interview questions to gather the data that will allow me to analyze every UX evaluation method using my theory-based research framework described in the previous chapter.

The third pillar represents a single case study – the application of analyzed methods on UX of the prototype of the website of the case company – Sophia. I used each method that I described in the theoretical part to evaluate the UX of the Sophia prototype. In case study methodology the researcher should define the criteria for analysis (Aberdeen, 2013). I analyzed the results of applying UX evaluation methods and concluded how effective that particular method was for the Sophia case. I recorded test results and they can be found in the Appendices. I decided intentionally not to use some methods for Sophia and I provide the explanation later in this study. The Sophia UX design process was based on the design-thinking methodology. Thus, it followed the phases of UX design described above in the section “UX design”. Consequently, I analyzed every evaluation method in relevance to the following phases of UX design process such as information architecture, sketches, low fidelity wireframe, high fidelity wire frame, interactive prototype, and MVP (see the section “UX design”).

Interviews with field practitioners and study of Sopia case allowed me to get empirical data on the UX evaluation methods. The idea was to compare empirical results to relevant theoretical knowledge in the field. The empirical study approach fits the research objectives because it helped to verify or deny assumed conclusions that I made after the theory-based description of each method. However, the fundamental academic literature about UX and usability inspection methods date back to the 1980-90s and since then the web systems have changed dramatically. Thus, I needed to obtain up to date knowledge of how practitioners evaluate UX of modern websites. In order to achieve this, I carried out the expert interviews.

My research approach structure is visualized in the Figure 8 below.

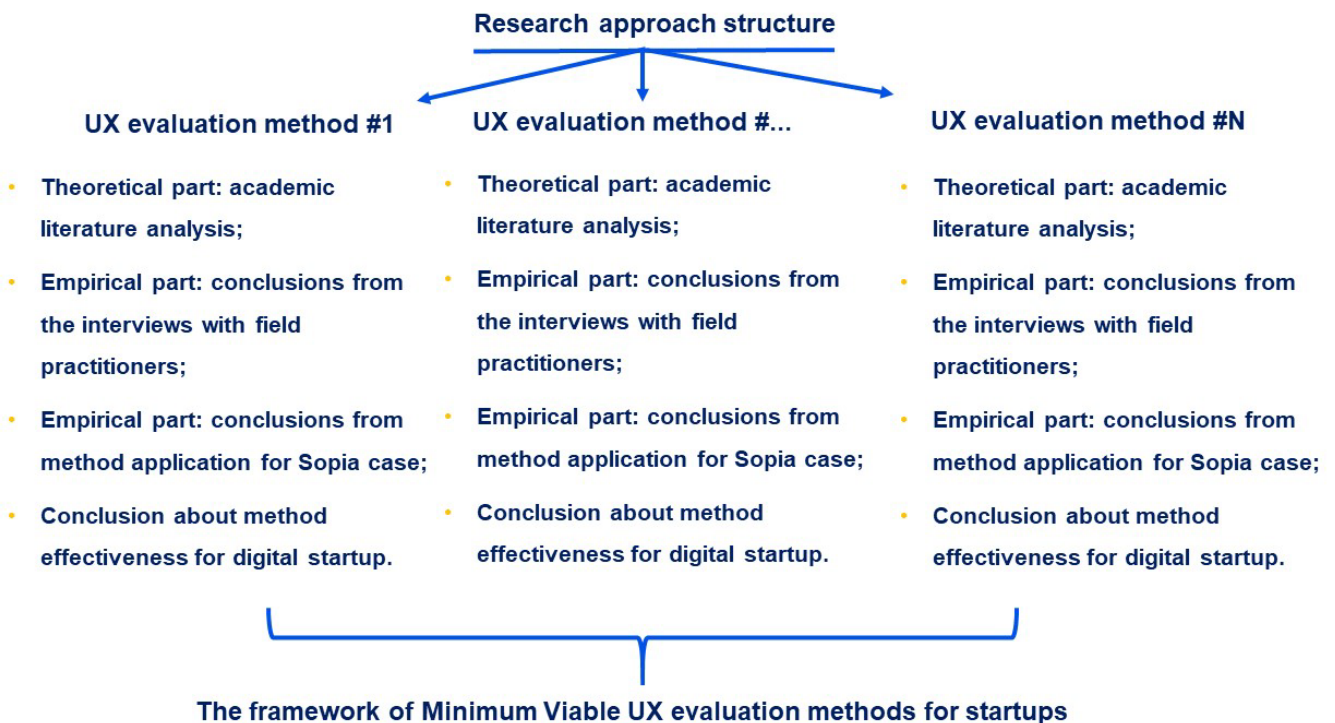


Figure 8 - Research methodology structure

My research methodology affected the structure of the thesis, changing it from the traditional sequence of theoretical and empirical parts. In my study, the theoretical and empirical parts are mixed. The logic of the study is structured based on methods. I choose the method, carried out

theoretical research of the method origins, formal definition, reasons for using it; then through a series of interviews I analyzed how practitioners use or don't use the method and what is the reason; and, lastly, I applied the method in the Sophia case. I repeated the same process for each UX testing method that I was investigating in my study or I explained why I did not apply the method in the Sophia case. Thus, my study has sections that I labelled by the names of the methods. Each section ends with the conclusion about method effectiveness to evaluate the website if the company is a startup. The conclusion is grounded on three levels of analysis: academic literature analysis, interview results and analysis of method application for the Sophia website. As an outcome of my research, I aimed to develop the framework of Minimum Viable UX evaluation methods for startups based on conclusions from each section describing the methods.

To follow the structure described above I need to introduce the Sophia case before evaluation methods analysis, otherwise it might be difficult to understand how I applied methods for the prototype of the Sophia website. The fact that I tested methods on the prototype of the website set the frames for my study in a way that I did narrow down my research only to those methods that could work for the digital products and among digital products only for websites.

An interesting observation about the availability of academic literature about UX evaluation methods is that the most commonly cited articles and books from the field date back to the 1990s. The recent academic literature about UX evaluation methods is relatively limited. That puts constraints on the analysis, as the field changed dramatically from the 1990s. Consequently, in investigating the origins of the methods I referred to the literature from the 1990s and after, but to understand current practice I referred to the data from interviews with field practitioners currently working on UX design.

4. DESCRIPTION OF SOPHIA

All startup constraints described above are fully applicable to the case company – Sophia. Sophia Oy is a startup aiming to operate in the rental housing market in Finland. I took Sophia as a case study for this study for many reasons. The first reason is that Sophia is a digital platform that matches tenants and landlords and its website is the main product. My work in Sophia and the practical need to find appropriate UX evaluation methods was an initial stimulus to start to investigate the topic

of this study. I was in the company almost from the first days of its existence and can track its progress. I also had access to information about the company.

In the case of Sophia, the website is the platform for the rental housing market in Finland. That means that the website is the main product, not just a digital representation of Sophia's contacts. In Finland currently, there are many platforms operating in the same market. Thus, the high-quality UX could be one of the ways to distinguish Sophia from its competitors. The UX of Sophia's website is crucial for business development.

The idea to create Sophia came from company founders¹ who rent out their apartments. Being private landlords they assessed the current process of renting out the apartments to be old-fashioned, ineffective and rigid. To solve this problem, they decided to create a startup. The foundation of the startup is the algorithm that matches the tenant's wishes about the apartment to rent and landlord's offers.

4.1. Description of Sophia's UX design process

The actual work on Sophia development started in May 2018, when Sophia received funding from the Kiradigi project and its founders created a team to work on the company's progress. The team consisted of two founders, two service designers (me as one of the designers) and two programmers. Later, closer to the stage of MVP development, the team became bigger: two more coders and a UI designer were involved. The service design team became the UX design team when we started to work on the prototype of Sophia's website. In the UX design process I took decisions and responsibility for what kind of testing methods to use. I had relevant education but did not have practical experience in UX design. The rest of the team was working on other tasks to develop our startup and I was delegated the responsibility for making decisions about testing methods. Our service-design team reported the results of our work to the whole startup team.

From the beginning of the startup development, the whole team decided to stick to the logic of the UX design process that I described in the section "UX design". To visualize the Sophia UX design process I made the following figure:

¹ I will not use their names in the Master Thesis

Sopia's development life span

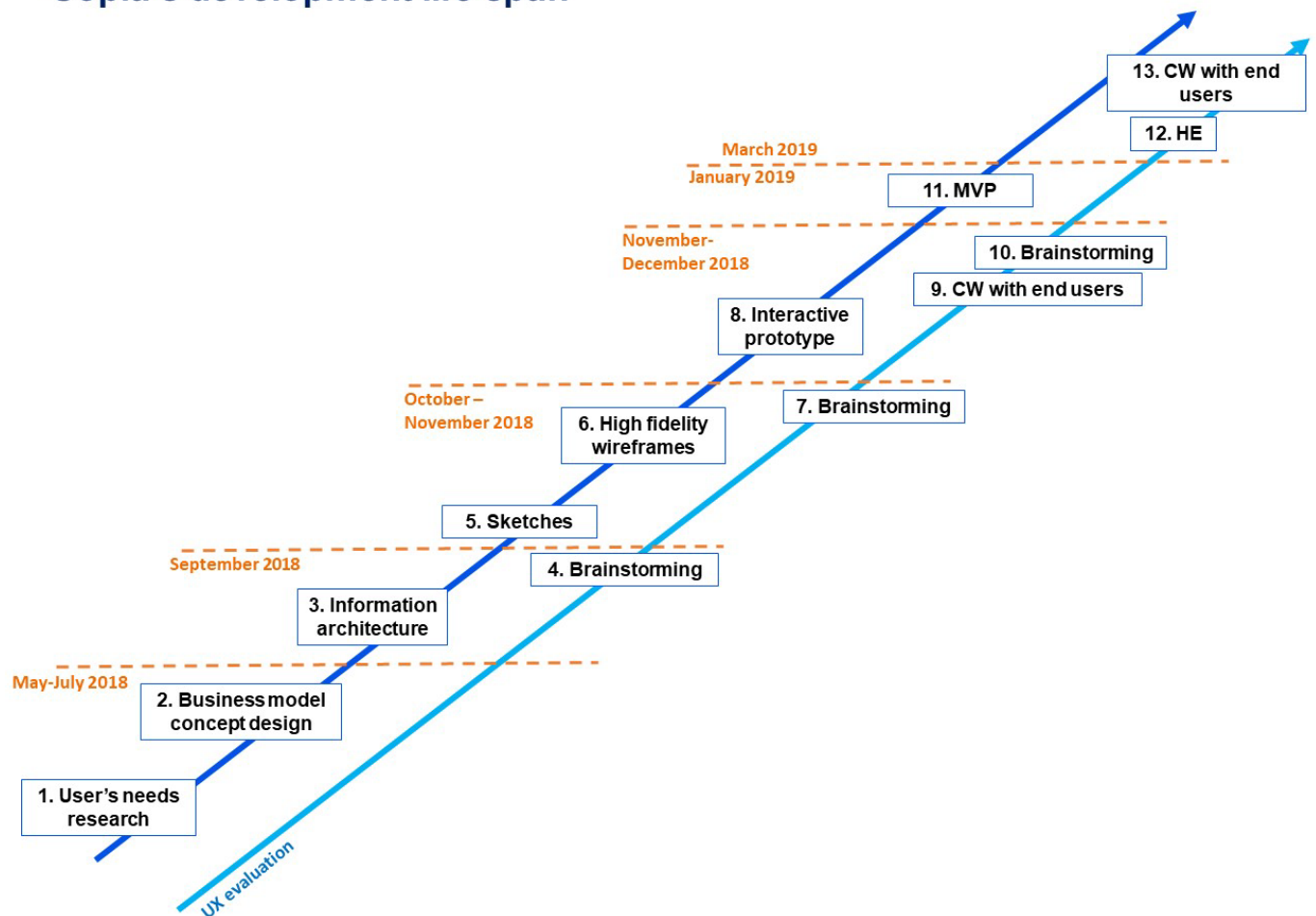


Figure 9 – Sopia's development life span.

The bright blue arrow on the left shows the sequence of UX design actions that Sopia's team made in order to proceed with the product development. The light blue arrow on the right shows the UX evaluation methods that we used. I added the numbers to each action to illustrate the order of UX design steps that we made. However, it should be taken into account that not every category represents a single action. For example, the #1 user's needs research represents 33 interviews or #9 CW with end users means five tests. I added the arrows to visualize that further development is still ongoing, as well as UX testing. The goal for further years (2019 - 2021) is to understand the flaws of the current MVP and to create a better version. The scope of this thesis was to analyze the process until the creation of MVP before January 2019. However, after MVP was ready the design team continued UX evaluation of the MVP (#12 and #13). Below is a description of the

whole process of Sophia development. In the description I use “design team” to refer to myself and my colleague as UX designers in this project. When I use “big team” I mean all people who were involved in Sophia’s MVP development.

We as designers started with user research. We made 33 open-ended interviews with tenants and landlords. Our aim was to cover different profiles of potential users of the platform. From the tenant side we interviewed foreign people, employed, unemployed, students, people just divorced, people living alone, families with children, etc. From the landlord side, we interviewed landlords from different regions of Finland, landlords with many apartments to rent out and landlords renting out their home when they move abroad. The focus of interviews was to understand current issues that both parties face while renting or renting out the apartments. After interviews, we organised a team-workshop to analyze the data that we gathered – a brainstorming session with an affinity diagram exercise.

The goal of the user research phase was to get to know current problems on the market and be able to design our solution in order to solve these problems. As soon as we had the user insights summary, we organised a workshop (brainstorming session with a modified Business Model Canvas exercise) with all team members to create a concept of our solution and to think about the business model that will support it. Also, at that point, we decided on the scope of our MVP which would include only a few core functions.

Our next step was the design of information architecture (IA) (#3) – “backbone of the site” (Nielsen Norman Group), a helpful tool to visualize which feature follows what, what are the features on a specific page, and what are the steps that the user has to take to accomplish the task. The IA is a very useful tool for the whole team to understand the website architecture. In the Sophia case, the design team created the IA and presented it to the big team. The IA provided a clear visualization of the concept. The whole Sophia team performed brainstorming (#4) several times to test the IA, to allow Sophia project to proceed on to the next stage.

After the approval of basic IA suitable for MVP, we started to make separate sketches for the landing page (#5). We created six versions of the landing page interface design, using Adobe XD as the main software (#6). After this, we arranged the workshop (#7) with the whole Sophia team to choose the most appealing interface design for our solution. During the workshop, the team decided to concentrate on two particular designs: the most consistent and traditional for the rental

housing market, providing the feeling of home and peace, and the abstract, that aimed to create the emotion that the user is dealing with an innovative solution. Based on the decision made, we designed two interactive prototypes with two different interface designs (#8). “The traditional” version was supposed to help the tenant to find an apartment, the “abstract” version was guiding the landlord (see in appendix). Simultaneously, we started to test interactive prototypes on end users (#9). The design team started with a cognitive walkthrough – we gave the users tasks and observed how they completed them. However, the first testing already revealed serious mistakes in prototypes that blocked the possibility of accomplishing the task. This stage resulted in a way that the design team carried out prototypes’ improvement after each testing. Sophia’s team chose the core interface design based on the user’s feedback. In November 2018, a UI designer joined our design team and made a professional high fidelity prototype with the chosen interface design.

The design team had to gather all possible feedback and information about UX mistakes in the prototype to analyze them and provide a detailed summary to the programmers before they started to code the MVP. To summarize all the findings we organised two brainstorming sessions with the big team (#10). As soon as the big team approved the summary, the coders started to build the actual product. MVP was almost ready in the middle of January 2019 (#11). However, after the team tested it, team members found several bugs that went back to the programmers. I carried out a heuristic evaluation of our MVP simultaneously with the new set of CW (#12 and #13).

The Sophia team created the first version of MVP over nine months, which is a good pace, especially taking into account the team size and that team members didn’t work full-time. At the time when I am writing my master thesis, Sophia’s team is working on MVP improvement and, simultaneously seeking funding. Sophia’s UX design process and Sophia’s typical startup constraints represent a good empirical environment to test the tools that I am researching. The conclusions that I drew after investigating the applicability of certain tools to the Sophia’s product could be extrapolated to a certain extent to other startups developing their website.

5. SELECTION OF INTERVIEWEES AND THE LOGIC OF INTERVIEW QUESTIONS

I described above that my research approach had three pillars – the theoretical study of UX testing tools and two empirical studies: the Sopia case and the data from interviews with field practitioners. When I was contacting people for the interviews, I chose those who work in startup companies at different stages of development and have to develop UX design for them or for the company clients. I did not focus on the industry where the company operates. UX practitioners I interviewed were not willing to share their names and asked not to mention their companies' names and companies' detailed description in this study. Thus, I provided very limited information about interviewees and used the number while citing their comments.

Besides interviews with people working in or running startups, I interviewed Evgenia Litvinova who was working as UX research assistant in Aalto University and was responsible for the product development cycle with a strong emphasis on UX. In addition, I interviewed Mikael Leppä, Design Director at Wärtsilä. He leads their UX design team. Wärtsilä is a big established Finnish company that manufactures and services power sources and other equipment in the marine and energy markets. The company was established in 1834 and has around 19,000 employees. The company is located in 70 countries. Wärtsilä works in a stable and predictable market that is far from the vulnerable startup environment. I interviewed Evgenia and Mikael to compare the practice of UX evaluation in small startups with the educational environment and a big company without startup constraints and thus able to allocate more resources to UX design.

Table 2 – Description of interviewees

#	Tester job description	Company description	Company localization
Startups			
1.	<ul style="list-style-type: none"> • User Research; • Ideation; • Workshop; • Design Audit; • UX/UI design. 	A startup that operates as an online platform that provides a software testing service. Growing startup. 60 employees.	Tallinn, Estonia
2.	<ul style="list-style-type: none"> • Academic research; • Benchmarking; 	A startup from the well-being industry that provides well-	Helsinki, Finland

	<ul style="list-style-type: none"> • User interviews; • A/B testing; • Wireframing; • UI designing; • Interaction designing; • Prototyping; • Copywriting. 	being devices and a platform to support it.	
3.	<ul style="list-style-type: none"> • UI design; • User testing; • User interviews; • Prototyping. 	A startup providing a financial service platform for entrepreneurs and freelancers.	Helsinki, Finland
4.	<ul style="list-style-type: none"> • UI design • UX design • Programming 	A startup providing an application that enables easy to make 3D modelling and design. The website works as a marketing channel. Currently 8 people in the team.	Helsinki, Finland
5.	<ul style="list-style-type: none"> • Sketches • UX design • Programming 	A startup that sells Chinese manufacturing equipment, maintenance and training. The website works as a shop.	Helsinki, Finland
Educational environment			
6.	<ul style="list-style-type: none"> • User research • Sketching • Wireframing • Interactive prototyping • Usability evaluation 	Aalto University. However, the work of the interviewee was similar to the work in a startup.	Espoo, Finland
Big established company with a stable market			
7.	<ul style="list-style-type: none"> • Director of Design team with a strong focus on UX design 	Wärtsilä	Helsinki, Finland

I interviewed seven UX practitioners. The questions I was asking had the same logic as the theory-based research framework I created to conduct my study. I asked separately about each method I was investigating. I asked the same questions for each method that I analyzed. The questions were intended to collect the opinions of practitioners about the usefulness of each method taking into consideration startups constraints. The main body of the questions derived from the generic

parameters of the research framework. I considered the practitioners as experts because their everyday work includes UX design and UX evaluation; they have practical knowledge in the field. The questions were open-ended without embedded answers. This approach allows the collection of qualitative data to align them with overall research logic.

The table #3 below represents the questions I was asking and explains how they are attributed to the theory-based research framework.

Table 3 – Interview questions

The question	The parameters of the theory-based research framework
What is your opinion about the usefulness of the method for a digital startup?	The question is attributed to the constraints of startup such as time pressure, scarcity of financial resources, scarcity of human resources to make evaluation.
When is it a good time to apply it (to the existing product, an early prototype, MVP or beta version of the product)?	The question refers to the phase of UX design process and asks about the phase when the investigated evaluation tool could be used: information architecture, sketches, low fidelity wireframe, high fidelity wire frame, interactive prototype, and MVP.
How effective is the instrument to reveal usability problems?	The question opens up the issue of interpretation of evaluation results. Through this question I was able find out what kind of data could be gathered during the evaluation: qualitative or quantitative? How easy is to interpret them? Types of problems identified and how many tests should be done to get reliable data?
How cost-efficient is the method?	<p>The question covers all three phases of the evaluation process: preparation, evaluation and interpretation. If the method assumes involvement of end users, that means that in preparation phase design team has to make efforts to arrange the group of end users to conduct evaluation. That resulted in additional costs and time. Another issue is the requirement for qualification of instructor(s) – could the design team conduct the evaluation on its own or should it arrange the involvement of usability experts? The question also touches on the prerequisites for the input to the method such as equipment, environment, materials.</p> <p>Referring to the evaluation phase the question asks for data to cover the following parameters: duration of evaluation and</p>

	<p>amount of tests to get reliable data. Both characteristics will influence the costs needed to conduct evaluation.</p> <p>In the interpretation phase the involvement of experts to interpret results will also affect the cost of the method.</p>
Is there any software that can substitute the method guided by evaluator?	The question refers to the limited resources of the startups. I am asking if there are any means to make evaluation cheaper and faster using specific software.
Do you apply the method at work? Could you elaborate why you apply or you do not apply it?	The closing question brought the information about UX designers' practical experience in using or not using the researched tool. I applied gathered information to compare the usage of UX testing methods among practitioners

I am aware that the sample of interviewees is not fully representative: respondents work in different fields and in startups at different stages of development. I interviewed only one respondent (respondent #6) who was working in the University environment, and only one (respondent #7) who is working in a big company. However, interviews with UX practitioners form only one part of empirical analysis and the second part is supported by the Sophia case. Another point is that already seven interviews allowed me to notice common patterns and practices of developing UX in the case of startups, university and a big company. Interviews brought me valuable qualitative data that was the point of this research. The comparison of the respondents practice with our process in the Sophia case allowed me to make conclusions based on gathered information.

6. STUDY OF UX EVALUATION METHODS

6.1. Heuristic evaluation

6.1.1. Academic literature analysis of HE

Jakob Nielsen and Rolf Molich presented heuristic evaluation (HE) in 1990 as usability evaluation method where usability experts check an interface of the system and try to define if the design is good or bad, or if there are any interface design mistakes (Molich & Nielsen, 1990). Nielsen attributed heuristic evaluation to the “so-called discount usability engineering” methods that do not require excessive resources to be carried out. He claimed that heuristic evaluation is “cheap, fast and easy to use” (Nielsen & Mack, 1994, p. 25). Nielsen advocated for the use of heuristic evaluation as a minimum usability inspection method if the developer team finds other UX

evaluation methods to be intimidating, too costly, and too effortful and time-consuming to apply (Nielsen & Mack, 1994). He defined ten "heuristics" – ten general principles for each digital product. His main idea was that each website should work based on these heuristics. If the website does not respond to all of the heuristics that means that the usability of the system is not good enough. The system compliance with ten heuristics is a minimum requirement to create good UX, and the starting point to create great UX. “HE is guided by heuristic principles to identify user interface designs that violate these principles” (Khajouei et al., 2017, p.2). These heuristic maxims are used by evaluators as a framework to reveal the potential UX issues users may face.

The “traditional” list of the heuristics is presented on the website of Norman Nielsen group (1994) – the UX consulting company organized by two core figures in the field of UX design – Jakob Nilsen and Don Norman (Norman Nielsen group, 1994). In this study, I referred several times to the materials from this website because I found them relevant as both creators and editors of the website are considered important contributors to the field of UX design and UX evaluation methods.

Table 4 – the interpretation of heuristics

The name of heuristic	My interpretation
1.Visibility of system status	The system should communicate the current state to the user that he/she will control the process and take appropriate actions to reach the goal.
2.Match between system and the real world	The terms (the meaning of functions, labelling of buttons etc.) that the system uses to communicate with the users should be familiar to the user and should not differ from common conventions in order not to confuse the user.
3. User control and freedom	The system should provide the user with the possibility to leave the unwanted domain, category and service. The system should support undo and redo functions.
4. Consistency and standards	The system should have a consistent style and vocabulary. Every concept should have one concrete definition, not a multitude, to prevent user confusion.

5. Error prevention	Sometimes users make a mistake. Before choosing a critical action (pay, register, upload etc.) the system should double check that the user is confident to perform this action.
6. Recognition rather than recall	The system should include visual elements that help users easily recognize the concepts (pictures, common phrases, icons etc.) rather than try to extract the information from memory (recall)
7. Flexibility and efficiency of use	The heuristic that takes into account one of the basic concepts of usability – easiness to learn. Designers should design the system in a way to smooth the process of first-time usage for the novice user and to help him/her to learn the system from the first usage. Simultaneously, the system should have an accelerated process to accomplish the task for experienced users.
8. Aesthetic and minimalist design	The system should incorporate simple, clear and consistent information and visuals rather than confusing excessive information and irrelevant functions in all possible formats.
9. Help users recognize, diagnose, and recover from errors	If the user makes a mistake, the system should provide him/her with a polite, precise and visible message about the error and the ways to correct it.
10. Help and documentation	The system should provide help and guidance in different formats: text, videos, chat box etc.

I perceive heuristic evaluation to be a simple inspection method because the "heuristics", that Jakob Nielsen created, seem to be obvious. It is easy to understand and agree with the ten heuristics because they are attributed to people's common sense when we are using the website. Heuristic evaluation is usually carried out by usability experts reviewing the design of a user interface and making an opinion about its compliance with a set of predefined heuristics. The outcome of heuristic evaluation depends on the qualification of people performing the evaluation (Khajouei et al., 2017). Jakob Nielsen states that anyone can perform that kind of evaluation after special training, not only experts: the method is so easy that anyone can learn it during a half-day seminar (Nielsen & Mack, 1994).

However, to get valuable results from the evaluation, three to five people should separately perform heuristic evaluation of the product and compare the results after evaluation. One person is not enough because the results will be highly subjective.

Curve showing the proportion of usability problems in an interface found by heuristic evaluation using various number of evaluators.

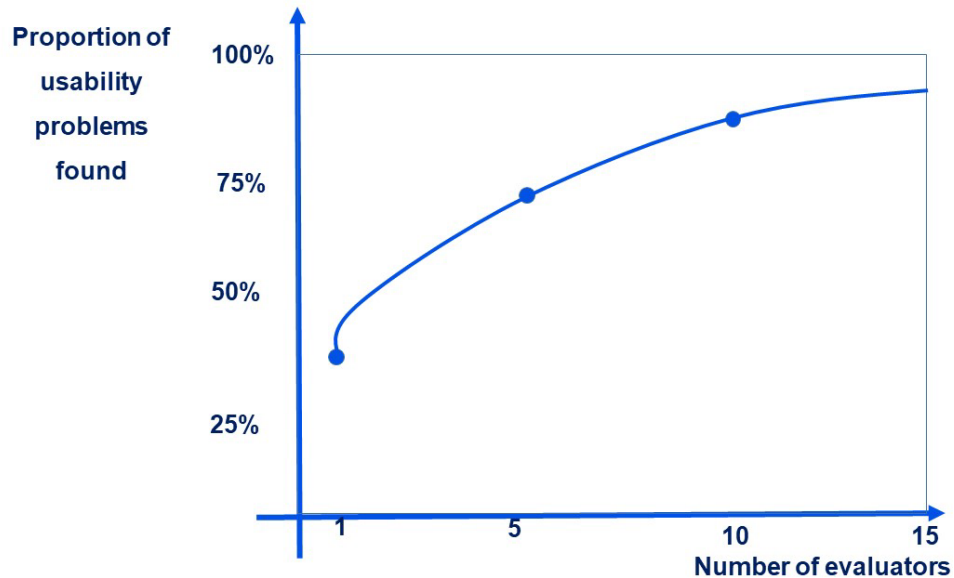


Figure 10 - The suggested number of evaluators to perform heuristic evaluation (Nielsen, Mack, 1994, p. 33).

Based on the curve made by Jakob Nielsen, already one expert can reveal around 40% of usability problems. Five experts will cover 75% of the problems. However, based on the curve it is almost impossible to cover all the problems. The reason is that “it is definitely true that part of usability issues are so easy to reveal that they are exposed by almost anyone, but there are also certain issues that are found by very few experts” (Nielsen, Mack, 1994, p. 26). From that, it could be concluded the HE alone will not be sufficient to understand the flaws of the website and it should be combined with additional UX evaluation methods.

6.1.2. Empirical part 1– conclusions from the interviews with practitioners.

Out of five respondents representing startup community, three were familiar with heuristic evaluation and one was using it at his work. The other two respondents were not familiar with the method. Those interviewees who know the method but don't use it expressed a general comment

that it is a specific method that they don't have time, resources and motivation to learn and start using in their job. The quote from respondent #1, who knows the tool but doesn't use it: "For the startup environment, there is no time and resource to check every detail one by one. Most of the time we only check a few items and fix the bug". The respondent pointed out the common situation for startup teams – resources scarcity. The startup is oriented to develop MVP as soon as possible and that puts a limitation on learning and exploring different UX testing methods including heuristic evaluation.

The opinion of the respondent who uses the method is the opposite of those interviewees who do not: "Heuristics can be risky but come with the benefit of speed. Startups are inherently risky and you must move fast so heuristic evaluation is often useful" (Interviewee #2). I can assume that the reason why the majority of startups I interviewed did not use heuristic evaluation is their limited knowledge about UX testing tools variation and the lack of time and motivation to study the available tools for testing UX.

I asked the same questions about heuristic evaluation to the member of staff working in Aalto University (Interviewee #6) and to the UX director in Wärtsilä (Interviewee #7). Both knew the method. The Aalto staff member answered that she used the method to evaluate the prototype when she didn't have any resources to conduct tests with end users. The respondent from Wärtsilä said that his team didn't use the method as such because "We just design everything having them (heuristics) in mind. We incorporate them in all the templates from sketches to a clickable prototype. It is natural for us" (Interviewee #7). However, he pointed out that he thought knowledge of the 10 heuristics is very important for the startup environment. It is the method "that you can use without end users. It is less time consuming and requires less commitment – so quite cost-efficient" (Interviewee #7).

Thus, from all the seven interviews, two respondents use the method: one employee of a startup and a member of staff working in the university environment. Those two respondents that use the tool in their work both mentioned that the method is effective to apply in the early stages of product development when the design team can easily make changes. Both respondents and the representative of the big company defined the method as cost-efficient, meaning that the benefit of using the method is higher than the resources spent to run it.

One respondent pointed out that heuristic evaluation is more effective in revealing issues related to the interface design (ID), not exactly usability (Interviewee #2). Another interviewee, on the contrary, mentioned that heuristic evaluation is not very effective in revealing User Interface (UI) problems because of the big difference in how experts who perform evaluation and the user perceive the product. Thus, she suggests testing the aesthetics of ID with end users instead of using heuristic evaluation performed by experts (Interviewee #6).

The same respondent pointed out that she applied heuristic evaluation when she did not have access to the real users and she found herself in the situation when she should evaluate her own work as a designer with heuristic evaluation. She said that applying heuristic evaluation was better than doing nothing, however, “it was indeed very hard to be critical and evaluate your own ideas” (Interviewee #6).

The representative of the big company said that heuristics could be a good communication tool when the design team has to explain the website configuration to external people: partners, stakeholders etc. “We refer to heuristics when we talk to the stakeholders. We point back to heuristics that it is an internationally accepted UX practice. We use it as evidence” (Interviewee #7).

One question in the interviews concerned the validity of the original version of ten heuristics developed by Jakob Nielsen in 1990. The Startup practitioner answered that he is using the modified version of heuristics (Interviewee #2). The staff member at Aalto University pointed out that she used the original version, not because it was optimal, but because as she said “That was rather at the beginning of my career, so I didn’t feel comfortable changing the original heuristics and used them as they were” (Interviewee #6). The representative of the big company thought that the original version was valid (Interviewee #7).

Final important observations from the interviews – two respondents who used the method did not invite external usability experts to perform the evaluation. In both cases, the evaluation was carried out by internal resources of the design team (Interviewee #2) or solely by UX designer (interviewee #6).

6.1.3. Empirical part 2 - Heuristic evaluation of Sophia MVP

In the case of Sophia, we started by thinking about the heuristics of the Sophia website prototype already at the stage of low fidelity wireframing. I had an education in UX, so I studied heuristics before. We tried to keep in mind all the 10 basic heuristics developed by Jacob Nielsen. However, I performed real heuristic evaluation in the phase after MVP development.

The intended goal was that two team members of the design team carry out heuristic evaluation separately to compare the results and create a common summary. However, one member encountered family issues that prevented her from performing the evaluation. Sophia is a startup at an early stage of development, thus, Sophia could not afford to hire external professional evaluators. Consequently, I decided that at the current stage of Sophia's development the second heuristic evaluation is not crucial.

HE results for Sophia MVP

Evaluator #1 (Lidia Borisova).

Date of heuristic evaluation – the 28 January 2019.

Project status – MVP is ready but still needs minor corrections. Heuristic evaluation was done before MVP was presented to end users.

Rating scale (Nielsen, Mack, 1994, p. 49)

0 – I don't agree that this is a usability problem at all

1 Cosmetic problem only – need not be fixed unless extra time is available on project

2 Minor usability problem – fixing this should be given low priority

3 Major usability problem – important to fix, so should be given high priority

4 Usability catastrophe – imperative to fix this before product can be released

Table 5 - Heuristic evaluation results for Sophia MVP

The name of heuristic	Rate	Explanation
-----------------------	------	-------------

1. Visibility of system status	3	The visibility of the system is insufficient.
2. Match between system and the real world	2	The language of the system reminds about other services on the rental housing market. However, a few mismatches in vocabulary were found.
3. User control and freedom	0	The heuristic is working due to “back” buttons, menu dashboard and by showing me the location that the user types to let the user verify if he/she puts the information correctly (tenant’s side).
4. Consistency and standards	3	The website has a design consistency: colours, type fonts, shapes of the bars have the same style. However, the feature with double matching that distinguishes our service from competitors might be difficult to understand for novice users as they didn’t see that double matching before.
5. Error prevention	0	This feature is supported by showing the user the location that he/she types to let the user verify if he/she puts the information correctly (tenant’s side). The same verification statement the system provides at landlord’s side. In addition, the possibility for the landlord to make the announcement public or to hide it if he/she wants to change the information. In addition, in both profiles (tenant’s and landlord’s) MVP has a function to modify information.
6. Recognition rather than recall	0	The walk through of the system is fast and easy for both sides. It is not overwhelmed by information. The navigation bar on the left helps the user to understand at which place in the process he or she is.
7. Flexibility and efficiency of use	0	When the user has an account, he directly falls into the matches page where he/she can easily switch from matches to his/her profile information. That function provides the feeling of flexibility.
8. Aesthetic and minimalist design	0	Current version of the interface design is minimalistic and has design consistency. However, that it due to the fact that now MVP doesn’t contain irrelevant information. Opposite, MVP doesn’t contain enough of relevant information. When the relevant information such as tutorials, texts, will be added the heuristic of aesthetic and minimalist design should be checked again.
9. Help users recognize, diagnose, and	0	The same answer as for heuristic #5.

recover from errors		
10. Help and documentation	n/a	In my opinion, irrelevant heuristic at this early design stage of the project.

More elaborated heuristic evaluation results are in the appendice.

During the evaluation, I faced the challenge of checking several heuristics. For example, being a designer, I might be biased evaluating the match of the product language and real-world users' language (heuristic #2). I cannot deny my professional vocabulary and cannot look to the labels of the system from the point of view of an outsider. Another point is that I am so familiar with the service that I cannot see the confusions and ambiguities that a novice user might face. The same challenges were pointed out by one of the respondents I interviewed (interviewee #6).

Despite the serious limitation of the method when only one person carried out the evaluation and taking into account that I was familiar with the product, nevertheless, heuristic evaluation is effective. It helped me to reveal a few usability mistakes that our team could easily fix at MVP before launching the coding of the next version of MVP with real users. Thus, the Sophia case showed that the method seems to be cost-effective, because even one designer can perform the evaluation and reveal some usability mistakes. However, mistakes that I discovered were not critical for task accomplishment, but more for overall interface design of the system.

By carrying out heuristic evaluation of Sophia's MVP, I saw that the smaller is the scope of MVP, the fewer mistakes we can find. For example, our MVP focused on testing the process of matching the tenant's requirements for an apartment and the landlord's offer. However, the process of payment or making an e-contract was not yet developed. For example, it is crucial to test the payment feature on the possibility to prevent errors (the heuristic #5) because it is a financial transaction process that should be fast, easy and at the same time secure. However, we were not able to do that because our MVP did not provide that feature yet. I faced the same challenge carrying out the evaluation based on the 10th heuristic: help and documentation function. Sophia's MVP does not have any help and documentation function. This feature, that does not directly affect the task accomplishment, is usually developed at the end, when all other functions are already developed and tested on end users. Thus, for Sophia, it is irrelevant to test the "help and

documentation” heuristic for the current MVP development stage. The conclusion from that reasoning is that when a new version of MVP with more features comes out, we should carry out heuristic evaluation heuristic evaluation again.

However, I think the 10 heuristics are overlapping with each other. For example, heuristic #5 “error prevention” and heuristic #9 “help users recognize, diagnose, and recover from errors” are evaluating the same features of the system. The possibility to prevent errors, to modify information, and to check the information is an integral part of the user control. The consistency and standards (heuristic #4) and match between system and the real world (heuristic #2) produce the same overlap. This duplication resulted in repetition of the information in the evaluation results. Consequently, for future heuristic evaluation of the next version of MVP I consider revision of the 10 original heuristics and come up with a shorter list.

6.1.4. Conclusion

My theory-based research framework presented earlier in the corresponding chapter enabled me to summarize my findings of heuristic evaluation method. The structure of the framework allows me to describe the holistic process of heuristic evaluation from the preparation phase to the interpretation phase taking into account startup context. The framework includes general parameters of evaluation method such as special preparation before the evaluation in terms of equipment, environment, materials; the number of tests to be done to get reliable results; the involvement of end users; the qualification of evaluator; the efficiency to reveal usability mistakes and types of problems identified; the number of tests to get reliable data. In addition, the framework visualizes the phase of UX design process when the analyzed method would bring the most value.

Testing

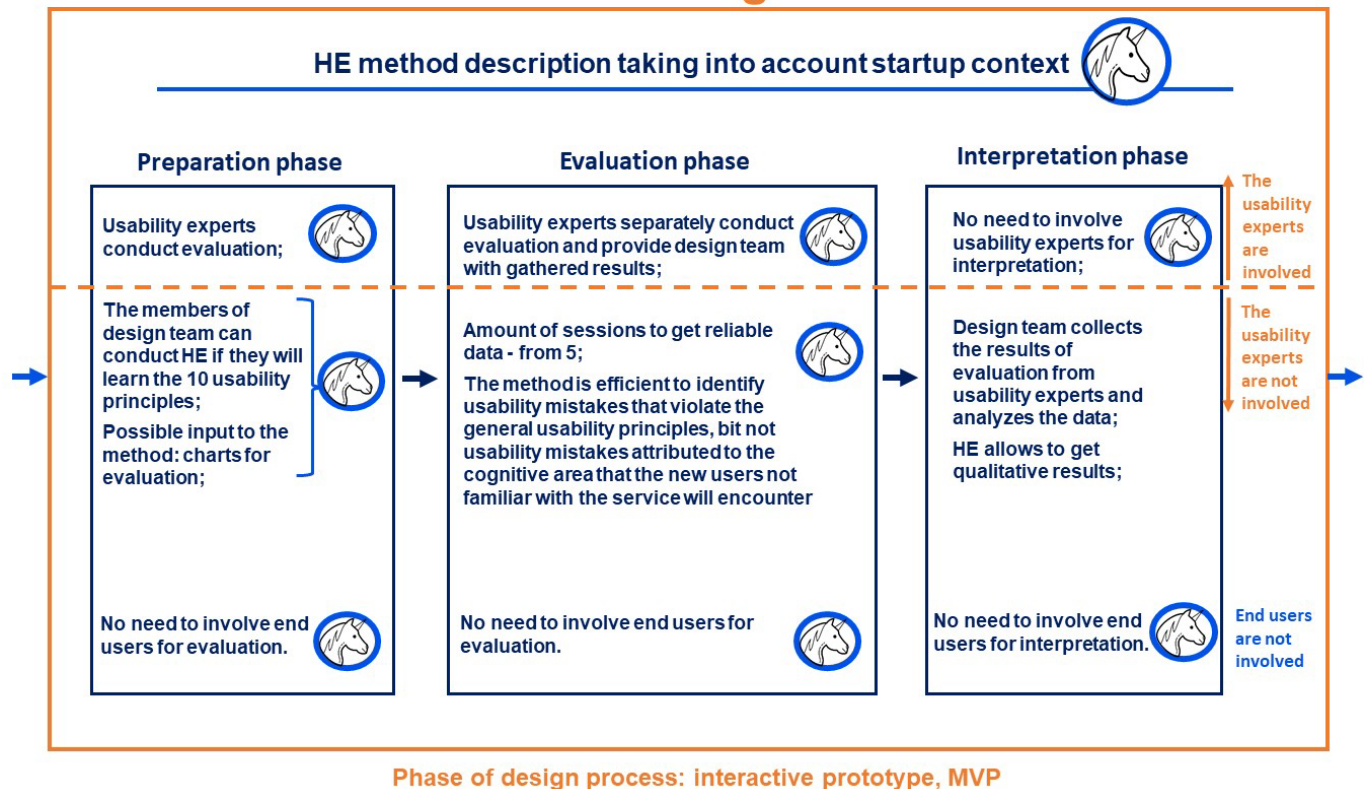


Figure 11 – The description of heuristic evaluation method using theory-based research framework to describe the UX evaluation method involving startup constraints

My multi-faceted research of heuristic evaluation revealed the different understanding of the process in certain parameters among researchers and UX practitioners. Based on my empirical analysis, heuristic evaluation is not a very popular method among startups: only one startup employee uses it in his work. If we take into consideration the broader sample of respondents with the representative of academia and the interviewee from a big company, the pattern will be slightly different. The researcher in Aalto used the method in her work, while the director of Wäertsilä claimed that his team designs everything having the 10 heuristics in mind, so there is no need for external evaluation. The observation that two respondents out of seven use heuristic evaluation and find it useful and cost-effective leads to the assumption that the method is not considered to be beneficial for product development in the case of the startup environment. My own experience with the Sophia demonstrated that heuristic evaluation helped me to reveal some usability problems of our first version of MVP, but not the issues critical for the overall functioning of the system.

The observations from empirical studies contradict the academic vision of the purpose of heuristic evaluation in the website creation process. Nielsen wrote that the design team while choosing between a minimum set of evaluation methods should stick to the heuristic evaluation (Nielsen & Mack, 1994). Nevertheless, the data from interviews demonstrate that heuristic evaluation is perceived more as a supplementary method to find usability problems, but not self-sufficient and able to substitute other UX evaluation methods. However, both academic sources (Nielsen & Mack, 1994), empirical study and Sophia case study support that the method is relatively cheap, fast and easy to perform, thus, can be attributed to discount usability evaluation methods.

One important disagreement with the Nielsen's guidelines on how to perform evaluation is the statement by Nielsen that external evaluators (at least three) should independently evaluate the system. However, in the Sophia case and based on the answers from the interviews, I can observe that the designer or design team can carry out evaluation without the help of external professionals. This will save the startup a lot of resources such as time and money. Respondents, myself and academics – we all agree that the method is relatively easy to use and doesn't require complex training compared to other evaluation methods.

My general observation is that in academic literature (Nielsen & Mack, 1994) the practice of applying heuristic evaluation is usually related to already developed systems or at least MVP, not prototypes. As I mentioned above, some heuristics are usually attributed to the well-developed product, not the early prototype or MVP that has to demonstrate only core functions. However, the common conclusion from empirical part #1 is that the method is useful in the early stages of product development. That opposes my evaluation in the Sophia case because I intentionally made heuristic evaluation only when MVP was ready.

Another point that emerged after both empirical analyses is that heuristic evaluation should be modified, as the original version is not optimal. One of the interviewees wrote that he uses the modified version (Interviewee #2) and another would do that if she felt more confident (interviewee #6). I did not modify the version, not feeling qualified enough. However, I experienced that some positions in the traditional edition of ten heuristics repeat themselves.

The next conclusion from the research is that in order to avoid heuristic evaluation and save time and money, the design team working in startup can study the 10 heuristics and keep them in mind while developing the product from sketches to the MVP. In Sophia, our design team partly followed

this principle while creating sketches, high fidelity wireframes and clickable prototypes. In Wärtsilä it is the common practice of every member of the design team.

The following consideration is that heuristic evaluation could be an instrument to communicate the UX design of the product to external people as it is in the case of Wärtsilä.

One parameter that was equally considered by researchers and practitioners is that it is difficult for designers to apply heuristic evaluation to the product that they designed – the vision is subjective and not critical enough. I personally experienced that in case of Sopia. One of the respondents offered the same comment (Interviewee #6). Nielsen, the designer of the method, described that in order to get reliable data the evaluation should be performed separately by at least three persons and after the results can be analyzed and combined. Thus, the involvement of additional people (at least two) will help to overcome the challenge described above.

6.2. Cognitive Walkthrough

6.2.1. Academic literature analysis of CW

The cognitive walkthrough (CW) is a task-oriented usability inspection method that “focuses on a user’s cognitive activities; specifically, the goals and knowledge of a user while performing a specific task” (Wharton et al, 1992, p.1). Wharton, Rieman, Lewis and Polson developed cognitive walkthrough in 1990 (Nielsen & Mack, 1994). Nielsen included the method in his book “Usability inspection methods” in 1994 (Nielsen & Mack, 1994). The main idea of the method is that the evaluator gives a task to the user and observes how the user uses the system to accomplish the task. The participation of end-users is not required, but possible. If end users are not involved in the evaluation, the person who walks through the system should behave from the perspective of the end users. In addition, the evaluation could be carried out in a group. For a group usability evaluation, the design team presents the prototype of the system or the system itself to a group of colleagues, usually after a certain milestone such as interactive prototype or MVP. The design team uses the gathered feedback to improve the design of the next revision (Nielsen & Mack, 1994). The evaluator has to mark the mistakes in the product that prevent the user from walking to the desired goal without interference. Evaluators study how easy and fast it is for new users to achieve goals with the system. Design team can apply cognitive walkthrough to review complex user interactions and tasks (Khajouei et al., 2017).

Cognitive walkthrough is attributed to the cognitive model of the system user (Mahatody et al., 2010). The definition of cognition is the mental action of gathering knowledge and understanding that includes perception and judgment. “Cognition includes all conscious and unconscious processes by which knowledge is accumulated, such as perceiving, recognizing, conceiving, and reasoning” (Encyclopedia Britannica). Thus, cognitive walkthrough is a usability evaluation tool that concentrates on checking a system design for ease of learning, especially by exploration (Nielsen & Mack, 1994). Designers need to check how easy it is to learn the product (cognitive model) because users, interacting with a certain digital product for the first time, often aim to learn it by themselves avoiding manuals. Consequently, cognitive walkthrough shows how easy it is to explore the system for the novice user.

To accomplish cognitive walkthrough, the evaluator has a scenario or a few scenarios. Scenario means a certain task or a set of tasks that a user should accomplish. Examples of scenarios: register in the system; find a specific item and buy it; find a company’s contact information. In the case of a rental housing platform, the scenarios could be: for tenants - find an apartment for rent; for landlords – find a tenant to rent out an apartment. Evaluators should carefully design the scenarios to represent the main goals that users should achieve using the system. Scenarios should be created in a way to cover all potential actions that user might take to accomplish a task. Hence, it is recommended to design a reasonable number of scenarios to secure that all users’ tasks are covered (Mahatody et al., 2010).

While creating scenarios the evaluator faces the challenge of prioritization which tasks to choose for scenarios. It is crucial to choose tasks correctly. Yet, cognitive walkthrough does not include guidance on how to prioritize tasks. “Any interface of even moderate complexity supports dozens or hundreds of tasks and task variants, and only a small fraction of them can be evaluated”. (Wharton et al., 1992, p.3)

Evaluators can be designers, users or professional usability experts (Mahatody et al., 2010). The process is as follows: the facilitator gives a scenario for the user to walk through the website. Users do not have to possess special training before testing the system. Ideally, the user should not be familiar with the system. Cognitive walkthrough method supports the logic of iterative UX design process. The method is designed to be used iteratively from early phases of the design cycle to

more developed phases. The evaluation could be performed by either software developers or usability specialists (Wharton et al., 1992)

There are two options to carry out cognitive walkthrough – to ask the user to think aloud or just to observe the process. If we ask the user to comment aloud on his/her actions that means the cognitive walkthrough is performed with the think-aloud protocol. Think-aloud protocol is the method when during task accomplishment, users describe aloud the process. Protocol welcomes users to explain out loud what they are focusing on, thinking, and feeling, and what grabs their attention on their journey to accomplish tasks. Yen and Bakken (2009) state that such parameters as functionality, features, processes, user interface, user-system interactions and manager-staff communication can be checked with the protocol.

Simultaneously, the facilitator might ask a person who is testing the website to answer the following questions at each step (Yen and Bakken, 2009, p. 2):

1. What do you like the most about the system and why?
2. What do you like the least about the system and why?
3. Do you have any suggestions for improving the system?

The person, based on his / her experience, comes with a success or failure story for each step (Wharton et al., 1992).

A sufficient number of users to test is 20 persons to secure valuable evaluation results (Meier et al., 2017). Test persons could also be recorded by a web camera. More recent academic research suggests that even five tests with potential users will reveal some usability problems (Cockton & Woolrych, 2002).

The cognitive walkthrough method has been actively developed over time. Practitioners and researchers proposed their modified versions of cognitive walkthrough, but the core of the process remains the same – “the method simulates the cognitive behaviour of the user by responding to questions related to the user's cognitive model" (Mahatody et al., 2010, p. 2).

6.2.2. Empirical part 1– conclusions from the interviews with practitioners.

Cognitive walkthrough is a common way to test UX for startups. From seven respondents, six claimed that they use it in practice: “it is the most useful tool for the usability audit” (Interviewee #1) or “It is a very valuable method – it gives an in-depth understanding of customers and end users” (Interviewee #7). One respondent who does not use the method pointed out that the method could be effective for startups with a different situation than his own. He said that if the startup team is not familiar with the market where the startup is operating, and if there is no clear understanding of the users, cognitive walkthrough can bring valuable results: “cognitive walkthrough is useful for a startup who doesn’t understand the market and the customer. By carrying out tests, startupers can get relevant knowledge to develop the product” (Interviewee #5). The interviewee mentioned that in the case where the website is the main product – such as a platform business, cognitive walkthrough would be effective in revealing UX mistakes. In this situation, the website was a marketing channel, but not the main product.

From six respondents who used the method, five used it with the involvement of end users and one respondent asked his design team members to perform the role of potential end users (interviewee #4). For the purpose of cognitive walkthrough, both approaches are correct: with and without end users. His colleagues were not yet familiar with the website: “I asked three people in the team to check the website. They were not familiar with the website, so their vision was “fresh” in a sense” (Interviewee #4). During testing, members of the design team could partly perform as potential end users if they encounter the website for the first time. The observation that the CEO of that startup among all evaluation methods chose brainstorming and cognitive walkthrough with team members demonstrates that both methods bring him the most value into the design process.

The representative of academia (Interviewee #6) and the respondent of the big company (Interviewee #7) said that they regularly used cognitive walkthrough in their work. Thus, the pattern is the same as in the case of startups.

All the respondents who used cognitive walkthrough asked the users to think aloud while trying to achieve the goals settled in the scenarios. There is no best timing to start cognitive walkthrough, based on practitioners’ opinion. The method could be run multiple times through the product development process from early stages to the MVP; one can use cognitive walkthrough anytime. Interviewee #6 mentioned that entrepreneurs could use the method to collect requirements about

business idea by demonstrating the users the early prototype (Interviewee #6). However, interviewee #7 said that it is difficult to perform cognitive walkthrough when you have just the sketches or low fidelity wireframes of the future product. The level of visual representation of the system in that case is too abstract and, consequently, end users cannot understand it well enough to provide constructive feedback: “We cannot expect them to comment sketch or low fidelity wireframes. We can obtain valuable feedback only for clickable visual prototypes”. Respondent #4 said that designers should perform cognitive walkthrough at the stage of MVP, not earlier: “If the website is the main product, I would start testing when the MVP (website as MVP) is ready” (Interviewee #4). That statement resonates with the answer of the respondent #5 who claimed that tests with end users are effective when the startup is developing version 2.0 of the product.

Talking about the cost-effectiveness of the method, all respondents except one answered that the method is expensive because you need to find end users and guide the tests with them. Testing takes a lot of time, so “recruiting participants and time are the biggest expense” (Interviewee #2). “In general, I would consider the method to be expensive for early stage startups as it takes time” (Interviewee #4). Cognitive walkthrough is especially expensive for a big company because the company produces a niche product (systems for chips), so to arrange the tests with potential users sometimes designers need to travel.

The expenses that startups incur to find end users to test with are the main obstacle for some startups not to perform cognitive walkthrough. “For startups time is money” (Interviewee #4) and if the startup owner or employee thinks that they know the market, they will not spend resources on UX evaluation as such, either on cognitive walkthrough.

The challenge to arrange cognitive walkthrough tests that will yield valuable results is the accurate choice of the end users: “We put a lot of effort into selecting the right segment that will form a good representation of the users” (Interviewee #7). In addition, it is important how the designer facilitates the cognitive walkthrough. The potential user should be willing to share his / her honest thoughts about the system: “It is important to have a warm up before the study so that a user feels comfortable and actually does think aloud, especially critical and negative thoughts” (Interviewee #6).

Some respondents mentioned different software that can substitute cognitive walkthrough. Such software products do not facilitate face to face testing with end users like in case of cognitive

walkthrough, but by tracking the task accomplishment allow to collect quantitative data how fast and easy the service might be for end users:

- Fullstory - <https://www.fullstory.com/>
- Usetrace - <https://usetrace.com/>
- Usabilityhub - <https://usabilityhub.com/>

Interviewee #3 often uses Usabilityhub to check usability and UI problems. However, interviewee #2 emphasized that he doesn't "believe any software can fully replace the evaluator" (interviewee #2).

Interviewee #7 claimed that his team sometimes uses Skype video call and video recording of the screen to conduct the test with the users remotely. Skype does not substitute the evaluator as the designer is still guiding the users through the system, but skype saves the travel expenses and facilitates the recording of evaluation.

According to the practitioners, the limitation of the method is that it provides unreliable data: subjective user's perception and sometimes unnatural behaviour: "users do not always behave the way that they naturally would when doing a cognitive walkthrough" (interviewee #2) and "Such tests with end users are subjective and you cannot fully rely on the data gathered" (Interviewee #4).

Those practitioners who did not use cognitive walkthrough mentioned that they already had a strong opinion about what they wanted, so they perceived the talks with users as unnecessary tasks that will cost time. They pointed out that in order to create a good website they used existing benchmarks and copy their approach: "Existing templates representing many modern popular UX designs helped me a lot in building my website without testing its prototype with end users. I skipped the phase of the prototype. Templates are an effective way to create a product with average quality; this is enough for my website" (Interviewee #5).

The majority of respondents did not specifically indicate the required amount of users to conduct cognitive walkthrough. However, interviewee #7 said that in his experience five to ten people is enough to yield valuable insights: "After 10 people we start to get repetitive answers" (Interviewee #7).

However, one of the respondents who used cognitive walkthrough emphasized the importance of cognitive walkthrough as a tool to overcome the design team limitation of vision: “in my experience, developers in a startup often focus on their idea and don't necessarily spend enough resources on checking if users need the idea or get the idea. This could be conscious (we don't have time, let's just try to do it and then see) or unconscious (we think we know our users well, they'll definitely like it/will definitely know how to use it). Even a low-budget user study could solve this problem” (Interviewee #6).

6.2.3. Empirical part 2 - The cognitive walkthrough tests of Sophia prototypes

Taking into account the importance of creating a high quality UX, our design team started to make cognitive walkthrough as soon as we had two clickable prototypes. We did not test sketches because they were too undeveloped to show them to potential end users. We did not make cognitive walkthrough with high fidelity wireframes, either, because our wireframes demonstrated the different versions of the interface design but not the core functions of the platform that were crucial for checking the usability of website.

As soon as we had first drafts of prototypes, we started to test them. However, the process passed multiple iterations. From the first test with the user, we revealed many crucial mistakes that we had to redesign with both prototypes. Therefore, apparently, after each user testing, we were fixing minor and major problems. Our users were people from our network who were not at all familiar with our service. We gave our testers scenarios: scenario #1 - to find an apartment for rent, scenario #2 – to make an announcement of the new apartment to rent out using the second prototype. Two service design team members from Sophia did the tests: my colleague and I. Both of us, except for my education in UX design, did not have special training in cognitive walkthrough evaluation. The length of the test was approximately from 30 minutes to 1,5 hours . The more tests we did, the fewer problems remained, and the simpler were the problems. The first test, for example, showed that the prototype does not work at all (in the designer's language that type of problem is labelled “catastrophe” – when the task accomplishment is impossible). However, the latest tests focused on the details of the prototype: labelling of features, the layout of the website, the sequence of actions etc. With cognitive walkthrough, we checked both the aesthetics of the interface and the usability. We tested Sophia's prototypes with six users.

In general, all users' comments were subjective. However, when the majority of testers pointed to the same UX issue it became clear that, despite the subjectivity of evaluation, if we did not correct the mistake, it could be common to the majority of all potential users.

The majority of the interviewed practitioners pointed out that cognitive walkthrough is expensive as it requires time and money to acquire users to test the product. However, in the case of Sophia the situation was different. We used our network to make the tests. Our users were people we know. We did not pay anything to the users and it did not take too much time to find those people. We were able to use the resources of our network because Sophia offers a service for the mass market: renting or renting out an apartment. It is not a niche product. However, considering future tests, when the resources of our networks are finished it might cost some money for Sophia to acquire new users for tests.

Cognitive walkthrough was the main tool for us to make significant progress in our service. Every user test revealed the mistakes that were hidden from our eyes, as we were biased by our own work and our vision was limited. The method was easy to use without special training because our main role as evaluators was to give the user the scenarios, observe how the user performs and make notes about usability and aesthetic mistakes that the user encounters on his/her journey with the service. Thus, we found the method to be cost-efficient as we did not hire a qualified evaluator but were able to carry out evaluation using our own resources.

Despite the availability of specific software to make cognitive walkthrough, we carried out tests by meeting with our users face to face. The reason is that we did not have special qualifications in such software and Sophia did not have the financial resources to buy such programs. In addition, we found personal communication effective. After the testing we had a general discussion with every user about the current flaws in the prototypes. This discussion was very valuable for us, as the users' experiences were still fresh enough and they were able to provide us with insights about our service design and usability.

We continued to carry out cognitive walkthrough tests with the MVP of Sophia service.

6.2.4. Conclusion

I used again my theory-based research framework to summarize my findings of the cognitive walkthrough evaluation method – Figure 12. The framework enables outlining the holistic cognitive walkthrough evaluation process taking into account all actions from the preparation phase to the interpretation phase to perform evaluation. The framework allows us to consider the applicability of cognitive walkthrough evaluation method in the startup context.

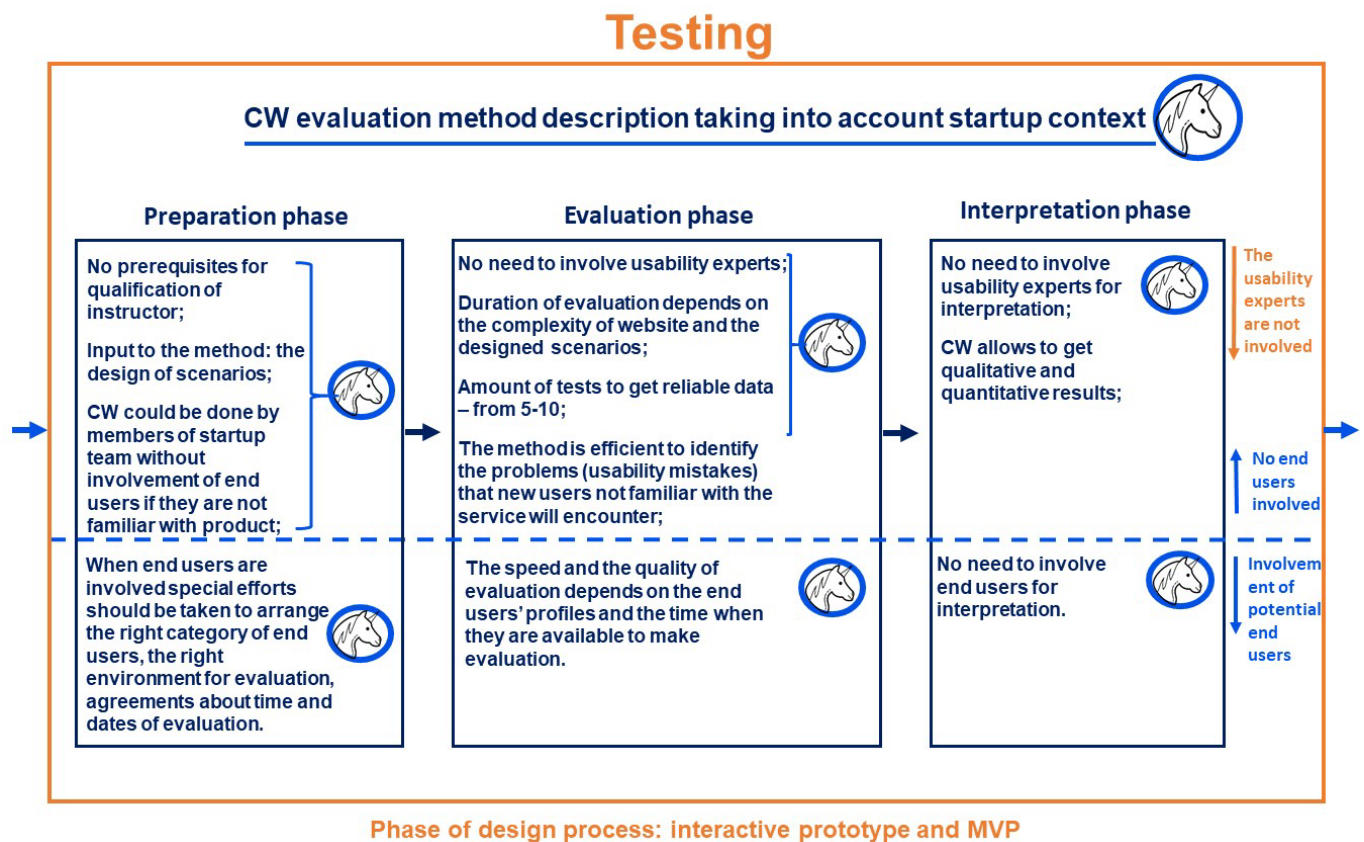


Figure 12 – The description of cognitive walkthrough method using theory-based research framework to describe the UX evaluation method involving startup constraints

My theory and practice based research of cognitive walkthrough revealed a common agreement between researchers and field practitioners about the value of cognitive walkthrough process in UX evaluation. Based on combined analysis, cognitive walkthrough is one of the most popular tools to test UX. Five out of seven practitioners from the startups, academia and a big company I

interviewed use it. One respondent conducted cognitive walkthrough with his team members, not potential end users. One respondent who didn't use it, believes that the tool is effective under certain conditions: when the startup team just enters the market and is not very confident about their customers' needs, expectations, behaviours etc. In that case, cognitive walkthrough allows gathering sufficient data to develop appropriate UX that will satisfy customers. Based on the Sophia experience, the tool brought us the most valuable results to improve our prototype and MVP.

Both academic literature (Cockton & Woolrych 2002) and data from practitioners demonstrate that the number of users to test with could start from five people. One interviewee claimed that even three tests were enough to reveal UX mistakes. The respondent who carried out cognitive walkthrough with his team members engaged three participants to evaluate the website.

A common observations from all three types of analysis is that the method is easy to use. No special training is required for designers to facilitate the testing. In the case of Sophia, our design team gave a task to the end users. All respondents also conducted the evaluation using their own resources, as Interviewee #2 stated: "cognitive walkthroughs are easy to do". The crucial aspect is that users with whom the testing is done should have no prior knowledge of the system to be able to look at it with fresh eyes.

Another common agreement is that cognitive walkthrough could be used multiple times during the design process: designers collect the summary of UX mistakes after the first cognitive walkthrough, they then can fix those mistakes and conduct the cognitive walkthrough again. Cognitive walkthrough is aligned with the logic of iterative UX design process that is considered to be effective for startups (see Chapter 2).

From the literature analysis and from the interviews the method could be defined as expensive, because the startup has to invest time to find end users to tests the product. Sometimes entrepreneurs may conduct cognitive walkthrough without expenditure using their own network, but in any case the evaluator has to find time to sit together with the users and to observe the walkthrough process. Since time is one of the most crucial resources for startups, the method is perceived to be costly. In order to conduct efficient evaluation, entrepreneurs could pay attention to selecting the sample of end-users. To obtain valuable cognitive walkthrough results, end-users should represent the potential end-user profile.

Currently, special software cannot substitute for the tests completely. When the method was developed in 1994, the IT industry didn't yet provide any programs. Now different systems to conduct evaluation exist. However, their application is still limited, as the systems cannot fully gather in-depth qualitative data about the UX issues that the evaluator who observes cognitive walkthrough can.

6.3. The system usability scale (SUS)

6.3.1. Academic literature analysis of SUS.

The system usability scale (or SUS) is a “quick and dirty survey scale” that enables the usability practitioners to quickly and easily evaluate the usability of a given system (Bangor et al., 2008, p.1). The tool was developed by Brooke in 1996. Brooke created the method as a measure for the subjective perception of users of the usability of the product. “We wanted a tool that would allow us to take a quick snapshot of people's satisfaction with using the systems that we were asking them to use, and that we could use as a yardstick to compare one system with another or to compare a new version of a system with an earlier incarnation”(Brooke, 2013, p.5). Brooke mentioned two main system usability scale objectives, such as to be able to measure people's subjective perceptions of the product and to make it in a very short period of time (Brooke, 2013).

The system usability scale survey has 10 positions that the user has to evaluate while interacting with the system.

The system usability scale survey has 10 questions with 5 response options (McLellan et al., 2012, p.3):

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.

8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

The response format is a five point Likert scale. According to Brooke, the optimal amount of users for evaluation is from eight to ten people (Brooke, 2013). However, at the website of Norman Nielsen Group it is said that to obtain reliable data the sample of users should be from 20 to 30 (Norman Nielsen Group, 2018). This sample is much bigger than the minimum of five users to conduct heuristic evaluation (Nielsen & Mack, 1994) or cognitive walkthrough (Cockton & Woolrych, 2002).

Afterwards, the evaluator calculates the score in the following way:

- Each item's score contribution values from 0 to 4;
- For the odd numbers (the items that have positive wording) the evaluator should subtract one from the user evaluation;
- For the even numbers (the items that have negative wording) the score contribution is 5 minus the scale position;
- The scores are summarized;
- Finally, the evaluator multiplies the sum of the scores by 2.5 to obtain the overall value of SUS (Brooke, 2013, p.7)

When the final score is calculated the higher score indicates better usability (Bangor et al., 2008). The results of 80 or higher are attributed to great usability (Norman Nielsen Group, 2018). System usability scale provides a single score, and there is no sense in referring to the scoring of each category separately. That means that the evaluator should not investigate each category separately, but look at the score in sum: Brooke cautioned that scores for individual positions are not valuable on their own (Bangor et al., 2008).

Researchers do not have an agreed opinion about system usability scale scoring and score interpretation. Some (Bangor et al., 2008) state that system usability scale is rather quick and easy to use by interviewees and evaluators. In addition, the system usability scale outcome in a format of a single score on a scale is easily grasped by almost anyone from project managers to software developers (Bangor et al., 2008). Others discuss the difficulty in calculating and interpreting the

system usability scale score. Even the creator of system usability scale, John Brooke states in his article that certain researchers perceive the scoring of system usability scale difficult and confusing since the usage of both positive and negative categories adds complexity and can lead to errors (Brooke, 2013).

System usability scale can be effective when designers evaluate their product with industry competitors. In addition, with system usability scale, designers can track the overall improvement in usability perception during the product development life span. However, the method will not provide information about the particular components of the system that should be improved. Brooke didn't suppose each individual question to have diagnostic value in itself or to be attributed to the certain features or function of evaluated system (Brooke, 2013). System usability scale does not indicate which component of the system the design team should improve, but shows a general assessment of how the users perceive usability.

Finally, the system may not be suited to the evaluation of modern digital products. System usability scale was originally created to evaluate massive, integrated office systems that differ from the digital products we use today (Brooke, 2013). However, the tool has particular benefits: “the survey is technology agnostic, making it flexible enough to assess a wide range of interface technologies, from interactive voice response systems (IVRs) and novel hardware platforms to the more traditional computer interfaces and Web sites” (Bangor et al., 2008, p.1).

6.3.2. Empirical part 1– conclusions from the interviews with practitioners.

None of the respondents used system usability scale in their work. Only one had knowledge of the tool and did not use it intentionally. The other respondents were not familiar with the term and, consequently, with the method. The interviewee who knew the method mentioned that system usability scale, in his opinion, was slightly outdated and cumbersome in the modern age. “It is not enjoyable for the user to fill out” (interviewee #2). In addition, the respondent pointed out that system usability scale does not establish which aspect of the UX is problematic. It gives the overall score, but not a concrete understanding of UX mistakes.

Three respondents checked the information about system usability scale after my question and all others answered that they think the method could bring certain value when MVP is ready, not at the development stage of the product. In addition, the method could be useful in comparing the

solution with competitive systems: “It should be a good and cost-effective way to benchmark an existing solution (see how bad it is) or an MVP (see how good a new solution is) “(Interviewee #6). This statement is aligned with the logic of the method’s creator Brooke that the method allows the comparison of different websites and different versions of the same product.

6.3.3. Empirical part 2 - Sopia case

My initial choice to analyze system usability scale as one of the potential UX evaluation methods came from my UX studies in Aalto where we learnt about the method. We were presented the method in line with other UX evaluation tools such as tree testing, heuristic evaluation and cognitive walkthrough. From my studies, I formed an initial assumption that all studied methods have a similar contribution in evaluation of UX issues. At that point of time, I perceived system usability scale to be as effective to reveal UX issues as other methods analyzed in this study. In addition, I thought that time to conduct system usability scale is relatively similar to the requirements for evaluation using other methods.

However, later, trying to understand how I can apply the method to evaluate Sopia website, I faced certain challenges. On the website of Norman Nielsen Group it is said that system usability scale is difficult to understand for novice UX practitioners (Norman Nielsen Group, 2018). The beginners in UX do not understand the aim of the tool, the process, the scoring system and its interpretation. For them, the method is perceived as complex and ambiguous. In the case of Sopia I shared the same sentiment. Personally, I partly grasped the process of system usability scale only after spending a significant amount of hours analyzing the research literature on the topic. When we were rushing to develop our prototype – the situation familiar for the majority of startups – our team did not have time to learn how to use system usability scale. I intentionally put the method aside. While comparing cognitive walkthrough and system usability scale, I decided to proceed with cognitive walkthrough and heuristic evaluation as these methods seemed to us to be more clear, easy to perform and flexible.

The system usability scale questions partly repeat the heuristic evaluation questions and cover the same topics. Thus, I considered that in the Sopia case the system usability scale evaluation method would not bring special value into the product development process.

Another reason why our team did not apply system usability scale evaluation to measure the usability of Sophia's prototype, and later the MVP, is that system usability scale evaluation would provide us with only a general score for the system, without specification of particular usability problems as it was in the case of cognitive walkthrough and heuristic evaluation. However, in order to progress with our service, we needed a detailed understanding of where the usability problems are. For a startup, it is crucial to understand as early as possible what kind of usability mistakes the prototype has and to fix them before putting resources into coding. This I considered as a limitation of the method, because we could not address specific problems and improve them later, as in the cognitive walkthrough.

In addition, I did not find system usability scale to be useful when the product is in the early stages of development. When we started to make the evaluation, we had just two prototypes with a limited amount of features and many usability mistakes. System usability scale would not help us to reveal these mistakes and correct them.

6.3.4. Conclusion

Using my theory-based research framework, I summarized my findings of system usability scale evaluation method – Figure 13. The structure of the framework allows me to describe the holistic process of system usability scale from the preparation phase to the interpretation phase taking into account startup context.

Testing

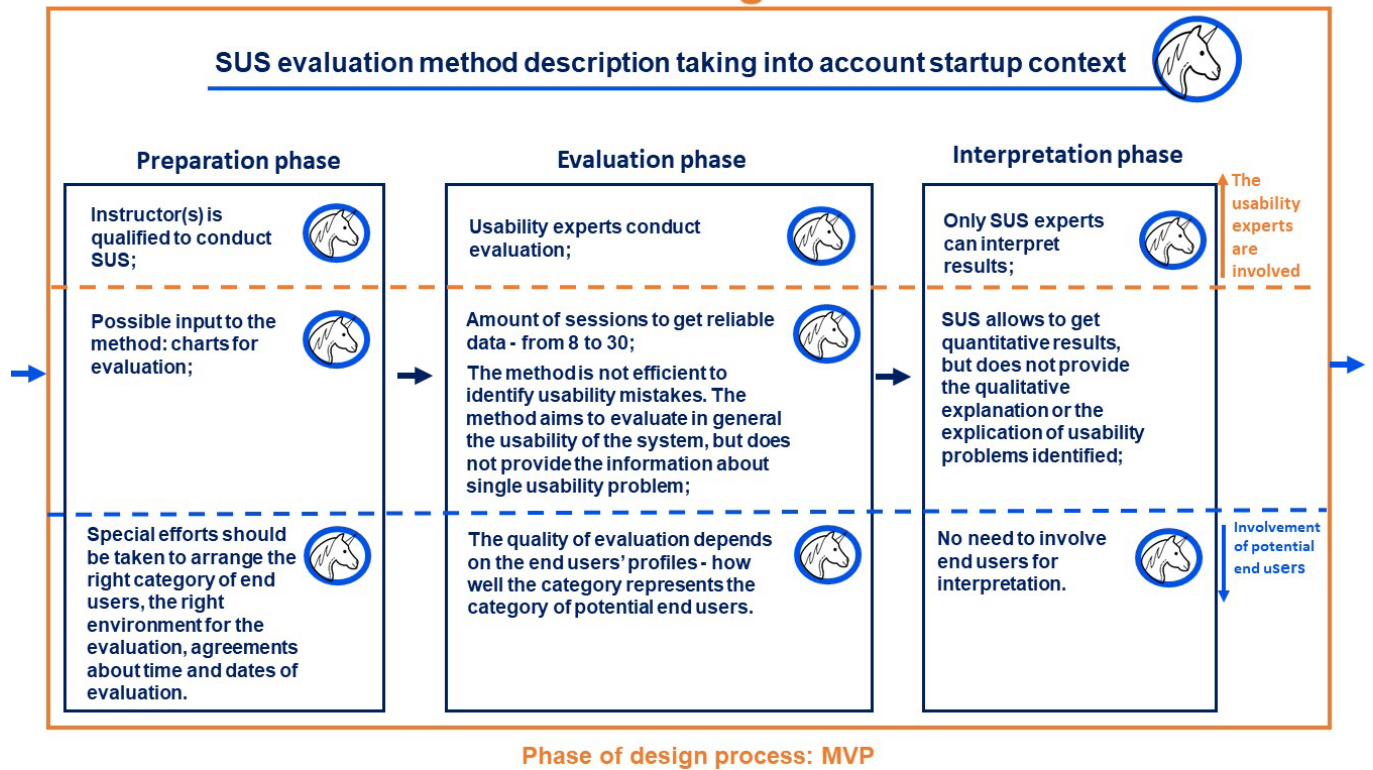


Figure 13 – The description of system usability scale method using theory-based research framework to describe the UX evaluation method involving startup constraints

My multi-faceted research of system usability scale disclosed the difference between researchers and UX practitioners in understanding of the process and value of system usability scale in UX evaluation of early stage startups. Researchers claim that system usability scale enables the usability practitioners to quickly and easily evaluate the usability of a given system (Bangor et al., 2008). However, based on my empirical analysis and Sophia case study none of the practitioners used the method or had plans to do so. The observation that the majority of my respondents are not even familiar with the term might signal that in general in the industry the method is not very popular: “no one here uses system usability scale as a word. Or at least I never heard anyone mentioning it once” (Interviewee #3). In the case of Sophia, we also neglected the tool despite the fact that I knew the method. From both empirical analyses, I can conclude that the startup environment does not support the usage of system usability scale. The time constraints and the

necessity to put all the resources into MVP development do not allow for the dedication of the effort and time to apply system usability scale.

Empirical findings contradict the academics' point of view. Researchers defined the method as a quick and easy way to measure usability. However, from empirical analysis and comparing with other UX evaluation methods that I investigated in this study, system usability scale is perceived to be difficult and time-consuming. First, the evaluator should have special training to be able to conduct the evaluation. Second, researchers state that the minimum sample for evaluation is eight users, which is a big investment in terms of time and money for startups. And, lastly, the method provides only a general score that could not be elaborated in a more detailed description of the UX problems. However, for startups the main reason for evaluation is to be able to track usability mistakes and correct them as soon as possible before the programming of the MVP.

6.4. Tree testing tool

6.4.1. Academic literature analysis of the tree testing tool.

The Norman Nielsen group provides the following definition for the tree testing method: “A tree test evaluates a hierarchical category structure, or tree, by having users find the locations in the tree where specific tasks can be completed” (Norman Nielsen group, 2017). Donna Spencer developed the method in 2003. Spencer developed tree testing as a user research method out of the need to regularly check hierarchy structures, detached from the interface (Le et al., 2014). The notion of the tree determines the tree of consecutive steps (functionalities) in the website, i.e. which functionality hides behind which label. The term describing the tree of the website is information architecture (IA), which can be defined as “the structure or map of information which allows others to find their personal paths to knowledge” (Morville, 2008, p.18).

The goal of the test is to find overlaps of information and confusing labelling (Norman Nielsen group, 2017). During the evaluation process users have to navigate through an abstract hierarchical tree of the product taxonomy to complete given user's tasks. The metrics that the evaluator applies to understand how well the tree is designed are completion time, task accuracy, and path length (Le et al., 2014). The recommended number of tests is 15 (Martin & Hanington, 2012).

The tree is a helpful instrument in the early stages of website development because it enables evaluation of the website structure before going into the prototype design. The tree test could be

done before prototypes of interface design, i.e. before designing page layouts or navigation menus. Tree testing in such an early stage of UX design process enables “inexpensive exploration and refinement of the menu categories and labels” (Norman Nielsen Group, 2017). The visualization of information architecture and its testing at the early phases of the UX design process support the whole system development (Murugesan & Deshpande, 2001).

The process of testing is relatively easy, but requires that the evaluator would carry out the evaluation method (Martin & Hanington, 2012). The evaluator does not need the prototype of the product or even the sketches of wireframes. To conduct a tree test the evaluator must have the website information architecture (or hierarchical menu) – the tree and the scenarios for participants. The scenarios for participants have the same meaning as the scenarios in cognitive walkthrough: a certain task or set of tasks that a user should accomplish (Norman Nielsen group, 2017). Tree test scores system goals and user tasks, and possible problem areas. The method is helpful to organize correct categories and that they would “reflect the mental model of your audience” and would be labelled with words that make the most sense to the potential users (Martin & Hanington, 2012, p.26).

Martin and Hanington (2012) pointed to the importance of correct labelling to avoid vague terminology or multiple meanings that labels could be associated with. The terminology could guide the behavior. Users prefer actions that they think will accomplish their purposes. If an action seems to lead away from a purpose, users will not choose it. People follow a strategy called ‘label-following’ when they select actions whose labels are undoubtedly related to their purposes. For instance, if users have the purpose of archiving a document they will promptly select an action labelled archive and less promptly one labelled disk maintenance (Nielsen and Mack, 1994).

The method focuses only on evaluating labels of functionalities of the website, so no other aspects that form the usability and aesthetic of the website are assessed (Norman Nielsen group, 2017). In a tree test, the test participants “navigate the disembodied representation of the UI, consisting of only menu labels and their sublevels while trying to complete representative tasks. This approach isolates the conceptual component of the navigation structure from the UI, though it is recognized that elements of the UI also contributes towards navigation within the HIT system” (Le et al., 2014, p.1).

The method can be performed by different software programs. There are remote tree testing tools such as Treejack, UserZoom, Plainframe. The benefit of these methods is that they allow larger-scale testing (Le et al., 2014). The software perfectly records the user's clicks.

However, the programs I mentioned above cannot ask open-ended questions and qualitative data are lost (Norman Nielsen Group, 2017). Thus, sometimes the UX inspector can see that there are problems in the website structure, but he or she doesn't have the information on why the problems occur and what could be the resolution.

6.4.2. Empirical part 1– conclusions from the interviews with practitioners.

From the seven respondents, only one uses tree testing and three more interviewees are familiar with the method and terminology. One of the respondents who knows the method explains the reason for not using it in the following way: "I do not believe that in a tree test the user could establish an optimal hierarchy for labelling. Users often may create "another" category but as has been shown in usability tests such a category is very rarely interacted with by actual users" (Interviewee #2). He finds the method to be too specific and focused only on labelling. The respondent thinks that testing the structure of labels is not relevant for the startups's website. He mentioned that in case the startup wants to apply the method, the only right moment will be the early stage of product development when the design team is making the prototype of the website.

Another interviewee who knows the method but doesn't apply it at his work is the representative of Wärtsilä. He claims that their team always creates the tree for website labels and hierarchy but further uses it as a material for team discussions and brainstorming. He justified the absence of showing the tree to the end users by the too abstract level of the tree. End users, he said, will not be able to get the whole concept of the product, and thus, will not be able to provide valuable feedback.

However, the interviewee who uses the method stands by its effectiveness. He claims that UX designers don't fully understand the value of the tool and, thus, neglect it: "It should be really useful. And I would encourage everyone to do it. It's just that people aren't always interested in investing in it. And especially if the product works, even if it isn't optimal, priorities lie somewhere else" (Interviewee #3). He claims that the method will be useful at any stage of startup development. To properly apply the method the practitioner needs 3-4 weeks: "setting up the

research/handling stakeholders, prescreening participants, sessions with participants, presentation deck and communication - roughly 1 week each phase” (Interviewee #3). From his comment, I can conclude that the method does not seem to be cost-effective, because 3-4 weeks is a long time for startups.

One of the interviewees who does not use the tool explained that in the startup world you often use copy practice: “We copied the website tree (architecture) from the best websites. If you know that there is a company that spent so much money and resources to create perfect information architecture – you just copy it. It is not even important in which market the company is, whose websites you browse, operate. If you know they are good, they did a lot of work, test this work and they know that the outcome is perfect, so as a startuper the easiest way is to copy their tree” (Interviewee #4). Consequently, their design team didn’t have the necessity to test their tree as they copied it from other websites.

6.4.3. Empirical part 2 – The Sophia case

We carried out tree testing only partially: we created a tree (information architecture) and discussed it within our startup team. However, we did not carry out a real tree testing with users in a proper way. The tree allows us to visualize the amount of information that should be displayed on the website, the connections between different commands, the number of pages that the user should pass to accomplish his/her goal. The tree was a very useful tool for the whole team to understand the website architecture, to be able to discuss important things for Sophia’s progress. In the Sophia case our team would discuss the whole concept based on the tree because it provided a very clear visualization and, thus, grounds for discussion.

However, I intentionally skipped the real tree testing with end users, considering CW tests with prototypes more important. I wanted to test many different aspects of the prototype together, not only the labelling. I wanted to test the labels, the functionality, the interface design and how all these elements work with each other. With tree testing it would be impossible as the method focuses only on labelling.

In addition, we started the first tests with end users when the two versions of interactive prototypes were ready. Even the stages of sketches and wireframes we considered to be too abstract for end

users. Consequently, we were not ready to show our information architecture that precedes the phases mentioned above to end users, thinking that they will not be able to grasp the concept.

6.4.4. Conclusion

I used my theory-based research framework to summarize my findings of tree testing method. The framework enables visualizing the holistic tree testing process taking into account all actions from the preparation phase to the interpretation phase to perform evaluation. The framework enables considering the applicability of tree testing in startup context.

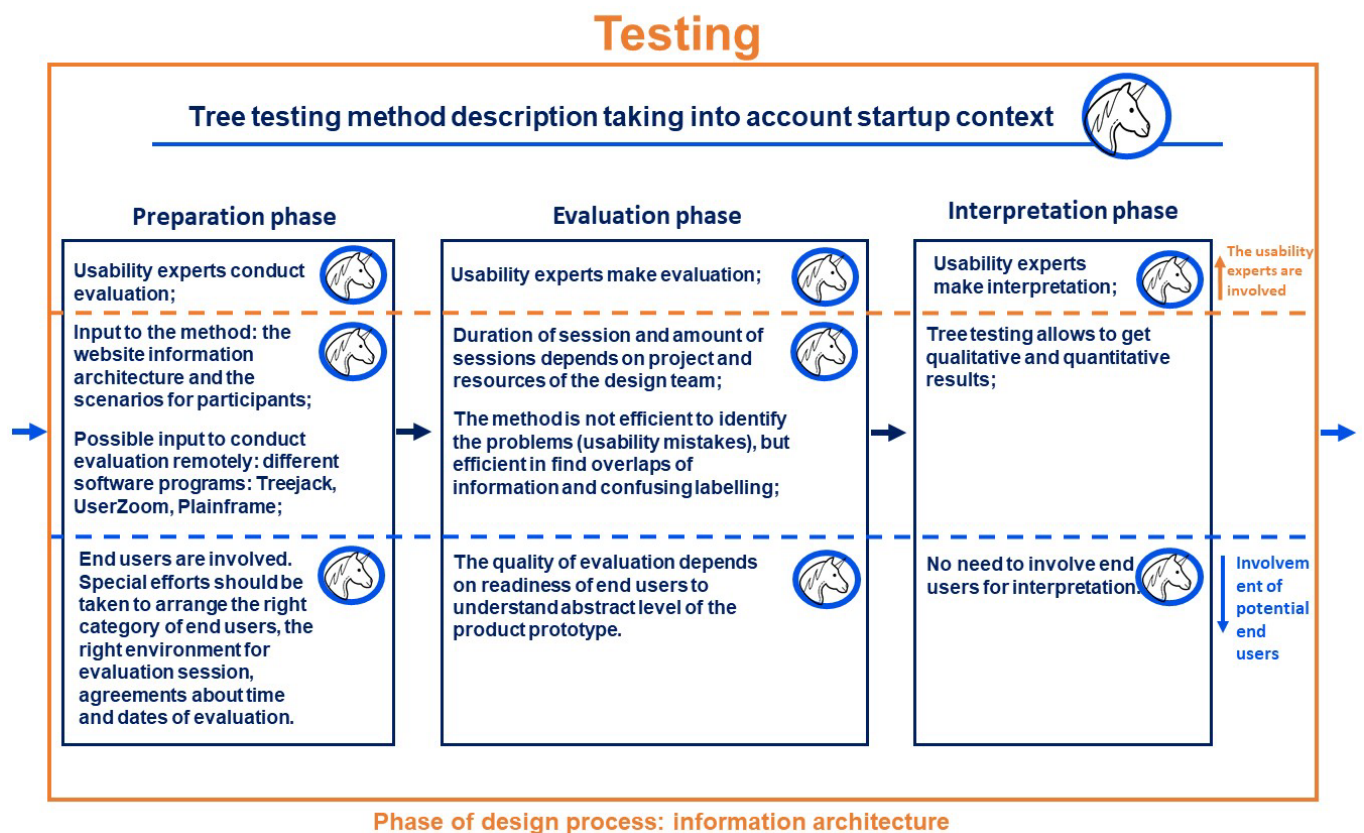


Figure 14 – The description of tree testing method using theory-based research framework to describe the UX evaluation method involving startup constraints

My theory and practice based research of tree testing revealed disagreement between researchers and field practitioners about the value of tree testing in UX evaluation process. The experience of practitioners exposed that tree testing would not be beneficial for startups when they evaluate the UX of their websites. Only one respondent out of seven uses tree testing in his work. My personal experience in Sophia also demonstrates that I did not find a specific potential in the method to investigate the UX of Sophia's prototypes. The reasons why tree testing could be neglected in UX evaluation in the case of startups are described below.

Tree testing is limited by checking only the labels and the hierarchy of the website's features. Researchers (Murugesan & Deshpande, 2001) state that tree testing is an important step in UX design. However, for the startup team it is crucial to find UX mistakes of all kinds, not only in the labelling. The goal of tree testing is to find overlaps of information and confusing labelling. Yet that could be easily checked with cognitive walkthrough. Trying to accomplish certain scenarios, users can easily see the mistakes in labelling and information repetition, but, in addition, reveal other important UX mistakes. That would be impossible with tree testing only.

Tree testing is time consuming because it requires end users to test the information architecture. As time is one of the most critical resources for startups, the majority of startups I interviewed do not apply tree testing.

Another reason for startups to put the method aside is that the level of information architecture is too abstract for end users and they might not be able to grasp the meaning of the process. Interviewee #7 said that all the levels before the interactive prototype are too abstract: "We are creating information architecture, but we do not come to the customers to test it. They will not understand information architecture because this level will be too abstract for them. We are using information architecture mostly among designers as a communication tool" (Interviewee #7).

Thus, the startup environment does not support the usage of tree testing: "Tree testing however seems to be quite specific (focused on structuring of labels) and this may not be that relevant for all digital startups" (Interviewee #2). The time allocated for the tests and the limitation of UX mistakes restricts the use of the method in the startup environment.

Tree testing could not be performed without information architecture. However, my Sophia experience and my observations from the interviews support the development of information architecture of the website. Based on my research, I agree with the academic statement that the

process of creation of information architecture at the early stages of the design process supports the whole system development (Murugesan & Deshpande, 2001), with the comment that startups will benefit from the information architecture design, but to save time and money no external testing is needed. Information architecture is a good visualization of the structure of the website that could provide good grounds to facilitate internal discussion with the design team participants, as we did in the Sophia case.

6.5. Brainstorming session

6.5.1. Theoretical part

Usually, the brainstorming session is not commonly attributed to the UX evaluation in academic literature. In the retrospective of the UX evaluation methods by Nielsen and Mack (Nielsen & Mack, R.1994) the brainstorming is not mentioned. While searching “Brainstorming in the UX design” the Google scholar search engine provides a very limited set of articles. However, the brainstorming is recognized by scholars as an inevitable part of the design thinking process that I took as framework in this research to describe UX design (Brown & Wyatt, 2010). This mismatch motivated me to investigate the role of brainstorming in the process of the UX evaluation.

Brainstorming can be defined as “an individual or group method for generating ideas, increasing creative efficacy, or finding solutions to the problem” (Wilson, 2013, p.6). The method is the oldest comparing with the other methods that I described before. Brainstorming first appeared in a book *Applied Imagination* of Alex Osborn published in 1953 as a methodology to create a lot of out-of-the-box and innovative ideas. The author defined two core rules of brainstorming: opting for quantity of ideas without thinking about quality, and avoiding judgment so that participants of the session would not censor ideas. Since this definition, brainstorming became a common way of generating ideas and innovative solutions and took on many different forms. However, the session still follows Osborn’s two rules (Burnett & Evans, 2016).

Brainstorming is a generic tool that can be used in many fields and for many reasons. In the case of UX design evaluation, the team can perform a brainstorming session for the following purposes:

- in the research phase to analyze data from user’s research and define customers’ profiles (personas) and customer’s needs;
- to ideate solutions in the development phase;

- in the development phase to test prototypes, find UX mistakes and decide how to improve them.

Within the scope of this study, I analyzed the effectiveness of brainstorming for the last purpose: to test prototypes, find UX mistakes and decide how to improve them.

The brainstorming in the development phase could be performed with or without end users. Designers can create a focus group with end users to discuss the good and bad qualities of the website or the design team might organize its own brainstorming sessions.

Regardless of the session format, to make the process more productive it is better to give the group "homework" to consider in advance. That allows people to come to the workshop already prepared to establish creative thinking (IDEO). No special preparation except the homework and welcoming environment that gives the feeling of psychological safety are needed. However, to facilitate better engagement it is good to have a pleasant environment with snacks, coffee and tea, markers and whiteboards (IDEO).

The brainstorming is a non-structured process that is adjusted every time for the specific needs of the session facilitators. However, there are certain basic rules for the session: the first principle is that during the brainstorming session it is prohibited to make judgments. "It is important to discourage anyone taking on the often obstructive, non-generative role of devil's advocate" (Brown & Wyatt, 2010, p. 6). Other rules for brainstorming are: motivate people to come up with wild ideas, develop ideas of others, stay focused on the main topic, have one conversation at a time, use visual means to explain (IDEO). The idea of brainstorming is to build an open environment for creative ideas.

However, there is a certain critique to the brainstorming with end users. For instance, the evidence that user's own interpretation of their reactions to the system is unreliable. That means that users are introspective about their experience with the system. They will reject innovative ideas if facilitator asks them. This makes brainstorming method unsuitable for significantly innovative systems. It is a mistake to rely on brainstorming in evaluation of "high-cognitive-friction products" (Cooper, 2004, p. 211).

As I wrote earlier, scholars do not provide any recommendations on how to use brainstorming in UX evaluation process, at which stage of the process, with or without end users. The main

conclusion from the theoretical analysis is that brainstorming is not seen in academia as a method to evaluate UX.

6.5.2. Empirical part 1– conclusions from the interviews with practitioners.

Based on interviews, six out of seven interviewees used brainstorming in their work. The majority of representatives of startups, the interviewee from academia and the interviewee from the big company are constantly using brainstorming in the process of UX design. I can conclude from this observation that the pattern is the same for the startup, academic and corporate environments.

The interviewee who does not currently use brainstorming is working alone, imposing limitations on the brainstorming, since for brainstorming you need to have a team. From interviews, the brainstorming method yielded contradictory comments. One interviewee stated that “we’re brainstorming all the time. Non stop. Designers with designers, designer with product owner/business lead, designer with customer, designer with developers, designer with analytics team. Really all the time” (Interviewee #3) Another interviewee was, however very critical: “In general, brainstorming is not very effective. If you put 10 people with the same background in the room and make a brainstorming session with them – the value will be zero” (Interviewee #4).

Despite his critical comment, interviewee #4 carried out a brainstorming session with his team to evaluate and improve the text of the website. He pointed out that for him it is not brainstorming itself that brings value, but the diversity of the team. He said: “if the people are diverse, it is not important how they will give feedback - during a brainstorming session or not, it will be effective”. Therefore, the key for him is not the method of gathering opinions – brainstorming or not, but to ensuring that the people will have different backgrounds. All other respondents find brainstorming to be an effective tool to proceed with UX design.

Practitioners claim that brainstorming is a very effective tool to generate ideas, but not specifically to reveal usability or website aesthetic problems: “In my opinion brainstorming is good for idea generation. I would not use it to reveal usability or design problems. Depending on participants, it might be a very weak version of heuristic evaluation, or a non-effective focus group” (Interviewee #6).

The purpose of brainstorming is affected by the stage of the project when practitioners apply it. The majority of respondents agreed that it is very effective in the beginning of the UX design process when design team collect initial data about the users and need to start sketching and wire framing first versions of the website: “Following a double diamond process model brainstorming is very effective in the beginning of a project and then once more perhaps in the middle of the project” (Interviewee #2).

From the interviews I observed that only two respondents carried out brainstorming sessions with potential end customers. Other interviewees who used the brainstorming applied it within their design teams. Interviewee #7 who reported using brainstorming with potential end users, claims that his team includes professional service designers who have good expertise in brainstorming.

To get the most out of the session brainstorming requires certain preparation: “What is important is to carefully do pre- and post- session homework: have a good detailed, not too wide, not too narrow problem to focus on, and bring ideas back to reality and users” (Interviewee #6). Besides the preparation it is very important to get the right people on board the sessions: “The trick is to get the right people to commit and be there” (Interviewee #7). That preparation and accurate choice of the participants will directly affect the cost-effectiveness of the method.

The cost effectiveness of the method is relative and depends on the number and status of participants: “the method is both cheap and expensive” (Interviewee #3). If only the design team arranges the brainstorming, the method is relatively cost-effective. However, if external people, especially users are invited, the method becomes more expensive. The design team in early stage startup often consists of the founders or people working for relatively smaller fees as in the case of established company.

One of the respondents claimed that brainstorming is an effective tool to facilitate communication inside the design team: “It’s more the glue in the company, where customer insights can be put to the table and tried out through brainstorming and convincing others at the same time. Or arguing for or against something” (Interviewee #3).

6.5.3. Empirical part 2 – The Sophia case

In the case of Sophia, brainstorming sessions were crucial for the progress of the project. We had brainstorming sessions several times: to discuss the results of user research, to design the business model of the product, to evaluate information architecture, to evaluate the wireframes of landing page, to check the UX design of the interactive prototype and to evaluate the MVP.

The scope of the brainstorming sessions from those with all team members to small ones with UX designers. During the brainstorming session we designed the information architecture of the website. We carried out brainstorming sessions to decide which interface design will set the visual style of Sophia's prototype. However, in the Sophia case we didn't carry out a brainstorming session with end users. We used the methodology only within the team. To reveal UX mistakes we preferred CW sessions with end users accompanied with discussion afterwards. We used brainstorming later on to find out the solutions to improve the revealed UX issues.

All brainstorming sessions were effective because they allowed us to share a common vision on the topic discussed and together agree on certain decisions to move forward. One of the reasons for the session effectiveness was that for each session we included the preliminary "homework": the summary of the interview insights, the high fidelity wireframes of landing pages, or the interactive prototype of MVP. Without such materials, the effectiveness of discussion could be lower.

Another reason for the effectiveness of our brainstorming workshops was that our team is diverse. We have people from different backgrounds, cultures, genders, ages, and education. Diversity allowed us to put on the table different opinions and visions.

6.5.4. Conclusion

Using my theory-based research framework, I summarized my findings of brainstorming as an evaluation method. The structure of the framework enables describing the holistic process of brainstorming from the preparation phase to the interpretation phase taking into account startup context.

Testing

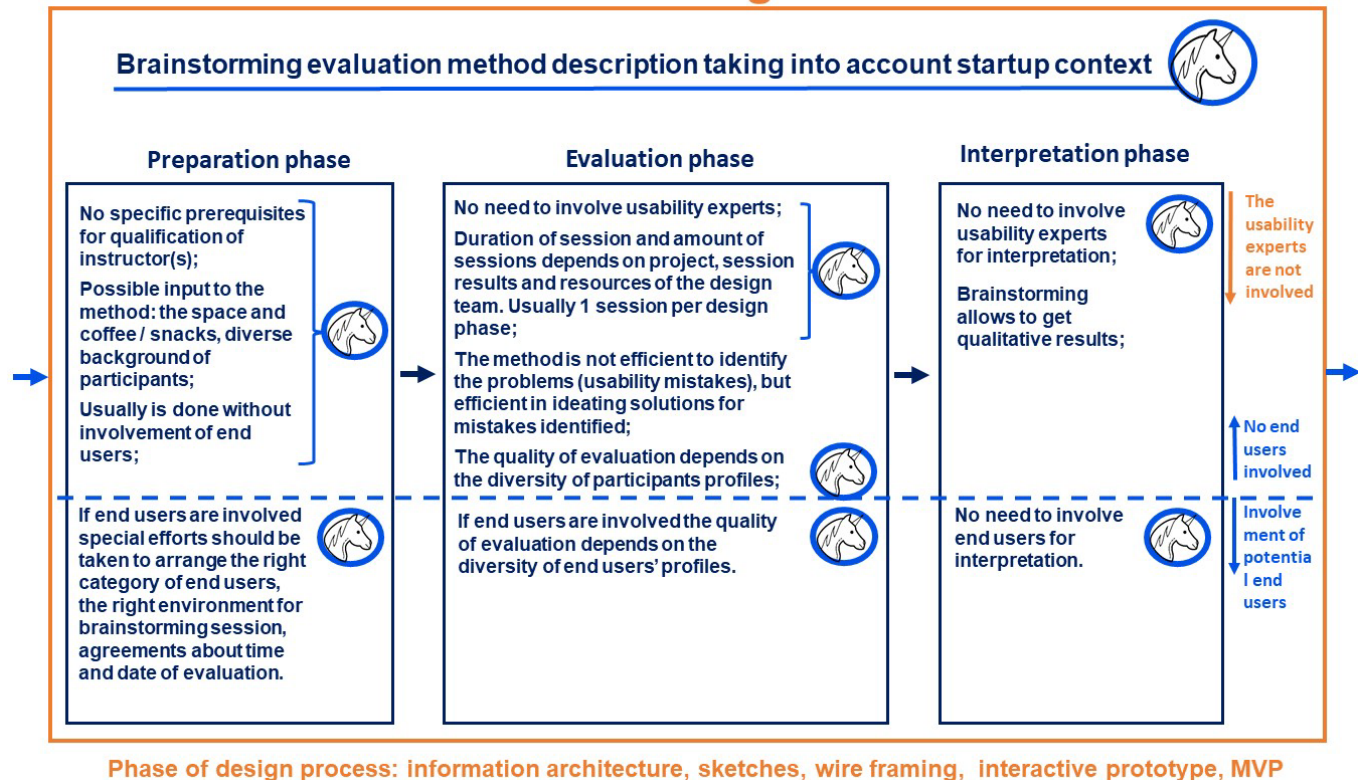


Figure 15 – The description of brainstorming method using theory-based research framework to describe the UX evaluation method involving startup constraints

The main conclusion from the theoretical analysis is that brainstorming is not seen in academia as a method to evaluate UX. The results of Sophia case and empirical analysis from interviews with field practitioners support the opinion of researchers.

Based on empirical analysis, brainstorming is generally an effective and popular method. All the respondents mentioned that the method is especially useful for concept creation after the phase of users' research that is aligned with the original purpose of the method creation. Our Sophia case proved that statement. This is the moment when the design team can analyze all the data about potential end users gathered during the interviews.

During the following stages, brainstorming can be effective in coming up with ways to fix revealed usability or UI problems. However, the method is not effective as such to perform the evaluation

to reveal usability or UI problems. The analysis of the academic literature provides the same conclusion.

Thus, designers could combine brainstorming with CW: “the brainstorming is effective to discuss how to fix usability mistakes revealed though other testing methods” (Interviewee #7). During CW, designers gather user’s feedback about usability and UI problems, during the brainstorming session the team finds the creative solutions to fix revealed problems. The team can repeat these cycles until the optimum in terms of UX website prototype is developed. The conclusion that brainstorming is not effective in revealing UX problems is aligned with the observation I made from the theoretical analysis.

Another observation from empirical analysis is that brainstorming can help to facilitate communication inside the design team. When the team runs the brainstorming session, it can establish a common understanding of the ongoing UX design process.

Brainstorming could be more effective if there is special preparation before the session and if participants are diverse to represent different opinions. The preparation work might include setting clear structure and goals for the process, choosing a group of participants relevant to the case, giving the participants tasks according to the goals before the session, and arranging an inspiring environment with coffee and snacks.

However, the extensive preparation work might require financial resources that could set certain limitations to startups. On the other hand, the better the quality of preparation, the more productive the session and the faster the startup will be able to move forward. Thus, the startup should carefully assess money and time commitment to lead the session.

6.6. The framework of Minimum Viable UX evaluation methods for digital startups

Based on theoretical and empirical research, the contribution of this thesis is to provide the framework of Minimum Viable UX evaluation methods for startups - the list of necessary UX evaluation tools that each startup, despite time, money and human resource constraints, should follow. Each evaluation method, based on my findings, is unavoidable to help the startup to progress with the product development.

To create such a framework, I produced a table that compares the usage of UX testing methods among practitioners. The numbers from one to seven refer to the interviewees. The directory “startups” represent the answers of the interviewees working in startups. The column “University” – the answer of the respondent who was working in Aalto University. The column “big company” – the answers of the UX lead in Wärtsilä – big established company. The column “Sopia” refers to the tools that we used in the Sopia case. The table visualizes which tools respondents prioritize in their work. The sample of seven interviewees already demonstrates some common patterns.

Table 6 - The usage of UX testing tools in UX design practice of interviewees and in the case of Sopia.

Evaluation tool	Startups					University	Established company	Sopia
HE	-	X	-	-	-	X	-	X
CW	X	X	X	X	-	X	X	X
SUS	-	-	-	-	-	-	-	-
Tree testing	-	-	X	-	-	-	-	-
Brainstorming	X	X	X	X	-	X	X	X

The table shows the quantitative analysis of the usage of specific UX testing tools among practitioners, not their opinions about the effectiveness of the tools. I choose the usage of tools instead of opinion as a base for quantitative analysis because I think that actions are more objective than reasoning. When a respondent claims that he/she uses the tool that demonstrates that he/she finds it effective. However, when I added the general opinions of respondents in the table about tool usage in digital startups, the difference is not significant. Only the numbers of positive responses for cognitive walkthrough and brainstorming methods increased. The updated table looks as follows:

Table 7 - The opinion about the effectiveness of UX testing tools from interview respondents and my conclusion in the case of Sophia.

Evaluation tool	Startups					University	Established company	Sopia
HE	-	X	-	-	-	X	-	X
CW	X	X	X	X	X	X	X	X
SUS	-	-	-	-	-	-	-	-
Tree testing	-	-	X	-	-	-	-	-
Brainstorming	X	X	X	X	X	X	X	X

Before interviewing the leader of the Wärtsilä team I was expecting to see that in such a big company with resources for UX design the practice will be much broader than in the case of small startups. I had the same expectation for the university environment. However, the patterns are the same in all three contexts. The big company, Aalto University and startups do not differ in their approaches to evaluate the UX design. The staff member working in Aalto University (Interviewee #6) claimed that her work had the same constraints as the work of UX designer or UX researcher in a startup. Her practice of UX evaluation proves the same.

Both tables show the predominance of certain tools. For example, brainstorming and CW have the highest scores. Thus, from that I can conclude that CW and brainstorming are necessary for the UX design process, while the necessity of HE is not obvious. The tables also demonstrate that designers could exclude from design process such tools as SUS and Tree testing because they are not relevant in the startup environment with its constraints.

Based on the research presented in the study, sections with conclusion after each UX evaluation method, and tables 6 and 7 shown above, I created a framework of the minimum viable evaluation methods (Table 8) for startups.

In my study, I was investigating how startups can evaluate the UX of their websites. Thus, I do not know if my framework will be applicable to other contexts, for example, mobile applications.

The framework represents the list of necessary UX evaluation tools that startup, despite time, money and human resource constraints, could follow to succeed in UX of its website. Each evaluation method in the framework, based on my findings, is beneficial for the startup to make progress. I investigated how cost-effective each method is, that means how effectively it reveals UX mistakes compared to the resources that startups could put into the evaluation process. I found out the methods, presented in the framework, to be relatively cost-effective.

In addition, I researched if, to perform the method, the evaluator should have special training or if the startup team can use its' own human resources. The novice UX designer can perform the methods I included in the framework. This is an important aspect for the startup environment, because startups at an early stage do not have a big team (sometimes a startup has only a founder) and possibility to hire additional professionals. Thus, the startup team could design UX using only internal resources and skills. Therefore, the UX evaluation methods should be relatively easy to learn and apply.

Further, I was studying when startup team should start the evaluation of the UX. I described each method using my theory-based research framework. The framework contains the parameter of UX design process when the method could be applied. The table 8 below summarizes my conclusions about when is the right moment to start using specific methods. In addition, taking into account the iterativity of the UX design process, I researched the need for repetition of evaluation with each particular method. To show the place of each evaluation method I use the double-diamond framework presented in chapter 3 and zoom in the part of the "Deliver" phase. I visualized the outcome of my research in the table 8 below.

Table 8 - The framework of Minimum Viable UX evaluation methods for startups

	Holistic UX process based on double-diamond model, part of the “Deliver” phase					
Evaluation tool	Information architecture	Sketches	Low fidelity prototype	High fidelity prototype	Interactive prototype	MVP
HE						
CW					X	X
SUS						
Tree testing						
Brainstorming	X	X	X	X	X	X

Explanation of the framework:

Heuristic evaluation (HE) – entrepreneurs carrying out UX design should be familiar with heuristics and just have it in mind while designing the product. Heuristics are a “self-evident” (Interviewee #7) set of basic principles to design the website. They are easy to learn and any non-professional UX designer can understand them relatively fast. Based on the research, the startup does not have to make a separate HE with at least five experts (the requirement from the author of the method), which would be time and money consuming for startup. The recommendation is “just design everything having them in mind ... incorporate them in all the templates from sketches to a clickable prototype” (Interviewee #7).

Cognitive walkthrough (CW) – cognitive walkthrough is the most common way to test UX for startups: “This is very useful. We do it all the time” (Interviewee #3). From seven respondents, five use it in practice and two claimed that they would use it if they did not know the market where their startup is operating. “Cognitive walkthrough is probably the most useful when designing a product or feature that you yourself are not a primary user of” (Interviewee #7). Cognitive walkthrough allows the gathering of sufficient data to develop appropriate UX that will satisfy customers. Based on the Sopia experience, the tool brought us the most valuable results to improve our prototype and MVP. The best moment to start cognitive walkthrough is when the first version

of the interactive prototype is ready. In the previous stages, the level of product representation (information architecture, sketch, wireframe) for the users will be too abstract. Another common agreement is that cognitive walkthrough could be used multiple times during the design process: designers collect the summary of UX mistakes after the first cognitive walkthrough, then they can fix those mistakes and conduct cognitive walkthrough again.

System usability scale (SUS) – in general, in the industry the method is not very popular: “No one here uses SUS as a word. Or at least I never heard anyone mentioning it once” (Interviewee #3). In the case of Sophia, we also neglected the tool because I found out the method to be complex and ambiguous and not bringing additional contribution in UX evaluation process in comparison to cognitive walkthrough. From both empirical analyses, I can conclude that the startup environment does not support the usage of the system usability scale.

Tree testing – entrepreneurs do not use the method because it is too complicated and abstract: “The method seems a bit cumbersome with non-negligible time spent in setting up” (Interviewee #2) and “tree testing is difficult to conduct with end users – too abstract level” (Interviewee #7). However, research supports the development of information architecture of the website – the integral part of the tree testing. Startups could benefit from the information architecture design, but to save time and money no external testing is needed.

Brainstorming - based on theoretical and empirical analysis, brainstorming is generally an effective and popular method. “It is nearly impossible to create something without a brainstorming session. So yes, brainstorming sessions are very useful” (Interviewee #2) and “It’s more the glue in the company, where customer insights can be put to the table and tried out through brainstorming and convincing others at the same time” (Interviewee #3). However, brainstorming is not the method to test for practitioners, but to generate ideas on how to improve the revealed mistakes: “It is not effective for revealing usability mistakes. It primarily serves as a tool to generate ideas on how to fix mistakes” (Interviewee #2). The brainstorming should be run during the whole process of UX design, however, not for evaluation with end users. The main purpose of the brainstorming is to find creative ideas to solve revealed usability and UI problems. Thus, designers could combine the brainstorming with cognitive walkthrough.

7. DISCUSSION

7.1. When should we start UX evaluation?

I was studying how startups should evaluate their website's UX. This research question includes a sub question when startup team should start the evaluation of the UX, at which phase of UX design process. In the framework above, I visualized at which moment of the product life span each UX evaluation tool is applicable. However, in academia and between practitioners the question of how to find the balance between prototype evaluation, MVP evaluation and real product evaluation is still open. If we will make a proper evaluation of the early prototype, as we did in the case of Sophia, we will fall into the situation where we will reveal so many mistakes. For example, in HE evaluation the “back” button should be at every page (heuristic 1 - visibility of the system). However, for the prototype and MVP it is not important whether we have a “back” button or not. The “back” button is not a core function that affects the value delivery to the user. It is not crucial for task performance.

Researchers state that it is rather simple to evaluate UX with existing systems, but it is more difficult to evaluate UX with an early product concept or a single feature that is only a prototype yet (Law et al., 2007). However, there is a controversial opinion: “The usability discipline tacitly hands the reins to the programmers, saying ‘You build it, and then I’ll test to see how well you have done. However, in this fast-moving, high-tech world, after it is built, it ships’” (Cooper, 2004, p.205). Post-facto UX evaluation does not have much impact on the system. This statement supports the logic of the iterative UX design process that UX evaluation should start as soon as possible with the first prototypes of product (Terence, 2003).

The good UX requires much effort and workload to complete it, but, later on, it will be easier to collect and analyze users’ feedback. The low quality UX of the prototype is less laborious than the high quality one, but the excessive workload will transfer to collecting and analyzing users’ feedback. Whilst evaluating business potential of the future system with real users to diminish risk, gathering valuable feedback with early system prototypes could be difficult. One challenge is that not fully developed or yet confusing user experience could weaken the user feedback and make the users focus mostly on the interface design. At worst, low quality UX could provoke the user to criticize the UX even if the system itself or the product idea were good (Hokkanen et al., 2016).

To illustrate that reasoning, I created figure 17 below to illustrate the correlation between the quality of UX and the quality of user's feedback during UX evaluation.

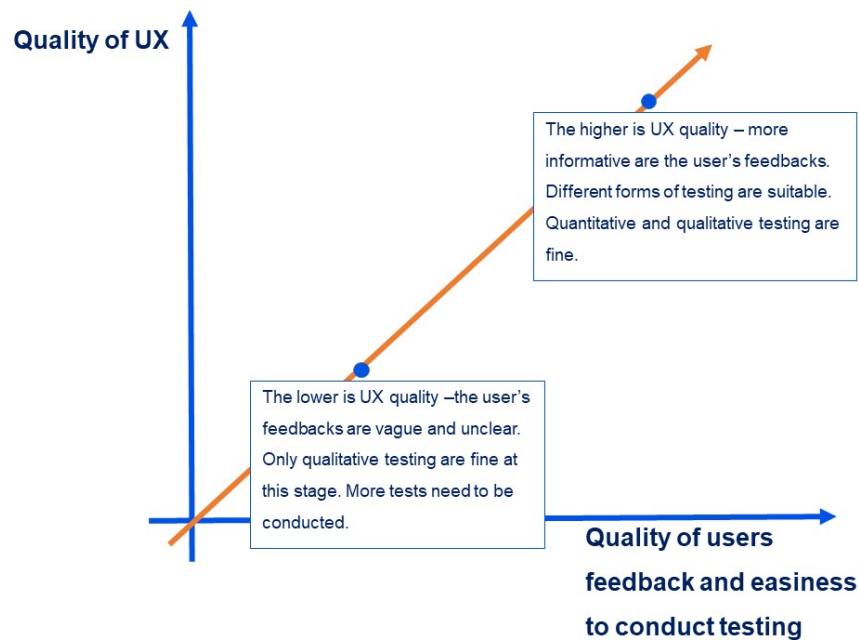


Figure 16 - The correlation between the quality of UX and the quality of user's feedback during UX evaluation

The graph visualizes the direct correlation between the quality of UX and the feedback from users. The higher is the quality of the UX of the product version (or prototype) – the higher is the quality of users' feedback. The high quality of users' feedback means that the users' comments address exactly the UX shortcomings in the system: how easy it is to use the system for novice user, how easy is to accomplish given task and how comfortable it is to use the system. The good quality of users' feedback allows getting a precise view on what should be improved in the system to develop a great UX.

Researchers do not provide a precise answer to the question when one should start the evaluation, but instead a set of recommendations. For example, startups should start talking with potential users as soon as possible. Talking with potential users is the most common approach to gather feedback for system improvements. However, the crucial prerequisite here is to find the right people to whom to show the early versions of the product (Hokkanen et al., 2015).

In the Sopia case, we selected a strategy to start UX evaluation at the earliest stages of product development before actual programming of the product. An early brainstorming session to discuss the landing page designs as well as first CW with end users already provided us with valuable results. We assumed that if we will be able to fix major mistakes in the prototype, the smaller will be the risk that we will have to reprogram the product.

I concluded for myself, that the best approach for evaluation, taking into account the logic of iterativity of the design process, is to compare the cost for correcting the mistakes in already programmed products to the costs of evaluating the prototype and the opportunity costs originating from the postponed entry to the market. Every case is individual and depends on the market situation, the resources and goals of the startup to a certain moment in the market.

7.2. Factors that affect UX

Another discussion topic is the role and interconnection of different aspects that all together form the UX. Many academics agree about the subjective nature of UX. That UX is influenced by such factors as “the user’s internal state, the context, and the perceived image of the product” (Väänänen-Vainio-Mattila, 2008, p.1). In the figure below, I map out different factors that affect the UX. For example, if the user is already loyal to the brand, he or she most likely will evaluate positively his UX with the product. A good example is Apple. Even if some Apple products often lack usability, they still attract customers because of brand loyalty (Cooper, 2004). Nielsen and Mack state that prior knowledge of the brand could be beneficial. The user easily learns and doesn’t forget the information associated to prior knowledge comparing to the completely new information (Nielsen & Mack, 1994).

Novice users and experienced users have different UX with the system. Users become used to the minor usability and aesthetics issues after encountering the system multiple times. This could be explained by the rationale that if a usability issue does not affect users’ goal completion, it becomes

less impactful over time as users get used to the product. But, these usability issues usually frustrate beginners (Yen & Bakken, 2009).

Another important aspect is the aesthetics of websites. The traditional notion of usability does not include aesthetic aspects. Generally, usability has underlined that for the novice user the process of interacting with the system should be smooth, fast and easy so that the user will not drop it. Saying differently, usability reflects how well the system fits the user's goals (Sutcliffe et al., 2009). However, aesthetics is a significant part of UX design and it impacts directly on the user perception of the interaction with the system. Users tend to forgive poor usability in more aesthetically appealing versions of system. Trying to connect usability with the aesthetics, we can talk about a halo effect – “by which judgement of one quality can spill over into another” (Sutcliffe et al., 2009, p.12).

The aesthetic is especially important for first-time users because it provokes an emotional response. However, if we are using the website more or less frequently, we tend not to pay attention to aesthetics, but expect the system to help us solve our tasks easily. Consequently, usability becomes more important at this point.

User's goals impact on user satisfaction: whether the users are seeking usability of the product or the pleasure and entertainment. Some websites are designed for pleasure – such as computer games, but some to solve concrete tasks, e.g. to pay a specific service, to order an item, to achieve a goal. In a work environment where digital systems use is compulsory, the users will evaluate the UX based primarily on the level to which the product enables productivity. Contrary, productivity would not be the main evaluation factor for computer games players or for those who browse the Internet for entertainment. Consequently, it would be reasonable to think that the user satisfaction would be based on quite distinctive criteria in the two environments (Lindgaard & Dudek, 2003). For instance, in the case of Sopia, we were creating a website that users do not access on an everyday basis but with specific purpose, which is to find or rent out an apartment. Thus, the goal of our UX was not to entertain the users, but to help them effectively accomplish their tasks such as to find an apartment or a tenant.

To visualize different factors that affect UX I created a figure 17 below:

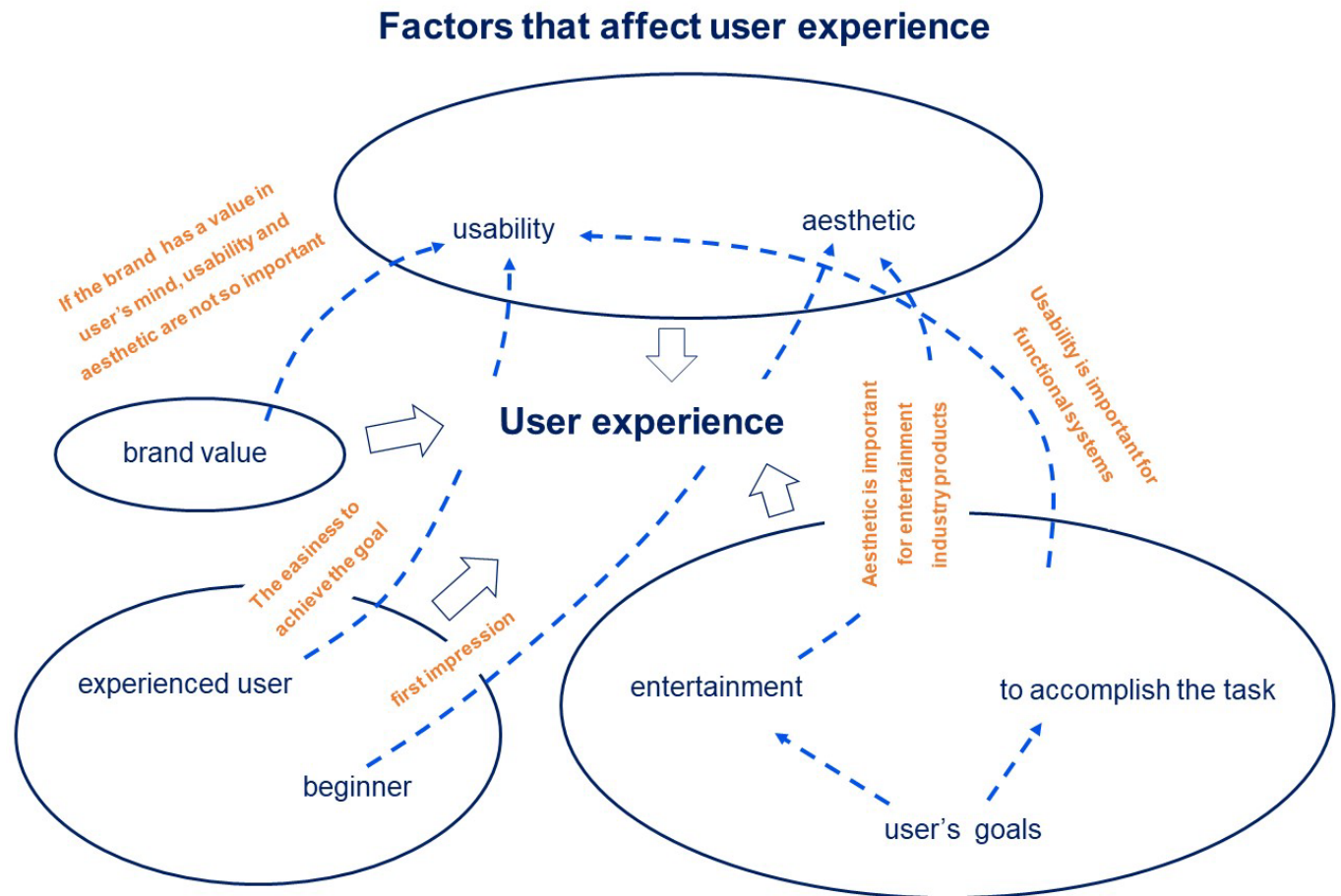


Figure 17 - Factors that impact UX

The UX could be defined as a flow of emotions that the user feels while interacting with your product, website, or system (Kraft, 2012). Emotions are formed by the quality of usability of the product and quality of aesthetics. The graph above shows that UX depends whether the user is beginner or experienced. If user sees the system for the first time, he/she will be more focused on aesthetics of the system comparing to the experienced user, who will concentrate more on usability of the product.

Another factor is brand value. If the company has well established brand, users, if they are already loyal customers, tend to excuse low quality usability and aesthetics.

Final set of factors that I visualized in the graph represents the types of user's goal while interacting with the product. If user wants entertainment, he/she will pay attention to the aesthetics of the

system. In contrary, if user has specific task to accomplish in a minimum possible amount of time, the prime attention will be on the usability of the system.

The discussion presented above revealed an important limitation in the Sophia case during the UX evaluation with end users. In the middle of the Sophia design process, we had two prototypes representing two different design concepts: one supposed to be more traditional with design cognates to other rental housing platforms; another design was more abstract and innovative in a way, without photos of happy people in the apartment from the landing page. During discussions with end-users after the CW tests we noticed that the majority of testers felt more comfortable with traditional versions, as it was more familiar for them. “I’ve seen that before!” – one of the quotes from the user. This observation proves that it is difficult to test an innovative concept because the users will not be able to associate the product with something that they already know.

The evaluation methods described above test, basically, the usability of the product. However, the user experience is broader than usability (as I explained earlier). Researchers argue that designers have to know how to design the products that will enhance people’s subjective experience and, consequently, be able to test how well the product is designed in relevance to the emotional design (Law et al., 2007).

The following methods to evaluate how aesthetics affect the UX of the users:

- Different variations of questionnaires;
- heart rate and dermal activity as physiological measures;
- EMG to assess facial expressions;
- an application that asks users to report their emotions and feelings at a given time of the day for a certain trial period;
- Interviews.

In the case of a startup, it is difficult to build a complex framework, especially with special medical devices to track the physical state of users. The brainstorming sessions and interviews with end users are more affordable.

8. CONCLUSION

8.1. Research summary

The importance of UX evaluation is commonly recognized: “Testing prototypes keeps you from charging down the wrong path, and spending way too much money and other resources on a potential solution only to discover that it doesn’t work like it should” (Ingle, R.,2013 p.11). However, the startup world and academic perception differ regarding the issue of UX evaluation. In the academia, researchers developed many methods to create high quality UX. However, many of them are not a priority for startups.

Based on my analysis, researchers do not consider evaluation method based on the time and resources that the method requires to be able to evaluate the UX. The main factor of effectiveness of investigated method for researchers is type and quality of UX mistakes revealed. However, startups have more parameters to take into consideration such as parameters described in my theory-based research framework (see corresponding chapter) below:

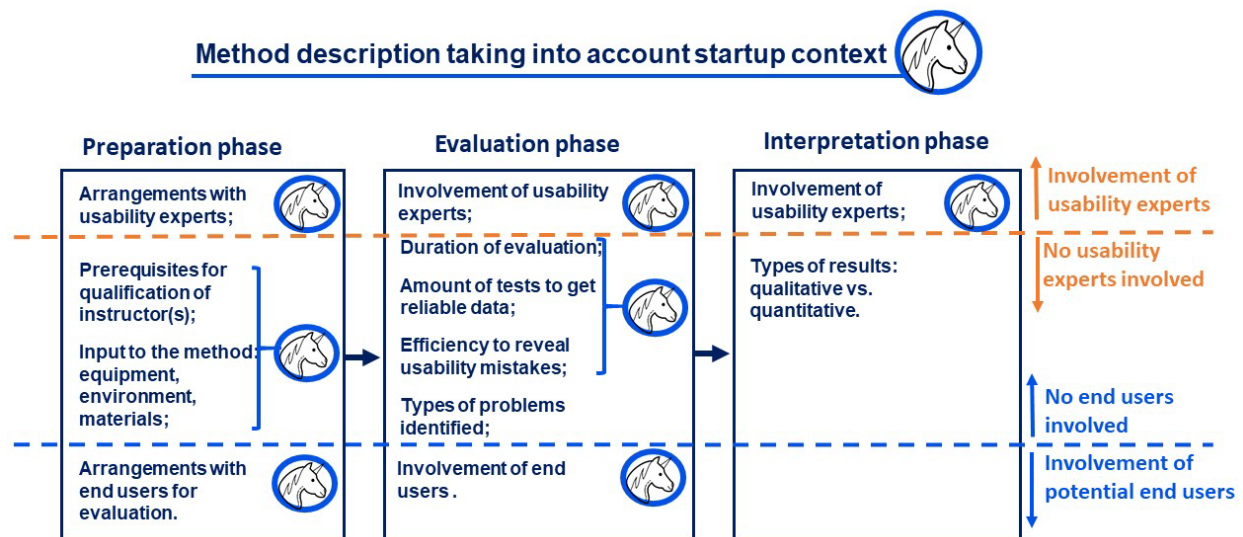


Figure 5 – Theory-based research framework to describe the UX evaluation method involving startup constraints

Startup constraints significantly affect the decisions of entrepreneurs to use or avoid the usage of a certain method. Startups have their perceived hierarchy of things to be carried out, such as the development of MVP as soon as possible. Consequently, they dismiss many UX testing methods,

thinking that they will not provide significant value compared to the amount of resources that they require. They do not take time to learn those methods because of lack of financial resources and time pressure. This dismissal of evaluation methods that require time, specific qualification of evaluator, specific preparation and substantial financial resources is a logical step if the company lives under the startup constraints.

Startup constraints guide startups to use the copying practice. Instead of putting time and human resources into UX evaluation, startups take the best UX market examples that are known to be excellent and copy them. It saves time and money and allows the creation of a good, but not unique, product.

Another way for startups to deal with UX evaluation is to use software to create a good website that already includes basic UX principles. For example, wordpress.com/ is commonly used. Their slogan is “create a website in minutes”. The UX evaluation methods themselves are now transforming into software that outsources the inspection work. In the future, the situation will be most likely that all the evaluation methods will be automated. An example is Treejack for tree testing and unmoderated cognitive walkthrough tests.

8.2. Practical implications

Based on my findings represented in the framework of Minimum Viable UX evaluation methods (see the corresponding section above) for startups, three methods from all five that I investigated could be useful in the startup context but in different format: heuristic evaluation, cognitive walkthrough and brainstorming. Practitioners tend to get rid of the complexity and special requirements of original usability inspection methods and are navigating towards more flexible, fast and simple tests. Cognitive walkthrough and brainstorming match this trend. Cognitive walkthrough allows getting the real potential end users on board, as they can provide valuable feedback and reveal flaws in the product prototype at an early stage. Brainstorming carried out after, inside the design team helps to find resolution for revealed usability problems.

In the majority of academic articles, the main conclusion is that it is useful to combine different testing methods, as they reveal different problems. Studies proved that both experts and end-users can reveal UX issues. However, they catch different UX perspectives. The results of researchers consistently demonstrate that usability experts report more commonly interface design problems

while end-users reveal serious interface interferences to their task accomplishment (Yen & Bakken, 2009). Methods that involve the participation of experts are good to identify the problems compared to the field expertise and benchmarks. However, the evaluation with end users helps to reveal crucial task performance problems. "To provide the most effective and thorough usability evaluation result, the combination of usability evaluation techniques from both expert and system user perspectives is recommended" (Yen & Bakken, 2009, p.4). This is why, based on my research, I would include the heuristic evaluation method in the evaluation toolbox. I would recommend the design team to study 10 basic usability heuristics to be able to conduct heuristic evaluation or design the products having heuristics in mind. The heuristics are easy to learn and any non-professional UX designer can understand them relatively quickly. This would fulfill the requirement to evaluate the product both from the eyes of experts and from the point of view of users.

8.3. Limitations of the study

This study has several limitations. First, the set of evaluation methods that I investigated is limited by four usability inspection methods such as heuristic evaluation, cognitive walkthrough, system usability scale and tree testing and a brainstorming method that is not attributed by scholars to usability evaluation. The main foundation for the selection of methods were my studies in Aalto University during UX training. However, I acknowledged that the methods I investigated represent only a minor part of all possible UX evaluation methods that include usability evaluation methods, aesthetics evaluation methods, user interface (UI) evaluation methods etc. Thus, the future research to collect, compare, analyze the efficiency of those methods in general and in startup environment is needed.

Another limitation of my study is that currently it is too early to evaluate the success of the case company - Sophia as the startup is at the early stage of development. The Sophia website is still under development and the design team keeps working on UX improvements. Consequently, the final results of UX evaluation of the prototype of Sophia website that our design team carried out would be seen only in a long-run perspective when Sophia will start serve real customers. Thus, it might be useful to further track Sophia's development to get an opinion about quality of UX of Sophia's system.

8.4. Suggestions for further research

The rise of startups are the dominant topics in much of the academic and business discourse in addition to the government debate (Davila et. al, 2015). At the same time, the entrepreneurial mindset started to influence established companies (Ries, 2017). Thus, the practice of running the business by analogy to the entrepreneurial approach is getting more popular. Typical startup constraints became important for new companies and for new projects in established companies. The data in this study about the practice of carrying out UX evaluation described by interviewee #6 and interviewee #7 who represent university and established company environment supports statement above.

Every new product, system, website possesses certain UX. The better UX is – the more chances the product has to succeed on the market. UX evaluation methods are crucial components that enable the entrepreneurial success (Sturm et al., 2017). UX evaluation protects entrepreneurs from moving down the wrong path, and spending limited resources on a potential UX only to discover later that it doesn't attract and retain users (Ingle, R., 2013).

However, my study revealed that entrepreneurs neglect many UX evaluation methods being affected by constraints of startup environment. Thus, I see the main potential for further research in systemic analysis of existing UX evaluation methods and their adaptation for entrepreneurial practices. Researchers could create a pool of easy to learn, fast and cost-effective UX evaluation methods that would fit to the iterative product development process.

9. REFERENCES

- Aberdeen, T. 2013 Review essay. Yin, R. K. (2009). Case study research: Design and methods. Canadian Journal of Action Research. Vol. 14:1. P. 69-71 DOI: <https://doi.org/10.33524/cjar.v14i1.73>
- Agile alliance. Available at: <https://www.agilealliance.org/agile-essentials/>. (Accessed: 7 October 2019)
- Bangor, A., Kortum, P. T., Miller, J. T. 2008. An Empirical Evaluation of the System Usability Scale Usability Scale. Intl. Journal of Human–Computer Interaction. Vol. 24:6. P. 574-594. ISSN: 1532-7590
- Berg, B. L. 2001. Qualitative research methods for the social sciences. 4th ed. Needham Heights, MA, US: Allyn & Bacon. ISBN 0-205-31847-9
- Bevan, N., Carter, J., & Harker, S. 2015. ISO 9241-11 Revised : What Have We Learnt About Usability Since 1998? In: Kurosu, M. (ed) Proceedings of the 2015, Human-computer interaction: 17th international conference, HCI International, Los Angeles, CA, USA, August 2-7, 2015. Springer International Publishing Switzerland 2015. Human-Computer Interaction, Part I, HCII 2015, LNCS 9169 P. 143–151. ISBN: 9783319209012
- Blažica, B. 2014. Use of UX and HCI tools among start-ups. In: Novak, F., Blažica, B., Bohak, C., Čehovin, L. (Ed.) Proceedings of the 17th International Multiconference information society – IS 2014. London, UK. 10.11-12.11.2014. Ljubljana, Slovenia: Institut Jožef Stefan. Vol. H. P. 22-25. ISBN 978-961-264-078-1
- Brooke, J. 2013. SUS : A Retrospective. Journal of usability studies. Vol. 8:9. P. 29–40. ISSN: 1931-3357
- Brown, B. T., Wyatt, J. 2010. Design Thinking for Social Innovation. Stanford Social Innovation Review. Stanford, CA, US: Leland Stanford Jr. University. ISBN: 94305-5015
- Burnett, B., Evans, D. 2016. Designing Your Life: How to Build a Well-Lived, Joyful Life. 1st Ed. New York, US: Knopf. ISBN-10: 9781101875322
- Cockton, G., Woolrych, A. 2002. Sale Must End : Should Discount Methods be Cleared off HCI' s Shelves ? Journal Interactions. Vol 9:5 P. 13–18. DOI: 10.1145/566981.566990

- Cooper, A. 2004 The inmates are running the asylum. Indianapolis, Indiana, US: SAMS publishing. ISBN 0-672-32614-0
- Creswell, J.W., 2009. Research design: Qualitative, quantitative, and mixed methods approaches. 4th Ed. Thousand Oaks, California, US: Sage. ISBN 978-1-4522-2609-5
- Davila, A., Foster, G., He, X., Shimizu, C. 2015. The rise and fall of startups: Creation and destruction of revenue and jobs by young companies. Australian Journal of Management. Vol. 40(1). P. 6 –35. DOI: 10.1177/0312896214525793
- Encyclopedia Britannica. Available at: <https://www.britannica.com/topic/cognition-thought-process>. (Accessed: 11 October 2019)
- Hassenzahl, M., Tractinsky, N. 2006. User experience - a research agenda. Journal Behaviour & Information Technology. Vol. 25:2. P. 91-97. ISSN: 0144-929X. DOI: 10.1080/01449290500330331
- Hertzum, M., Jacobsen, N. E. 2003. The Evaluator Effect: A Chilling Fact About Usability Evaluation Methods. International Journal of Human-Computer Interaction, Vol.15:1. P. 183-204. ISSN: 1532-7590
- Hokkanen, L., Kuusinen, K., & Kaisa, V. 2016. Minimum viable user experience : a framework for supporting product design in startups. In. H. Sharp and T. Hall (Eds.). International Conference on Agile Software Development XP 2016, Edinburgh, United Kingdom. 24-27.5.2016. Springer, LNBP Vol. 251, P. 66–78, 2016. ISBN: 978-3-319-33515-5
- Hokkanen, L., Kuusinen, K., & Väänänen, K. 2015. Early Product Design in Startups : towards a UX strategy. In: Part of the Lecture Notes in Computer Science book series LNCS. Vol. 9459 for the 16th International Conference on Product-Focused Software Process Improvement, PROFES 2015. Bolzano, Italy. 2.12 – 4.12.2015. Abrahamsson et al. (Eds.). Switzerland: Springer International Publishing. P. 217–224. DOI: 10.1007/978-3-319-26844-6
- IDEO. Available at: <https://www.ideo.com/pages/brainstorming> (Accessed: 15 October 2019)
- Ingle, B. R. 2013 Design Thinking for Entrepreneurs and Small Businesses: Putting the Power of Design to Work. 1st ed. Berkeley, CA, US: Apress. ISBN: 978-1-4302-6182-7

ISO 9241-210:2010. Available at: <https://www.iso.org/obp/ui/#iso:std:iso:9241:-210:ed-1:v1:en>. (Accessed: 6 October 2019)

Khajouei, R., Esfahani, M. Z., & Jahani, Y. 2017. Research and Applications Comparison of heuristic and cognitive walkthrough usability evaluation methods for evaluating health information systems. *Journal of the American Medical Informatics Association*, 24:1. P. 55-60. doi: 10.1093/jamia/ocw100 ISSN: 2749-7799

Kraft, C., 2012. *User Experience Innovation: User Centered Design that Works*. 2nd ed. Apress, New York. ISBN 10: 1430241497 / ISBN 13: 9781430241492

Law, E., Vermeeren, A., Hassenzahl, M., Blythe, M. 2007. Towards a UX Manifesto. In: *Proceedings of The 21st British HCI Group Annual Conference University of Lancaster*. Lancaster, UK. 3.09 – 7.09.2007. Lancaster, UK: COST. ISBN: 978-1-902505-95-4

Le, T., Chaudhuri, S., Chung, J., Thompson, H. J., & Demiris, G. 2014. Tree testing of hierarchical menu structures for health applications. Vol. 49:1. P. 198–205. ISSN: 1532-0464

Lindgaard, G., Dudek, C. 2003. What is this evasive beast we call user satisfaction ? *Interacting with Computers* Vol. 15. P. 429–452 ISSN: 0953-5438

Mahatody, T., Sagar, M., & Kolski, C. 2010. State of the Art on the Cognitive Walkthrough Method, Its Variants and Evolutions. *International Journal of Human-Computer Interaction*. Vol. 26:8. P. 741-785. DOI: 10.1080/10447311003781409

Martin, B., Hanington, B., 2012. *Universal methods of design. 100 Ways to research complex problems, develop innovative ideas, and design effective solutions*. Beverly, MA. US: Rockport publishers. ISBN: 978-1592537563

McLellan, S., Muddimer, A., Peres, S. C. 2012. The Effect of Experience on System Usability Scale Ratings. *Journal of Usability Studies*. Vol. 7:2. P. 56–67. DOI: 10.1177/1541931213571042

Meier, E., Böhm, P., & Wolff, C. 2017. Comparing Heuristic Walkthrough and User Studies in Evaluating Digital Appliances. In: Gäde, M., Trkulja, V., Petras, V. (Eds.): *Everything Changes, Everything Stays the Same? Understanding Information Spaces*. *Proceedings of the 15th International Symposium of Information Science (ISI 2017)*, Berlin, Germany. 13.03 —15.03. 2017. Glückstadt: Verlag Werner Hülsbusch, P. 146—157. DOI: 10.18452/1453

- Merholz, P. 2007. Peter in Conversation with Don Norman About UX & Innovation. Audio material. Available at: <https://vimeo.com/2963837> (Accessed: 6 October 2019)
- Molich, R., & Nielsen, J. 1990. Heuristic evaluation of user interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Seattle, Washington, USA. 1.04-05.04.1990. New York, NY, USA: ACM. P. 249–256. ISBN:0-201-50932-6
- Morville, P. 2008. Information Architecture on the World Wide Web. 1st ed. Sebastopol, Canada: O'Reilly Media. ISBN: 1-56592-282-4
- Murugesan, S., Deshpande, Y. (Eds.). 2001. Web Engineering. Managing Diversity and Complexity of Web Application Development. Berlin; Heidelberg; New York; Barcelona; Hong Kong; London; Milan; Paris; Singapore; Tokyo: Springer. Springer. ISBN 3-540-42130-0
- Nielsen, J., Mack, R. L. 1994. Usability Inspection Methods. New York, NY, USA: John Wiley & Sons, Inc. ISBN: 0-471-01877-5
- Nielsen J. 1995 Usability inspection methods. In: Proceedings of Conference Companion on Human Factors in Computing Systems. ACM. Denver, Colorado, USA. 7.05 - 11.05.1995. New York, NY, USA: ACM. P. 377-378. ISBN: 0-89791-755-3
- Norman Nielsen Group. Available at: <https://www.nngroup.com/articles/ia-vs-navigation/> (Accessed: 9 October 2019)
- Paternoster, N., Giardino, C., Unterkalmsteiner, M., & Gorschek, T. 2014. Software development in startup companies : a systematic mapping study. Journal of Information and Software Technology. Vol. 56. P. 1200–1218 ISSN 0950-5849
- Ries, E. 2017. The startup way: how modern companies use entrepreneurial management to transform culture and drive long-term growth. New York, US. Crown Publishing Group. ISBN 9781101903209
- Sutcliffe, A. 2010. Designing for User Engagement: Aesthetic and Attractive User Interfaces. Penn State University, PA, US: Morgan & Claypool Publishers series. (Synthesis Lectures on Human-Centered Informatics) ISBN: 9781608450244
- Terence, A. 2003. Criteria For Evaluating Usability Evaluation Methods. International Journal of Human-Computer Interaction. Vol. 15:1. P. 145-181.

<https://doi.org/10.1207/S15327590IJHC1501>

The British Design Council. Available at: <https://www.designcouncil.org.uk/news-opinion/design-process-what-double-diamond> (Accessed: 6 October 2019)

Tschimmel, K. 2012. Design Thinking as an effective toolkit for innovation. In: proceedings of the XXIII ISPIM Conference: Action for Innovation: Innovating from Experience. Barcelona, Spain. 17- 20.06.2012. ISBN 978-952-265-243-0

Unger, R., Chandler, C. 2012. A Project Guide to UX Design: For user experience designers in the field or in the making. 2nd ed. Berkeley, CA, US. New riders. ISBN 13: 978-0-321-81538-5. ISBN 10: 0-321-81538-6

Väänänen-Vainio-Mattila, K. 2008. Now let's do it in practice : user experience evaluation methods in product development. In: proceedings CHI'08 Extended Abstracts on Human Factors in Computing Systems. April 5.04 -10.04.2008 Florence, Italy: ACM New York, NY, USA. 2008. P. 3961–3964. ISBN: 978-1-60558-012-8

Wharton, C., Je, R., & Fran, A. 1992. Applying cognitive walkthroughs to more complex user interfaces: experiences, issues, and recommendations. In: Proceeding of the SIGCHI Conference on Human Factors in Computing Systems. Monterey, California, USA. 3.05 – 7.05.1992. New York, NY, USA: ACM. P. 381–388. ISBN: 0-89791-513-5

Wilson, C. 2013. Brainstorming and Beyond. A User-Centered Design Method. 1st Ed. Oxford, UK. Morgan Kaufmann. ISBN: 978-0-12-407157-5

Yen, P., & Bakken, S. 2009. A comparison of usability evaluation methods : heuristic evaluation versus end-user think-aloud protocol – an example from a web-based communication tool for nurse scheduling. In: proceeding of AMIA 2009, American Medical Informatics Association Annual Symposium. San Francisco, CA, USA, 14.11 -18.11.2009. P. 714–718. PMID: 20351946. PMCID: PMC2815403.

Zainal, Z., 2007. Case study as a research method. Journal Kemanusiaan Vol.5:1. EISSN: 2600-755X

10. APPENDICE

10.1. Heuristic evaluation results

Evaluator #1 (Lidia Borisova)

Date of HE – the 28 January 2019

Project status – MVP is ready but still needs minor correction. HE I did before MVP should be presented to end users.

Rating scale (Nielsen, Mack, 1994, p. 49)

0 – I don't agree that this is a usability problem at all

1 Cosmetic problem only – need not be fixed unless extra time is available on project

2 Minor usability problem – fixing this should be given low priority

3 Major usability problem – important to fix, so should be given high priority

4 Usability catastrophe – imperative to fix this before product can be released

Heuristic evaluation results based on the list of the heuristics:

1. Visibility Of System Status

Rate – 3. The visibility of the system is insufficient.

After clicking «get started» we are on the second page with the options to choose if I am a tenant or a landlord. However, the menu bar disappears at this point that doesn't give me the feeling of understanding where I am, or how to go back or have more options. Thus, the visibility of the system is insufficient.

At some pages MVP had a “back” button, at some – not. The “back” button should be set at every page of the system.

The system status is visible in all the pages when tenant had to fill the form - the bars on the left show the overall length of the process - 4 blocks to fill - and underline with which of the 4 blocks the user is interacting at the moment. In addition, the visibility is supported by the possibility to switch from one block to another in the order, that user chooses by him/herself.

However, the same visibility is absent on the landlord side. It is impossible for landlord to get back to the landing page before completing the whole process of signing up. The bar indicating the

length of the process (similarly to the tenant's side) appears only in the "make an announcement" category.

2. Match Between System And The Real World

Rate – 2. The language of the system reminds about other services on the rental housing market. However, a few mismatches in vocabulary were found.

Your profile page (tenant's side)

"Sopia needs this information to match you with the landlord" – might sound strange for the users. I would suggest adding tutorials at this moment to explain more why we need this information and what "matching with landlord" means in practice.

Page with matches (tenant's side)

"match with landlord" and "match with property" is not clear for novice users. We should add tutorials explaining the logic of the system. This "double matching" is the feature that makes our system different from competitors, but more efforts should be put to explain the feature to users.

Page with matches (landlord's side)

Same confusion as on the page with matches at tenant's side. Sopia's team should add tutorials explaining why there are 2 matching columns. Now "match with property" and "match with your requirements" don't provide a clear understanding that the first column is about apartment parameters, and second is about ideal tenant's profile.

Dashboard page (both tenant's and landlord's side)

"dashboard" – I didn't see this label on other websites. It is not clear for me what it is. I would suggest change it for "your profile".

Information page (both tenant's and landlord's side)

"information" label is not clear, because the user will need to understand information about what? Labelling is ambiguous

3. User Control And Freedom

Rate – 0. The heuristic is working due to "back" buttons, menu dashboard and by showing me the location that I typed to let me verify if I put information correctly (tenant's side).

This feature is supported by “back” buttons, menu dashboard and by showing me the location that I typed to let me verify if I put information correctly (tenant’s side). The same verification statement the system provides at landlord’s side. Therefore, the heuristic is working.

4. Consistency And Standards

Rate – 3. The website has a design consistency: colours, type fonts, shapes of the bars etc. However, the feature with double matching that distinguishes our service from competitors might be difficult to understand for novice users as they didn’t see that double matching before.

The Sophia’s team should not avoid this feature in order to make the website more standardized. The double matching forms Sophia’s unique value proposition for end users. However, so that to make the system less confusing for beginners, we should add tutorials explaining the logic behind the double matching feature. Because currently, the user might wonder what different words, call to action mean, do they mean the same thing or different.

5. Error Prevention

Rate - 0

This feature is supported by showing me the location that I typed to let me verify if I put information correctly (tenant’s side). The same verification statement the system provides at landlord’s side. In addition, the possibility for the landlord to make the announcement public or to hide it if he or she wants to change the information. In addition, in both profiles (tenant and landlord) MVP has a function to modify information. Consequently, the feature is working.

6. Recognition rather than recall

Rate - 0

In general, the walk through the system is fast and easy for both sides. It is not overwhelmed by information. The navigation bar on the left helps the user to understand at which place in the process he or she is.

The sign in function is also very easy and requires only a few clicks.

Thus, the heuristic is well developed in the system.

7. Flexibility and efficiency of use

Rate - 0

When the user has an account, he directly falls into the matches page where he/she can easily switch from matches to his/her profile information. That function provides the feeling of flexibility.

8. Aesthetic and minimalist design

Rate - 3

The website has a design consistency: colours, type fonts, shapes of the bars etc. Now, MVP doesn't contain irrelevant information. Opposite, MVP doesn't contain enough of relevant information. Sopia website has to have tutorials, text, describing the overall logic of the system and, in addition, more parameters for apartments to make more precise match.

9. Help users recognize, diagnose, and recover from errors

Rate – 0

The same answer as for heuristic #5.

10. Help and documentation

Irrelevant heuristic at this stage of the project.