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.
Usability Testing
A Practitioner’s Guide to Evaluating the User Experience
Synthesis Lectures on Human-Centered Informatics

Editor

John M. Carroll, Penn State University

Human-Centered Informatics (HCI) is the intersection of the cultural, the social, the cognitive, and the aesthetic with computing and information technology. It encompasses a huge range of issues, theories, technologies, designs, tools, environments, and human experiences in knowledge work, recreation and leisure activity, teaching and learning, and the potpourri of everyday life. The series publishes state-of-the-art syntheses, case studies, and tutorials in key areas. It shares the focus of leading international conferences in HCI.

Usability Testing: A Practitioner’s Guide to Evaluating the User Experience
Morten Hertzum

Geographical Design: Spatial Cognition and Geographical Information Science, Second Edition
Stephen C. Hirtle

Human-Computer Interactions in Museums
Eva Hornecker and Luigina Ciolfi

Encounters with HCI Pioneers: A Personal History and Photo Journal
Ben Shneiderman

Social Media and Civic Engagement: History, Theory, and Practice
Scott P. Robertson

The Art of Interaction: What HCI Can Learn from Interactive Art
Ernest Edmonds

Representation, Inclusion, and Innovation: Multidisciplinary Explorations
Clayton Lewis

Research in the Wild
Yvonne Rogers and Paul Marshall

Designing for Gesture and Tangible Interaction
Mary Lou Maher and Lina Lee
From Tool to Partner: The Evolution of Human-Computer Interaction
Jonathan Grudin

Qualitative HCI Research: Going behind the Scenes
Ann Blandford, Dominic Furniss, and Stephann Makri

Learner-Centred Design of Computing Education: Research on Computing for Everyone
Mark Guzdial

The Envisionment and Discovery Collaboratory (EDC): Explorations in Human-Centred Informatics with Tabletop Computing Environments
Ernesto G. Arias, Hal Eden, and Gerhard Fischer

Humanistic HCI
Jeffrey Bardzell and Shaowen Bardzell

The Paradigm Shift to Multimodality in Contemporary Computer Interfaces
Sharon Oviatt and Philip R. Cohen

Multitasking in the Digital Age
Gloria Mark

The Design of Implicit Interactions
Wendy Ju

Core-Task Design: A Practice-Theory Approach to Human Factors
Leena Norros, Paula Savioja, and Hanna Koskinen

An Anthropology of Services: Toward a Practice Approach to Designing Services
Jeanette Blomberg and Chuck Darrah

Proxemic Interactions: From Theory to Practice
Nicolai Marquardt and Saul Greenberg

Contextual Design: Evolved
Karen Holtzblatt and Hugh Beyer

Constructing Knowledge Art: An Experiential Perspective on Crafting Participatory Representations
Al Selvin and Simon Buckingham Shum

Spaces of Interaction, Places for Experience
David Benyon
Mobile Interactions in Context: A Designerly Way Toward Digital Ecology
Jesper Kjeldskov

Working Together Apart: Collaboration over the Internet
Judith S. Olson and Gary M. Olson

Surface Computing and Collaborative Analysis Work
Judith Brown, Jeff Wilson, Stevenson Gossage, Chris Hack, and Robert Biddle

How We Cope with Digital Technology
Phil Turner

Translating Euclid: Designing a Human-Centred Mathematics
Gerry Stahl

Adaptive Interaction: A Utility Maximisation Approach to Understanding Human Interaction with Technology
Stephen J. Payne and Andrew Howes

Making Claims: Knowledge Design, Capture, and Sharing in HCI
D. Scott McCrickard

HCI Theory: Classical, Modern, and Contemporary
Yvonne Rogers

Activity Theory in HCI: Fundamentals and Reflections
Victor Kaptelinin and Bonnie Nardi

Conceptual Models: Core to Good Design
Jeff Johnson and Austin Henderson

Geographical Design: Spatial Cognition and Geographical Information Science
Stephen C. Hirtle

User-Centred Agile Methods
Hugh Beyer

Experience-Centred Design: Designers, Users, and Communities in Dialogue
Peter Wright and John McCarthy

Experience Design: Technology for All the Right Reasons
Marc Hassenzahl

Designing and Evaluating Usable Technology in Industrial Research: Three Case Studies
Clare-Marie Karat and John Karat
Interacting with Information
Ann Blandford and Simon Attfield

Designing for User Engagement: Aesthetic and Attractive User Interfaces
Alistair Sutcliffe

Context-Aware Mobile Computing: Affordances of Space, Social Awareness, and Social Influence
Geri Gay

Studies of Work and the Workplace in HCI: Concepts and Techniques
Graham Button and Wes Sharrock

Semiotic Engineering Methods for Scientific Research in HCI
Clarisse Sieckenius de Souza and Carla Faria Leitão

Common Ground in Electronically Mediated Conversation
Andrew Monk
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
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
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.2  How Many Evaluators Are Needed?</td>
<td>67</td>
</tr>
<tr>
<td>7.3  Rate Problem Severity</td>
<td>69</td>
</tr>
<tr>
<td>7.4  Devise Redesign Proposals</td>
<td>72</td>
</tr>
<tr>
<td>7.5  Report Test Findings</td>
<td>73</td>
</tr>
<tr>
<td>8. Variations and Alternatives</td>
<td>77</td>
</tr>
<tr>
<td>8.1  Remote Usability Tests</td>
<td>77</td>
</tr>
<tr>
<td>8.2  Unmoderated Usability Tests</td>
<td>78</td>
</tr>
<tr>
<td>8.3  Field Usability Tests</td>
<td>79</td>
</tr>
<tr>
<td>8.4  Pairwise Usability Tests</td>
<td>80</td>
</tr>
<tr>
<td>8.5  Performance Testing</td>
<td>81</td>
</tr>
<tr>
<td>8.6  Usability Specification</td>
<td>82</td>
</tr>
<tr>
<td>8.7  Usability Inspection</td>
<td>84</td>
</tr>
<tr>
<td>References</td>
<td>87</td>
</tr>
<tr>
<td>Author Biography</td>
<td>103</td>
</tr>
</tbody>
</table>
I became interested in usability evaluation methods in the late 1990s when usability and user experience were still in the process of establishing themselves as central concerns in product development. This book has grown out of my research on this topic over the many years that have passed since then. Most of this research has been conducted in collaboration with colleagues who have contributed invaluably, and in various ways, to my thinking about usability testing. Without their contributions, I would not have been in a position to write this book. I am particularly indebted to Pia Borlund, Torkil Clemmensen, Erik Frøkjær, Kristin Due Holmegaard, Kasper Hornbæk, Niels Ebbe Jacobsen, Bonnie John, Kristina Bonde Kristoffersen, Jyoti Kumar, Rolf Molich, Qingxin Shi, Xianghong Sun, Hans Sønderstrup-Andersen, and Pradeep Yammiyavar. In addition, I wish to extend my thanks to the many usability professionals who, in spite of their busy schedules, have been prepared to take part in our empirical studies as evaluators.
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.

<table>
<thead>
<tr>
<th>OFFICIAL BALLOT, GENERAL ELECTION PALM BEACH COUNTY, FLORIDA NOVEMBER 7, 2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>(REPUBLICAN)</td>
</tr>
<tr>
<td>GEORGE W. BUSH - PRESIDENT</td>
</tr>
<tr>
<td>DICK CHENEY - VICE PRESIDENT</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>(DEMOCRATIC)</td>
</tr>
<tr>
<td>AL GORE - PRESIDENT</td>
</tr>
<tr>
<td>JOE LIEBERMAN - VICE PRESIDENT</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>(LIBERTARIAN)</td>
</tr>
<tr>
<td>HARRY BROWNE - PRESIDENT</td>
</tr>
<tr>
<td>ART OLIVIER - VICE PRESIDENT</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>(GREEN)</td>
</tr>
<tr>
<td>RALPH NADER - PRESIDENT</td>
</tr>
<tr>
<td>WINONA LaDUKE - VICE PRESIDENT</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>(SOCIALIST WORKERS)</td>
</tr>
<tr>
<td>JAMES HARRIS - PRESIDENT</td>
</tr>
<tr>
<td>MARGARAT TROWE - VICE PRESIDENT</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>(NATURAL LAW)</td>
</tr>
<tr>
<td>JOHN HAGELIN - PRESIDENT</td>
</tr>
<tr>
<td>NAT GOLDHABER - VICE PRESIDENT</td>
</tr>
<tr>
<td>13</td>
</tr>
</tbody>
</table>

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.

![Diagram](image.png)

**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).](image)

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.](image)

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.
Usability tests are a means to evaluate whether a product is usable and how the users experience its use. But what are usability and user experience? In this chapter, we define these two concepts. The definitions are adopted from the International Organization for Standardization and have, thus, gone through a process of proposal, comments, and refinement among experts in the field. In spite of this consensus process, alternative conceptualizations abound. To contextualize the definitions of usability and user experience, we discuss some of these alternative conceptualizations.

2.1 DEFINITIONS

According to Shackel (1984), the first recorded use of the term usability dates back to 1842. Today, the concept has entered everyday language and is ubiquitous in human-computer interaction. The concept of user experience is newer. It emerged in response to a recognition that affect, meaning, value, and the like are important to users’ interactions with products but possibly under-recognized by usability, which was seen as focusing mainly on cognition and performance (Hassenzahl and Tractinsky, 2006; Law et al., 2009). Today, the concepts of usability and user experience are often used in tandem. The two concepts overlap but emphasize different aspects of users’ interactions with a product. The International Organization for Standardization provides widely used definitions of both concepts (ISO 9241, 2010, p. 3).

- **Usability**: “extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.”

- **User experience**: “person’s perceptions and responses resulting from the use and/or anticipated use of a product, system or service.”

The definition of usability contains the four elements commonly used to describe a use situation, see Figure 2.1. By insisting that usability is only defined when these four elements are specified, the definition makes usability an attribute of the use situation, not of the product. Whenever we describe a product as usable, it is a shorthand for saying that a use situation (i.e., a specific configuration of product, users, tasks, and context) works well. It is worth noting the match between the basic components of a usability test (Figure 1.2) and the four elements of the use situation (Figure 2.1). In a usability test, specified users exercise a specified product by solving specified tasks, that is, they work toward specified goals. The one element that may be sacrificed is the context of
use, which is largely absent when the usability test is conducted in a lab-like setting away from the users’ real work.

![Figure 2.1: The use situation.](image)

The usability definition also stipulates what a usability test must establish. It must establish the extent to which the use situation possesses three qualities (ISO 9241, 2010, pp. 2–3): effectiveness (defined as the “accuracy and completeness with which users achieve specified goals”), efficiency (“resources expended in relation to the accuracy and completeness with which users achieve goals”), and satisfaction (“freedom from discomfort and positive attitudes towards the use of the product”).

The definition of user experience makes it a subjective perception. This experience includes “all the users’ emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviors and accomplishments that occur before, during and after use” (ISO 9241, 2010, p. 3). Notably, the definition mentions neither tasks/goals, nor context, except in the indirect manner of defining user experience as resulting from use, which presupposes a use situation. The primary focus of the definition is on the user and product. In addition, the definition extends use to include also anticipated use. By explicitly including anticipated use, user experience is also about expectations and the relations between (pre-use) expectations and (in-use) experiences. From an evaluation perspective, the inclusion of anticipated use recognizes the value of testing early on the basis of, for example, mockups and sketches.

### 2.2 OTHER VIEWS ON USABILITY AND USER EXPERIENCE

There is no shortage of alternative takes on usability and user experiences. Tractinsky (2018) argues that usability is an umbrella concept and, thus, vague and loose. Umbrella concepts are prevalent in diverse and context-sensitive domains that lack a unifying research paradigm. This appears an apt characterization of human-computer interaction (HCI), the research domain that encompasses usability testing. In the absence of a unifying paradigm, concepts tend to denote broad ideas and lack agreed-upon definition. By describing usability as an umbrella concept in spite of the ISO
9241 definition, Tractinsky (2018) emphasizes the presence of competing conceptualizations. He considers this state of affairs inferior to a definitive usability concept, which would define the exact attributes of the concept and provide validated instruments for its measurement. However, a definitive concept also entails the risk of becoming a formal quest for conceptual precision at the expense of practical relevance (Hertzum, 2018).

As an alternative to definitive concepts, Blumer (1954, p. 7) proposes sensitizing concepts, which “suggest directions along which to look.” Thus, a genuine understanding of the usability of a product requires a capacity for approaching usability from multiple points of view so that one becomes sensitized to the various elements that impact the use of the product. To this end, Hertzum (2010) presents six images of usability, see Figure 2.2. In spite of a shared essence, these images differ in mindset and perspective. Each image suggests a different direction along which to look and provides only a partial view of usability. It is only by bringing multiple images to bear that you become sensitized to the rich variety of elements that enters into the user experience. The six images of usability are as follows (Hertzum, 2010).

*Universal usability:* This usability image embraces the challenge of making products for everybody to use. People’s abilities, backgrounds, personal styles, technological environments, and so forth are diverse. Yet, all people may need, or want, access to information or some of the other opportunities provided by websites and other products (Stephanidis et al., 2012). Consistently excluding certain groups of people from these opportunities is incompatible with general notions of a fair society. Excluding sizable groups of people from the use of individual products may diminish the feasibility of developing these products. Therefore, it is a compelling goal to design products that are usable for all—a goal quite different from the “specified users” of the ISO 9241 (2010) definition. Universal usability is a grand challenge because it involves that products must be as inclusive as humans are diverse. Universal usability is particularly relevant in relation to the testing of general-purpose products, walk-up-and-use systems, and a variety of web applications, such as e-commerce, e-government, and e-health.

*Situational usability:* This image corresponds to the ISO 9241 (2010) definition of usability. It is based on the premise that users do not experience products in isolation but as part of a use situation. Consequently, usability must be understood in relation to specified situations with their users, tasks, and wider context of use (Figure 2.1). Simply put, the particulars of the use situation are imperative to whether a product is usable. This situatedness outweighs general usability principles. Draper (1993) considers the situational image of usability pessimistic because it rejects generalization beyond specified use situations. While this largely implies that universal usability is
an unattainable goal, it is consistent with basic HCI principles such as “know thy user” (Kim, 2015). Situational usability is directly applicable to the testing of products that are developed with specified use situations in mind, rather than for a market of diverse customers and use situations. Such products include bespoke products, which are commissioned by a customer and custom-built, or configured, for this particular customer’s situation.

**Figure 2.2:** The six images of usability, based on Hertzum (2010).

*Perceived usability:* According to this image, usability concerns the user’s subjective experience of a product. Perceived usability is truly user-centered, as opposed to use-centered, because it makes the individual user the final arbiter on usability. This image carries no particular focus on satisfaction but merely a focus on subjective assessments, as opposed to performance measures. People’s intention to use a product
depends, to a considerable extent, on their perceptions of the usefulness, ease of use, and enjoyment associated with using the product (Hornbæk and Hertzum, 2017). These three perceptions correspond roughly to effectiveness, efficiency, and satisfaction. They demonstrate how perceived usability extends beyond satisfaction and, for example, influences users’ intention to use a product, their ways of interacting with a product, and their future purchasing decisions (e.g., Han et al., 2001; Hornbæk and Hertzum, 2017). Perceived usability is especially relevant when evaluating products the use of which is discretionary.

*Hedonic usability*: This image holds that usability is about joy of use rather than ease of use. Hedonic usability is related to the satisfaction component in the ISO 9241 (2010) definition of usability, but while this component appears biased toward avoiding negative emotions, hedonic usability is all about producing positive emotions. This distinction is important because the qualities that help avoid negative emotions are different from those that produce positive emotions (Helander and Khalid, 2006). Hedonic usability is similar to perceived usability in its focus on subjective experience, but it is distinct from the other images in its exclusive focus on pleasure and emotion. Hedonic usability is particularly relevant to the evaluation of consumer appliances, online games, social media, and other products that involve having a good experience or expressing oneself. In addition, hedonic usability is relevant to e-commerce because the presence of hedonic qualities impacts buying decisions (Jordan, 1998).

*Organizational usability*: According to this image, usability implies groups of people collaborating in an organizational setting. None of the four previous images have mentioned collaboration or organizations, even though products such as information systems abound in organizations. While a product is expected to provide collective benefit to an organization, there may be individual users who do not benefit. Frequently, some users are tasked with additional work to enter or process information; other users reap the benefits that accrue from this additional work (Grudin, 1994). The uneven distribution of work and benefits means that different user groups may perceive the product quite differently. Elliott and Kling (1997) propose that organizational usability should be assessed at three levels: the user-product match, the organization-product match, and the environment-product match. The second and third levels, in particular, recognize that organizational usability is affected by the ways in which products make some competences obsolete, reroute information, create new roles, and so forth. This image of usability is relevant to the evaluation of products that range from groupware used by organizational subgroups at their own discretion to corporate systems used by all employees on a daily basis.
2. USABILITY AND USER EXPERIENCE

Cultural usability: This image emphasizes that usability takes on different meanings for users with different cultural backgrounds. Cultural usability can be defined as “the extent to which a computer system, especially in intercultural contexts of use, matches the cultural background of its users, such that it supports their activities effectively, efficiently, and pleasurably” (Hertzum, 2010, p. 584). Many product elements may be culture-dependent, including colors, graphics, language, and layout (Callahan, 2005; Marcus and Gould, 2012). For example, the color red is associated with danger in the U.S. but with happiness in China; in Egypt the color associated with happiness is yellow, which in the U.S. signals cowardice (Thorell and Smith, 1990). In addition to such interface-level differences, people’s cultural backgrounds influence their cognitive processes—the way they know the world (Nisbett et al., 2001). These differences in cognitive processes create cross-cultural differences in what constitutes a usable design. Cultural usability is particularly relevant to the evaluation of products for international audiences, including web applications. International audiences may be internal to a single country or organization.

The six images of usability serve to elaborate and contextualize any one definition of usability. Being aware of all six images makes it easier to appreciate the focus and limitations of each of them. While a single usability image will normally dominate in any given test, you should not focus exclusive on one image throughout a development project. Rather, you are advised to enrich your understanding of the usability of the product by, occasionally, applying an alternative usability image to challenge the dominant image. The rationale for this advice is that the usability images are interrelated, but different. They sensitize you to different aspects of the user experience. In choosing an image for a usability test, you should consider the product to be tested and the objective of the test.
Effective usability testing is dependent upon a good test method, including good fits between the test and the use context and between the test and the encompassing development process. In this chapter the requirements for effective usability testing are spelled out in terms of five maxims. However, these maxims are at odds with each other. Consequently, applied usability testing involves a number of tradeoffs through which the maxims are modified. Five such modifications are discussed. Appreciating the maxims as well as the modifications is pertinent to being able to conduct effective usability tests.

3.1 FIVE MAXIMS

Ideally, a usability test is robust, valid, complete, high-impact, and low-cost, see Figure 3.1. The three first of these maxims are akin to the general methodological requirements of reliability, validity, and generalizability (Shadish et al., 2002). The two last maxims are about incorporating the usability test in the product development process. Failure to achieve any of the five maxims threatens the quality and usefulness of the test. This is apparent when the maxims are described in more detail (see also, Hertzum, 1999).

![Figure 3.1: Maxims of usability testing.](image-url)
Robustness: A test method is robust when it produces fairly stable results across a range of variations in the test situation. Such variations may, for example, include the group of user representatives participating in the test as test users, the order in which the test tasks are presented to the users, the evaluator conducting the test, and whether the users think aloud concurrently, retrospectively or not at all. Unless the test method is robust, it cannot be assumed that a rerun of a test will yield essentially the same results. If the results vary considerably from one test of a product to another test of the same product, then the test is not convincing and the results may mislead (Lewis, 2001b). The more robust a usability test is, the more useful its results are.

Validity: A usability test differs from real use situations in that some aspects of the test are not part of real use, just as some aspects of real use are not part of the test. For example, users rarely think aloud during real use and the outcome of tasks rarely has consequences for users during usability tests. The closer the test situation is to real use, the more ecological the test. Ecological gaps between the test situation and real use threaten the validity of the test (Thomas and Kellogg, 1989). A test that lacks validity suffers from two shortcomings. First, the problems that occur during the test may not exist during real use. Second, the problems that hamper real use may not surface during the test. As a result, usability tests are misleading unless their results are valid.

Completeness: A test method that robustly reveals valid usability problems may still reveal only part of the full set of problems. For example, the test tasks may not cover the full functionality of the product or the test users may not represent the full diversity of the user population. An incomplete usability test merely gives a partial picture of the usability of the tested product. With a partial picture it is difficult to assess product quality and prioritize improvement efforts. Conversely, a usability test that is complete, at least with respect to major usability problems, provides direction and a firm footing.

Impact: The impact of a usability test is its ability to bring about effective changes in the tested product. That is, impact is about whether the development team is persuaded to make fixes in response to the identified usability problems and whether these fixes are effective at removing the usability problems (John and Marks, 1997; Law, 2006). The most direct impact of a usability test is in relation to the development team, which may—or may not—be persuaded that the reported usability problems warrant product revisions. However, a usability test may also have an impact on management, marketing, or service by persuading them to reallocate resources, modify selling points, or expand product documentation. Whenever a validly identified
problem is left unaddressed, an opportunity to improve the tested product is missed. That is, the effort that went into finding the problem is wasted.

**Cost:** Usability testing entails expenses for equipment, evaluator competences, compensation to users, and hours for running and analyzing test sessions. In addition to the cost of finding the usability problems, the total cost of a test also includes the cost of addressing the problems found. Several studies aim to justify the cost of usability tests by converting the estimated benefits of performing the tests into cost savings (e.g., Bias and Mayhew, 2005). However, the subjective experience of many developers is that usability work adds expenses, lengthens projects, and fails to prevent that new problems show up when the products are released. As a result, practitioners tend to show a strong preference for low-cost usability tests.

### 3.2 MODIFICATIONS IN PRACTICE

While the five maxims may sound indispensable, usability tests rarely achieve them in practice—for multiple, good reasons. The ideal depicted in Figure 3.1 masks that practitioners face conditions and tradeoffs that modify the maxims. Appreciating these conditions and tradeoffs is important to understanding usability testing. Specifically, the modifications of the maxims influence what you can conclude from a usability test. In interpreting test results, you should heed the following modifications.

First, **robustness tends to come at the cost of validity.** For example, in-the-lab usability tests provide a controlled environment in which various sources of variability can be kept to a minimum. The controlled environment yields robustness but at the expense of reducing naturally occurring variability, thereby lowering validity. Conversely, in-the-field usability tests are sensitive to the dynamic particulars of real use situations. While this ecological sensitivity bolsters validity, it also implies a lack of control that threatens robustness. That is, a rerun of the test will likely encounter different particulars and, therefore, yield different results. This tradeoff between robustness and validity is not specific to usability tests but common to empirical methods. It has led McGrath (1981, p. 179) to conclude that “all research strategies and methods are seriously flawed; often with their very strengths in regard to one desideratum functioning as serious weaknesses in regard to other, equally important, goals.”

Second, **validity is hard to assess.** As a result, it often receives scant attention. For robustness, we can run the test twice and compare the results; the overlap (or lack thereof) is a direct indication of the robustness of the test. For validity, we usually cannot know until much later (e.g., by comparing test results with hotline calls) whether the problems reported from a usability test actually confuse users, slow them down, or otherwise degrade their user experience. The research on the