MORGAN & CLAYPOOL PUBLISHERS

Usability Testing

A Practitioner's Guide to Evaluating the User Experience

Morten Hertzum

Synthesis Lectures on Human-Centered Informatics

John M. Carroll, *Series Editor*

Usability Testing: A Practitioner's Guide to Evaluating the User Experience
Morten Hertzum
www.morganclaypool.com

ISBN: 9781681737812 print
ISBN: 9781681737829 ebook
ISBN: 9781681737836 hardcover

DOI 10.2200/S00987ED1V01Y202001HCI045

A Publication in the Morgan and Claypool Publishers series
SYNTHESIS LECTURES ON HUMAN-CENTERED INFORMATICS
Lecture #45
Series Editor: John M. Carroll, Penn State University

Series ISSN 1946-7680 Print 1946-7699 Electronic

Usability Testing

A Practitioner's Guide to Evaluating the User Experience

Mortem Hertzum
University of Copenhagen, Copenhagen, Denmark

SYNTHESIS LECTURES ON HUMAN-CENTERED INFORMATICS #45

MORGAN & CLAYPOOL PUBLISHERS

ABSTRACT

It is all too common for products, such as consumer appliances, information systems, mobile apps, and websites, to cause trouble and frustration. For example, products are often difficult or dull to use, make tasks less flexible or more tedious, shift attention away from important or gratifying activities, and simply fail to deliver expected benefits or experiences. By identifying such trouble and frustration in the lab prior to widespread use, usability tests have proven a valuable method for informing redesign efforts. A usability test consists of having test users exercise a product and think aloud about their experience using it, while an evaluator observes the users and listens in on their thoughts. On this basis, the evaluator identifies usability problems and assesses the user experience. This book describes how to conduct usability tests. After providing context about concepts and testing, the main chapters of the book cover the steps involved in preparing for a usability test, executing the test sessions, and analyzing the test data. Throughout the chapters, concrete guidance is balanced against more complex issues with an impact on the robustness, validity, completeness, impact, and cost of a usability test. The book concludes with an outlook to variations of usability testing and alternatives to it.

KEYWORDS

usability testing, usability evaluation methods, usability, user experience, user testing, thinking aloud, user-centered design, human-computer interaction

Contents

CHAPTER 1

Introduction

Information technology (IT) has transformed society and continues to do so. Workplaces become increasingly distributed because IT products provide possibilities for communicating and collaborating across distance (Olson and Olson, 2014). Data is heralded as the new oil because IT products provide unprecedented possibilities for supporting decision-making by mining large quantities of data (Javornik et al., 2019). Leisure is increasingly spent indoors engaged in digital media and online games (Thulin and Vilhelmson, 2019). Cash is giving way to cards and other forms of digital payment (Arvidsson, 2019), thereby changing the way we experience money. All these changes presuppose well-functioning information systems, mobile apps, websites, and other IT products. To function well, the technical quality of these products must be good but so must their use-related quality. This book is about quality in use, that is, about usability and the user experience.

Products of low usability provide poor user experiences. These products annoy, confuse, delay, frustrate, mislead, stress, and otherwise inconvenience users. Accordingly, they may result in missed deadlines, unintended incidents, erroneous decisions, or failure to complete tasks altogether. Studies suggest that users may be wasting huge amounts of time as a result of frustrating experiences with IT products (Lazar et al., 2006). In addition, a usability problem may have had a decisive influence on the U.S. presidential election in November 2000 (Wand et al., 2001).

To exemplify what a usability problem may look like, Figure 1.1 shows the butterfly ballot used for the U.S. presidential election in Palm Beach, Florida. The butterfly ballot had candidate names on both sides and punch-holes down the middle. You cast your vote by marking the punch-hole that corresponds to your candidate. The usability problem is about establishing this correspondence. Bush voters had to match the first name on the ballot with the first punch-hole; this appears straightforward. Gore voters had to match the second name on the left-hand side of the ballot with the third punch-hole; it appears that they could easily have mismarked their ballots by instead marking the second punch-hole. If they marked the second punch-hole, they voted for Buchanan. Wand et al. (2001) estimate that over 2000 Gore voters may mistakenly have voted for Buchanan. Gore lost the election in Florida with a margin of less than 600 votes to Bush.

The butterfly ballot illustrates that low usability may influence grand-scale decisions but also that designers, occasionally, come up with low-usability solutions. Testing is needed to ensure good usability. A well-established method for this purpose is the usability test (Dumas and Fox, 2012; Lewis, 2012). Other means of usability evaluation have also been devised, including usability inspection methods, which are analytic rather than empirical (see, e.g., Cheng and Mustafa, 2015; Cockton et al., 2012; Nielsen and Mack, 1994). Usability tests and inspections yield feedback to

designers about the strengths and weaknesses of their designs. At the outset of projects, design ideas will typically be half-baked and incomplete. Later, and with the feedback from tests, flaws will be weeded out and the design refined. Petroski (1992) argues that the main driver in innovation is to improve on the flaws of existing designs, thereby assigning flaws and the process of finding them a key role in design.

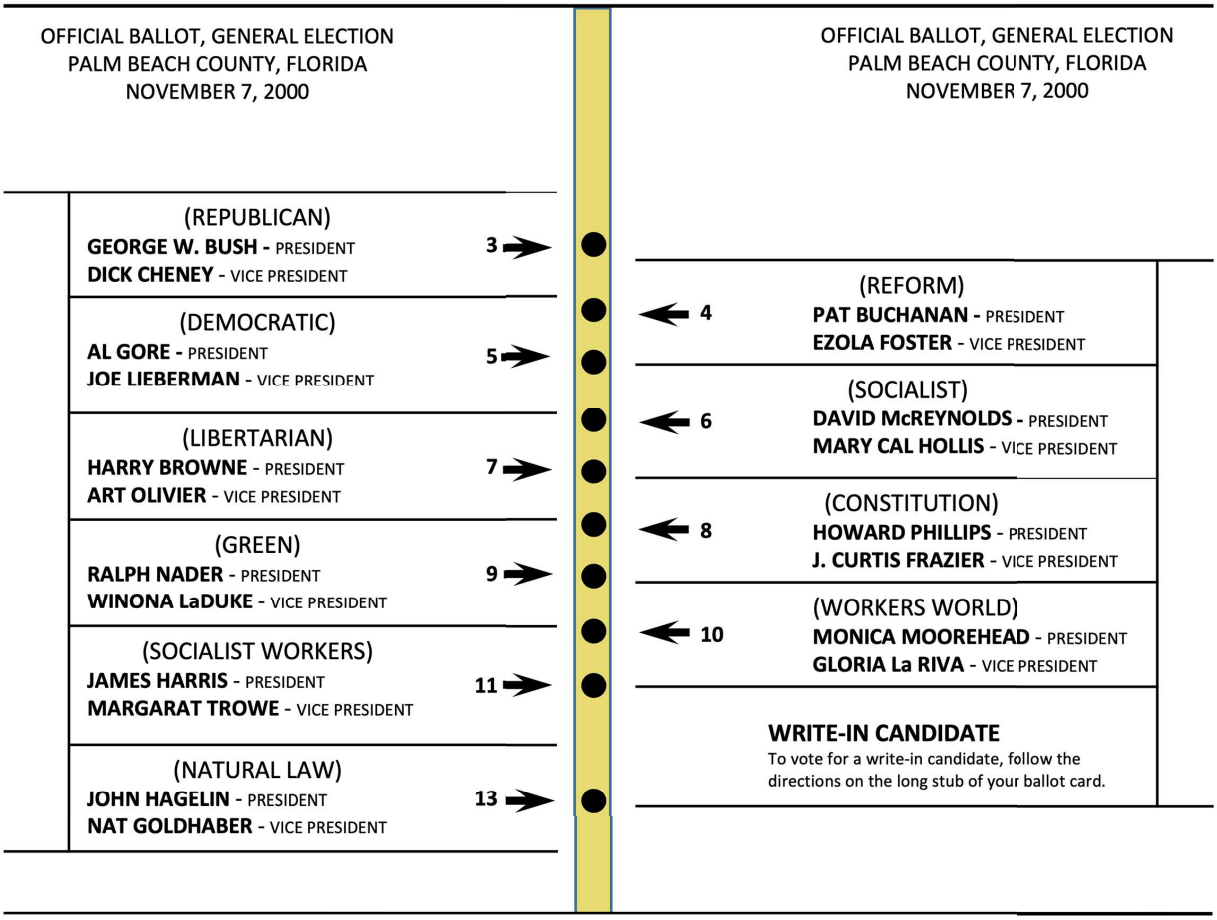

Figure 1.1: The butterfly ballot from the 2000 presidential election in Palm Beach, based on Wand et al. (2001).

1.1 THE BASIC COMPONENTS OF A USABILITY TEST

Usability testing dates back to the early 1970s (Bailey, 1972). An early and influential description of the usability test is the one by Lewis (1982), who called it the "thinking-aloud" method. In essence, a usability test consists of a user who exercises a product while thinking out loud and an evaluator who observes the user and listens in on the user's thoughts, see Figure 1.2. This basic setup allows for numerous variations. At this point, we simply note that a usability test comprises four main components (Clemmensen et al., 2009).

- *Instructions and tasks*: The users interact with the product on the basis of a set of instructions and a set of tasks prepared ahead of the test. The instructions include an explanation of how to think aloud; the tasks prescribe what the users should try to achieve with the product. Thereby, the tasks ensure that the users exercise the product in concrete detail.

- *Verbalization*: While solving the tasks, the users verbalize their thoughts—they think aloud. The verbalizations reveal how the users understand and experience the product. If the users fall silent for longer periods of time, they are prompted to resume verbalization. The users may also be asked to explain why they hesitate, what they expect, and how they assess their experience.

- *Reading the user*: The evaluator, or a group of evaluators, observes the users' interaction with the product and listens in on their thoughts. On this basis, the evaluator analyzes how well the product supports the users in accomplishing the tasks. This analysis results in the identification, description, and reporting of a set of usability problems.

- *Relationship between user and evaluator*: It is the evaluator's responsibility to establish a situation in which the user is able to exercise the product and feels free to make both positive and negative comments. Whether the user is at ease hinges on issues such as instructions, language, and indirect communication cues.

The four components are interrelated and presuppose that the evaluator is familiar with the product and its (intended) uses. These presuppositions mean that a usability test reaches into the preceding analysis and design activities as well as into the subsequent reanalysis and redesign activities. A usability test does not happen in isolation.

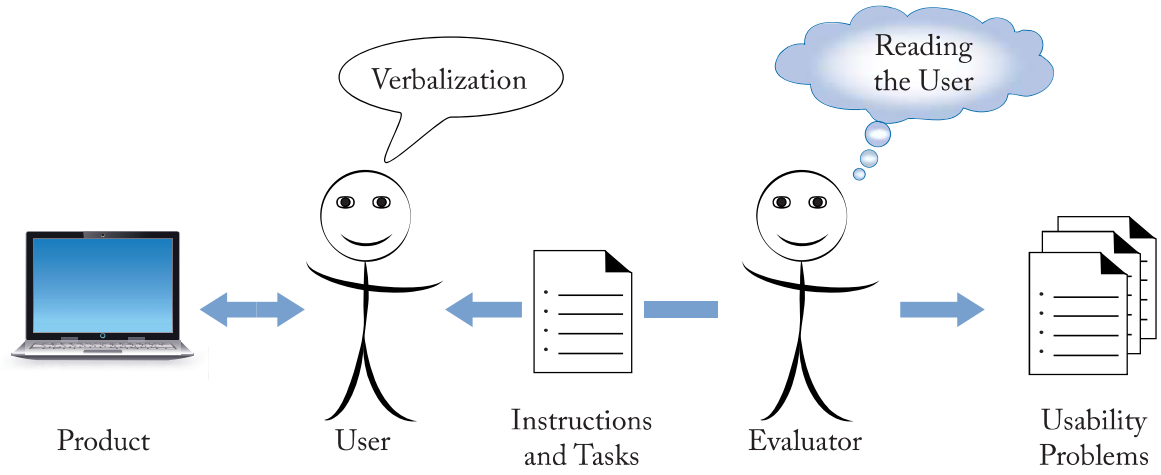

Figure 1.2: The basic components of a usability test, based on Clemmensen et al. (2009).

1.2 THE CONTEXT OF USABILITY TESTS

In the context of this book, usability testing is an activity in the process of product design. Thus, the purpose of usability testing is to inform design. It should, however, be noted that usability tests may also be conducted outside of design processes, for example to inform purchasing decisions or other choices among products that already exist in their final form.

Some models of the design process have evaluation as their pivotal activity. For example, the Wheel model (Hartson and Pyla, 2012) prescribes that the design process should cycle through the activities of analysis, design, implementation, and evaluation, see Figure 1.3. In this model, analysis is about understanding user needs; design is about creating conceptual designs and deciding interaction behavior; implementation is about prototyping; and evaluation is about checking whether the design is on track to meet user needs and requirements.

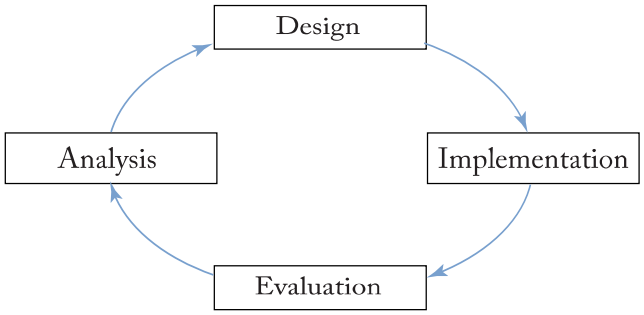

Figure 1.3: The Wheel model of the design process, based on Hartson and Pyla (2012).

The cyclic nature of the Wheel model means that analysis, design, and implementation decisions made during earlier cycles can, and should, be revisited and revised during later cycles, if an evaluation calls them into question. This way, the design process becomes agile and responsive to the insights that emerge as the process unfolds. In principle, any previous decision can be called into question by an evaluation such as a usability test. In practice, it will not be possible to reopen all decisions. The kinds of decisions that can be reopened are, however, not determined a priori by the Wheel model. Rather, it is left for the design team to determine on the basis of the particulars of the project. One of these particulars is cost, which limits—often drastically—the number of cycles that can be performed before the product must be released for use.

Other models of the design process consist of a linear sequence of activities and, thereby, restrict the possibilities for revisiting decisions made during earlier activities, see Figure 1.4. In these so-called waterfall models (Sommerville, 2016), the analysis of user needs is completed before the design begins, the design is completed before the implementation begins, and so forth. The possibilities for evaluation and iteration are, in principle, restricted to the individual phases. The model aims to prescribe that analysis decisions should not be reopened once the process has proceeded

to design, and so forth for the subsequent phases. In practice, it may prove necessary to return to a previous phase. While the exact phases differ across instances of the waterfall model, the phases depicted in Figure 1.4 are quite generic. Variants of the model tend to have more, and thus more narrowly scoped, phases rather than fewer phases. Evaluation is not brought to the fore as a separate activity but incorporated in the individual phases. With more narrowly scoped phases, within-phase evaluations become increasingly restricted.

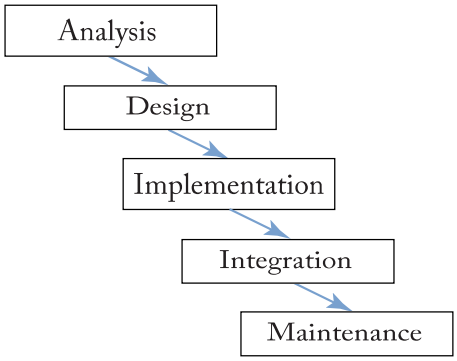

Figure 1.4: The waterfall model of the design process.

Agile methods like the Wheel model are commonplace in, for example, website development. In contrast, the linear, waterfall models are widespread in the development of safety-critical products. Neither the cyclic, nor the linear, models of the design process stipulate a specific number of usability tests. However, the cyclic models assign evaluation a more prominent position and presuppose multiple evaluations over the course of a design project. The waterfall models may not involve any usability testing, they may include some testing, or they may incorporate evaluation in all phases. Usability has become so important to product acceptance and success that usability testing is widespread in design projects (Alves et al., 2014; Gulliksen et al., 2006; Vredenburg et al., 2002), irrespective of whether the projects follow an agile, cyclic, or linear model.

1.3 A SUMMARY OF THE CHAPTERS THAT FOLLOW

The remaining chapters in this book elaborate Figure 1.2. If you are only interested in guidance on how to conduct usability tests, you can skip directly to Chapters 4–7. They are the main chapters of the book. Chapters 2 and 3 provide context about concepts and testing. Chapter 8 concludes the book by providing an outlook to variations of usability testing and alternatives to it.

We start with the concepts of *Usability and User Experience*. Chapter 2 will define what these two concepts mean. That is, it will begin to establish what you are looking for when you conduct usability tests. Multiple definitions exist of usability and user experience because these concepts are employed in diverse practical situations, are part of active research areas, and are influenced by the

continuous technological evolution. This book adopts the usability and user experience definitions endorsed by the International Organization for Standardization. To contextualize these definitions we also discuss some of the alternative conceptualizations of the two concepts.

Chapter 3, *Testing: Maxims and Modifications*, will expound what it requires for a usability test to be effective. Initially, these requirements are spelled out in terms of five maxims: robustness, validity, completeness, impact, and low cost. While a usability test ideally achieves all five maxims, they are in practice at odds with each other. As a consequence, usability testing involves a number of tradeoffs through which the maxims are modified. One of the modifications is that robustness tends to come at the cost of validity, thereby forcing a choice of either one or the other. A total of five modifications are discussed to accentuate the realities of applied usability testing. Appreciating the maxims as well as the modifications is pertinent to effective usability testing.

In Chapter 4, *Usability Testing: Step by Step*, the activities involved in conducting a usability test will be laid out. The activities are grouped into three phases: preparations, execution, and analysis. This chapter gives an overview of the three phases, which are covered in detail in the following chapters.

Chapter 5, *Preparations: Designing and Planning the Test*, is about the activities that precede the test sessions. The resulting test will depend on the specific purpose it is to serve and on the time and other resources available for conducting it. The activities in this phase consist of getting to know the domain and prototype, recruiting users, making test tasks, and setting up any equipment. Four key decisions in conducting these activities concern the fidelity of the prototype, the number of users needed, the specificity of the tasks, and the relative focus of the test on effectiveness, efficiency, and satisfaction.

Chapter 6, *Execution: Running the Test Sessions*, will cover the phase from the user arrives until the user has completed the activities involved in taking part in the test. The activities in this phase are welcoming and instructing the users, observing them and listening in on their thoughts, prompting them, taking notes, asking post-task questions, and thanking the users. Four key concerns in conducting these activities are how to make the users feel at ease, how to become sensitized to what they do and say, how and how much to prompt, and how to divide your attention among test-session moderation, on-the-fly analysis, and note-taking.

Chapter 7, *Analysis: Analyzing the Data and Reporting the Findings*, will cover how test data are turned into usability findings. The findings may include both positive and negative usability issues but your primary focus will normally be on the negative issues—the usability problems. This phase consists of analyzing the notes and other test data, rating problem severity, devising redesign proposals, and reporting the test findings. Conducting these activities involves three key concerns: What constitutes a usability problem? How many evaluators are needed? And how to ensure that the test has high impact on the continued development of the product?

Chapter 8, *Variations and Alternatives*, will conclude the book by describing ways in which usability testing can be varied. Seven variations and alternatives are covered: (1) remote usability testing, in which the user and evaluator are at different physical locations; (2) unmoderated usability tests, in which no evaluator is present during the sessions; (3) field usability testing, in which the users exercise the prototype in vivo rather than in a lab; (4) pairwise usability testing, in which thinking aloud is replaced with two users who solve the test tasks together; (5) performance testing, in which thinking aloud is performed retrospectively or not at all; (6) usability specification, in which the identification of usability problems is replaced with the assessment of whether the product meets preset usability targets; and (7) usability inspection, in which no users take part.

This book is intended for students and practitioners who need to learn, or refresh, how to conduct usability tests. The practitioners may be user-experience professionals who have usability work as their primary responsibility or designers who have usability testing as one of their responsibilities along with analysis, design, and implementation. Irrespective of background, usability testing is a nontrivial activity to perform. It is easy to do, but difficult to do well. Doing it well requires a reflective approach that recognizes the complexities, yet stays systematic. This book strives to provide such a reflective approach by balancing concrete, easy-to-follow guidance against more complex, important-to-consider issues. To achieve this balance, the book draws on more than three decades of research on usability evaluation.