educlash
Just Another Way To Learn

UX Evaluation

INTRODUCTION

Textbook Chapter 12. UX EVALUATIONINTRODUCTION

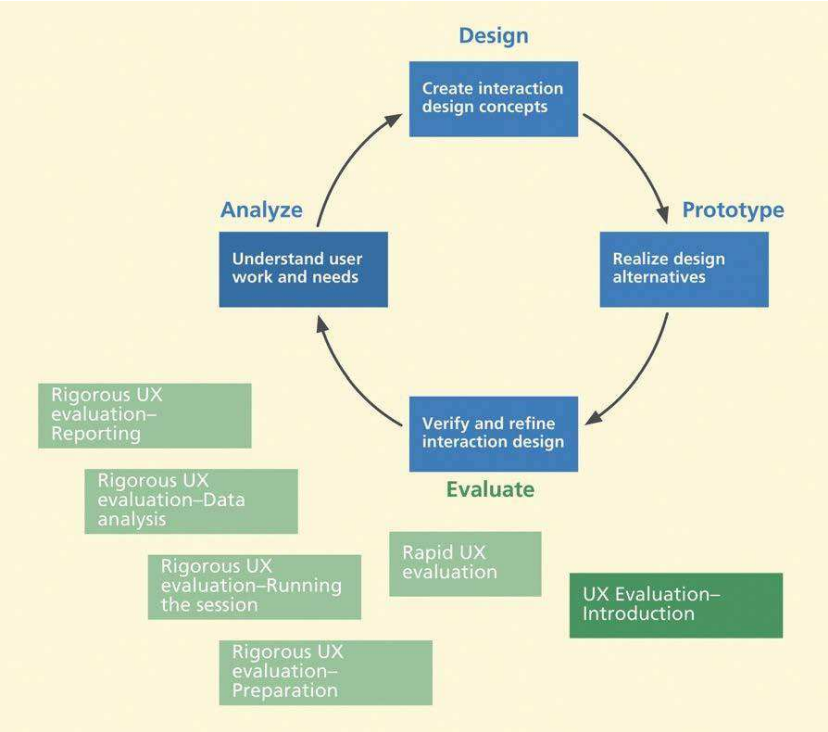

Figure 12-1
You are here, at the evaluation activity in the context of the overall Wheel lifecycle template.

INTRODUCTION

- Evaluate with a Prototype on Your Own Terms
- Measurability of User Experience
  - But can you evaluate usability or user experience? This may come as a surprise, but neither usability nor user experience is directly measurable.
  - we resort to measuring things we can measure and use those measurements as indicators of our more abstract and less measurable notions.
- User Testing? No!
  - UX evaluation must be an ego-free process; you are improving designs, not judging users, designers, or developers.

FORMATIVE VS. SUMMATIVE EVALUATION

- When the cook tastes the soup, that's formative; when the guests taste the soup, that's summative" (Stake, 2004, p. 17).
  - *Formative evaluation* is primarily diagnostic; it is about collecting qualitative data to identify and fix UX problems and their causes in the design.
  - *Summative evaluation* is about collecting quantitative data for assessing a level of quality due to a design, especially for assessing improvement in the user experience due to formative evaluation.

FORMATIVE VS. SUMMATIVE EVALUATION

- Qualitative Data
  - Qualitative data are nonnumeric and descriptive data, usually describing a UX problem or issue observed or experienced during usage.

- Quantitative Data
  - Quantitative data are numeric data, such as user performance metrics or opinion ratings.

FORMATIVE VS. SUMMATIVE EVALUATION

- Formal summative evaluation
  - is typified by an empirical competitive benchmark study based on formal, rigorous experimental design aimed at comparing design hypothesis factors. Formal summative evaluation is a kind of controlled hypothesis testing with an m by n factorial design with y independent variables, the results of which are subjected to statistical tests for significance. Formal summative evaluation is an important HCI skill, but we do not cover it in this book.
- Informal summative evaluation
  - is used, as a partner of formative evaluation, for quantitatively summing up or assessing UX levels using metrics for user performance (such as the time on task), for example, as indicators of progress in UX improvement, usually in comparison with pre-established UX target levels

FORMATIVE VS. SUMMATIVE EVALUATION

- Engineering Evaluation of UX: Formative Plus Informal Summative

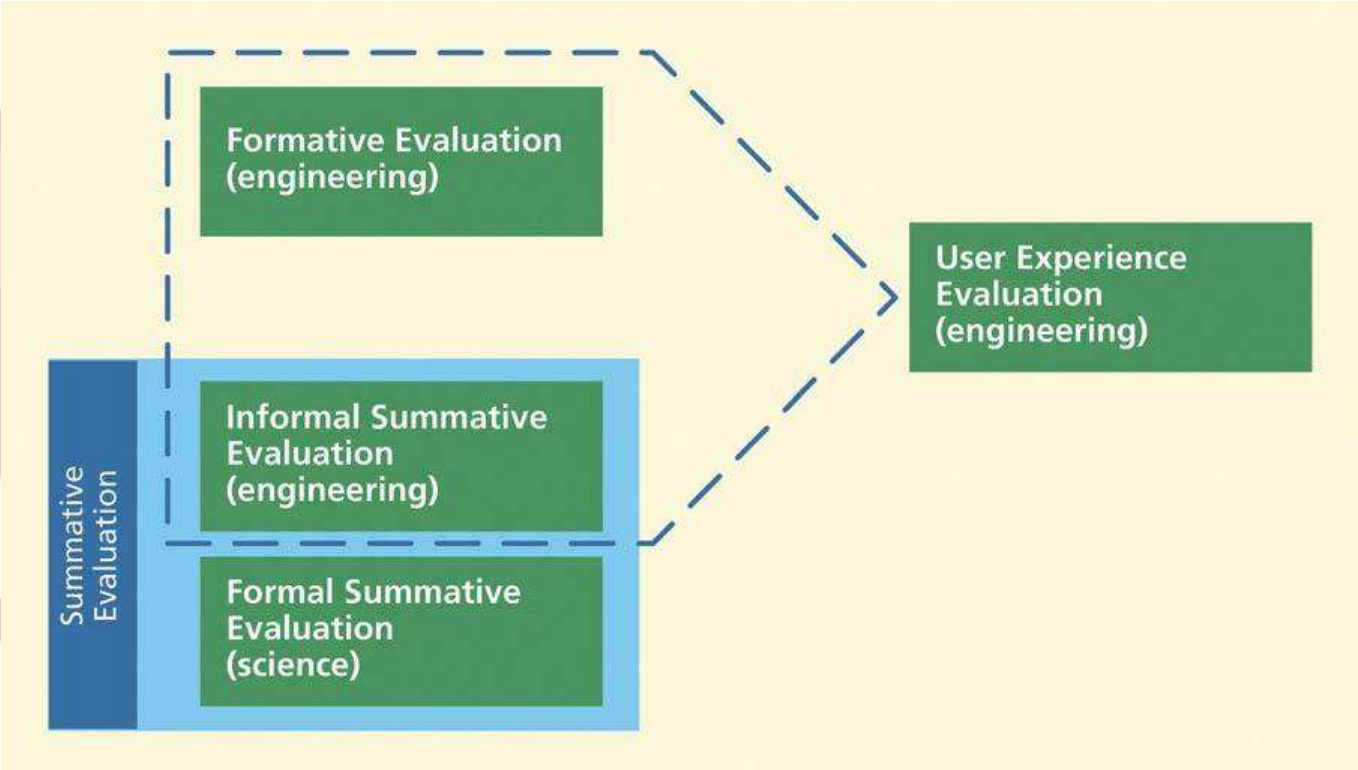

Figure 12-2
UX evaluation is a combination of formative and informal summative evaluation.

TYPES OF FORMATIVE AND INFORMAL SUMMATIVE EVALUATION METHODS

- Dimensions for Classifying Formative UX Evaluation Methods
  - empirical method vs. analytic method
  - rigorous method vs. rapid method
- Rigorous Method vs. Rapid Method
  - Choose a rigorous empirical method such as lab-based testing when you need effectiveness and thoroughness, but expect it to be more expensive and time-consuming.
  - Choose the lab-based method to assess quantitative UX measures and metrics, such as time-on-task and error rates, as indications of how well the user does in a performance-oriented context.
  - Choose lab-based testing if you need a controlled environment to limit distractions.
  - Choose empirical testing in the field if you need more realistic usage conditions for ecological validity than you can establish in a lab.

TYPES OF FORMATIVE AND INFORMAL SUMMATIVE EVALUATION METHODS

- UX evaluation methods can be faster and less expensive.
  - Choose a rapid evaluation method for speed and cost savings, but expect it to be (possibly acceptably) less effective.
  - Choose a rapid UX evaluation method for early stages of progress, when things are changing a lot, anyway, and investing in detailed evaluation is not warranted.
  - Choose a rapid method, such as a design walkthrough, an informal demonstration of design concepts, as a platform for getting initial reactions and early feedback from the rest of the design team, customers, and potential users.

TYPES OF FORMATIVE AND INFORMAL SUMMATIVE EVALUATION METHODS

- Where the Dimensions Intersect

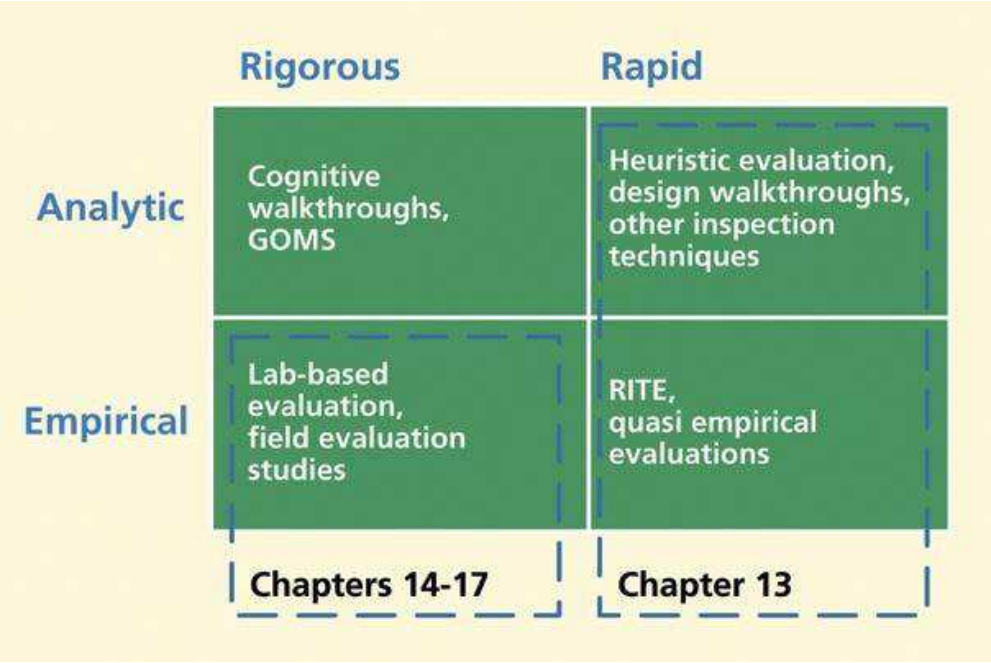

Figure 12-3
Sample UX evaluation methods at intersections between the dimensions of UX evaluation method types.

TYPES OF EVALUATION DATA

- Objective Data vs. Subjective Data
  - Objective UX data are data observed directly by either the evaluator or the participant. Subjective UX data represent opinions, judgments, and other subjective feedback usually from the user, concerning the user experience and satisfaction with the interaction design.
- Quantitative Data vs. Qualitative Data
  - Quantitative data are numeric data, such as data obtained by user performance metrics or opinion ratings. Quantitative data are the basis of an informal summative evaluation component and help the team assess UX achievements and monitor convergence toward UX targets, usually in comparison with the specified levels set in the UX targets

SOME DATA COLLECTION TECHNIQUES

- Critical Incident Identification
  - Critical incidents
    - the user's general activity or task
    - objects or artifacts involved
    - the specific user intention and action that led immediately to the critical incident
    - expectations of the user about what the system was supposed to do when the critical incident occurred
    - what happened instead
    - as much as possible about the mental and emotional state of the user
    - indication of whether the user could recover from the critical incident and, if so, a description of how the user did so
    - additional comments or suggested solutions to the problem

SOME DATA COLLECTION TECHNIQUES

- Relevance of critical incident data
- History of critical incident data
- Mostly used as a variation
- Critical incident reporting tools
- Who identifies critical incidents?
- Timing of critical incident data capture: The evaluator's awareness zone

SOME DATA COLLECTION TECHNIQUES

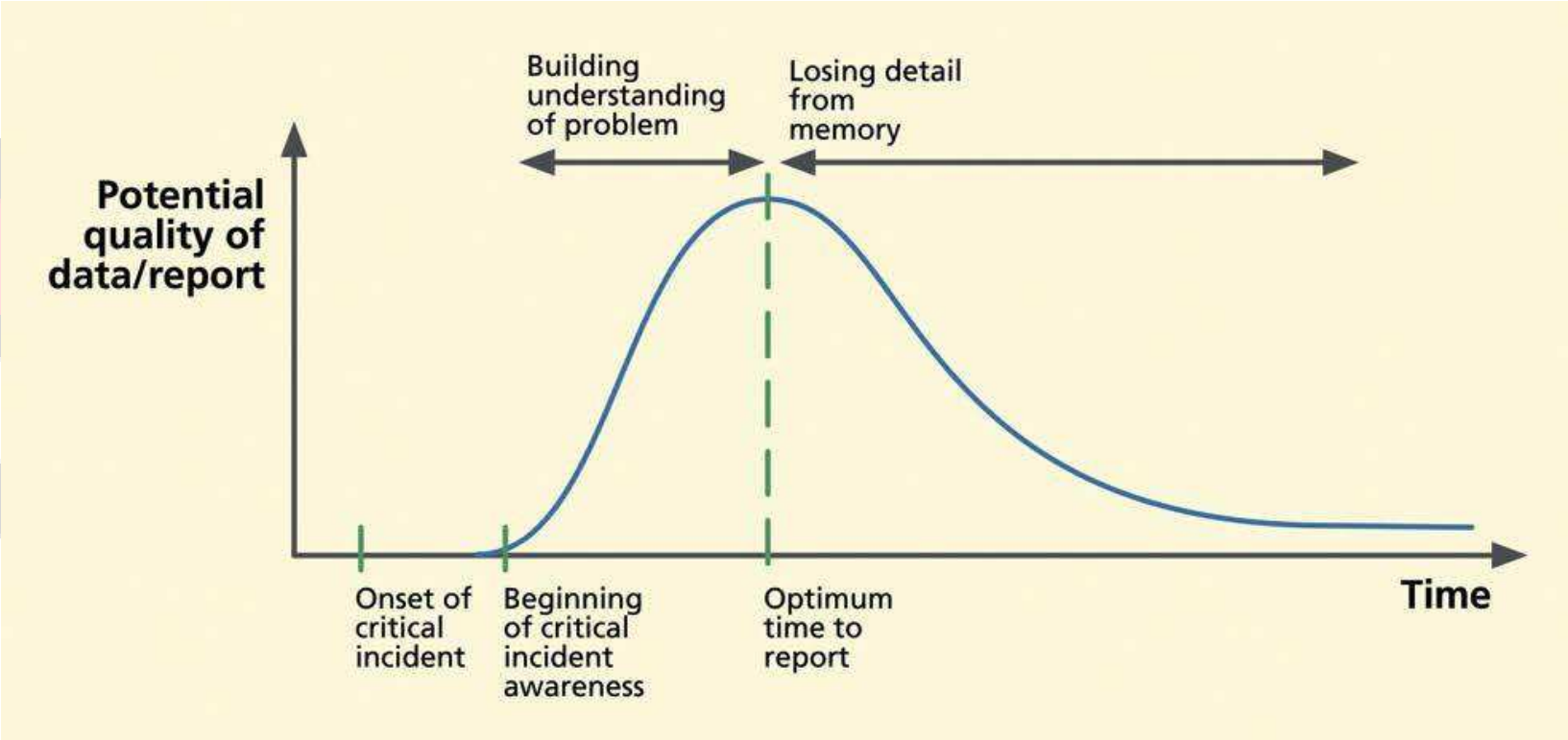

Figure 12-4

SOME DATA COLLECTION TECHNIQUES

Critical incident description detail vs. time after critical incident.

SOME DATA COLLECTION TECHNIQUES

- The Think-Aloud Technique
  - Why use the think-aloud technique?
  - What kind of participant works best?
  - How to manage the think-aloud protocol?
  - Retrospective think-aloud techniques
  - Co-discovery think-aloud techniques
  - Does thinking aloud affect quantitative task performance metrics in lab-based evaluation?

SOME DATA COLLECTION TECHNIQUES

- Questionnaires
  - Semantic differential scales
  - The Questionnaire for User Interface Satisfaction (QUIS)
  - The System Usability Scale (SUS)
  - The Usefulness, Satisfaction, and Ease of Use (USE) Questionnaire

SOME DATA COLLECTION TECHNIQUES

User Evaluation of Interactive Computer Systems

For each question, please circle the number that most appropriately reflects your impressions about this topic with respect to using this computer system or product.

1. Terminology relates to task domain	[distantly]	0 1 2 3 4 5 6 7 8 9 10	[closely]	NA
2. Instructions describing tasks	[confusing]	0 1 2 3 4 5 6 7 8 9 10	[clear]	NA
3. Instructions are consistent	[never]	0 1 2 3 4 5 6 7 8 9 10	[always]	NA
4. Operations relate to tasks	[distantly]	0 1 2 3 4 5 6 7 8 9 10	[closely]	NA
5. Informative feedback	[never]	0 1 2 3 4 5 6 7 8 9 10	[always]	NA
6. Display layouts simplify tasks	[never]	0 1 2 3 4 5 6 7 8 9 10	[always]	NA
7. Sequence of displays	[confusing]	0 1 2 3 4 5 6 7 8 9 10	[clear]	NA
8. Error messages are helpful	[never]	0 1 2 3 4 5 6 7 8 9 10	[always]	NA
9. Error correction	[confusing]	0 1 2 3 4 5 6 7 8 9 10	[clear]	NA
10. Learning the operation	[difficult]	0 1 2 3 4 5 6 7 8 9 10	[easy]	NA
11. Human memory limitations	[overwhelmed]	0 1 2 3 4 5 6 7 8 9 10	[are respected]	NA
12. Exploration of features	[discouraged]	0 1 2 3 4 5 6 7 8 9 10	[encouraged]	NA
13. Overall reactions	[terrible]	0 1 2 3 4 5 6 7 8 9 10	[wonderful]	NA
	[frustrating]	0 1 2 3 4 5 6 7 8 9 10	[satisfying]	NA
	[uninteresting]	0 1 2 3 4 5 6 7 8 9 10	[interesting]	NA
	[dull]	0 1 2 3 4 5 6 7 8 9 10	[stimulating]	NA
	[difficult]	0 1 2 3 4 5 6 7 8 9 10	[easy]	NA

Table 12-1
An excerpt from QUIS, with permission

SOME DATA COLLECTION TECHNIQUES

- The System Usability Scale (SUS)
  - *The questions are presented as simple declarative statements, each with a five point Likert scale anchored with "strongly disagree" and "strongly agree" and with values of 1 through 5. These 10 statements are:*
  - I think that I would like to use this system frequently
  - I found the system unnecessarily complex
  - I thought the system was easy to use
  - I think that I would need the support of a technical person to be able to use this system
  - I found the various functions in this system were well integrated
  - I thought there was too much inconsistency in this system
  - I would imagine that most people would learn to use this system very quickly
  - I found the system very cumbersome to use
  - I felt very confident using the system
  - I needed to learn a lot of things before I could get going with this system

SOME DATA COLLECTION TECHNIQUES

Usefulness	It helps me be more effective. It helps me be more productive. It is useful. It gives me more control over the activities in my life. It makes the things I want to accomplish easier to get done. It saves me time when I use it. It meets my needs. It does everything I would expect it to do.
Ease of use	It is easy to use. It is simple to use. It is user-friendly. It requires the fewest steps possible to accomplish what I want to do with it. It is flexible. Using it is effortless. I can use it without written instructions. I do not notice any inconsistencies as I use it. Both occasional and regular users would like it. I can recover from mistakes quickly and easily. I can use it successfully every time.
Ease of learning	I learned to use it quickly. I easily remember how to use it. It is easy to learn to use it. I quickly became skillful with it.
Satisfaction	I am satisfied with it. I would recommend it to a friend. It is fun to use. It works the way I want it to work. It is wonderful. I feel I need to have it. It is pleasant to use.

Table 12-2
Questions in USE questionnaire

SOME DATA COLLECTION TECHNIQUES

- Other questionnaires
  - General-purpose usability questionnaires:
    - Computer System Usability Questionnaire (CSUQ), developed by Jim Lewis (1995, 2002) at IBM, is well-regarded and available in the public domain.
    - Software Usability Measurement Inventory (SUMI) is "a rigorously tested and proven method of measuring software quality from the end user's point of view" (Human Factor Research Group, 1990). According to Usability Net, SUMI is "a mature questionnaire whose standardization base and manual have been regularly updated." It is applicable to a range of application types from desk-top applications to large domain-complex applications.
    - After Scenario Questionnaire (ASQ), developed by IBM, is available in the public domain (Bangor, Kortum, & Miller, 2008, p. 575).
    - Post-Study System Usability Questionnaire (PSSUQ), developed by IBM, is available in the public domain (Bangor, Kortum, & Miller, 2008, p. 575).

SOME DATA COLLECTION TECHNIQUES

- Web evaluation questionnaires:
  - Website Analysis and MeasureMent Inventory (WAMMI) is "a short but very reliable questionnaire that tells you what your visitors think about your web site" (Human Factor Research Group, 1996b).
- Multimedia system evaluation questionnaires:
  - Measuring the Usability of Multi-Media Systems (MUMMS) is a questionnaire "designed for evaluating quality of use of multimedia software products" (Human Factor Research Group, 1996a).
- Hedonic quality evaluation questionnaires:
  - The Lavie and Tractinsky (2004) questionnaire
  - The Kim and Moon (1998) questionnaire with differential emotions scale

SOME DATA COLLECTION TECHNIQUES

- Modifying questionnaires for your evaluation
  - choosing a subset of the questions
  - changing the wording in some of the questions
  - adding questions of your own to address specific areas of concern
  - using different scale values

SOME DATA COLLECTION TECHNIQUES

- Data Collection Techniques Especially for Evaluating Emotional Impact
  - Self-reported indicators of emotional impact
  - Questionnaires as a verbal self-reporting technique for collecting emotional impact data (AttrakDiff and others)
  - Observing physiological responses as indicators of emotional impact
  - Bio-metrics to detect physiological responses to emotional impact

SOME DATA COLLECTION TECHNIQUES

Scale Item	Semantic Anchors	
Pragmatic Quality 1	Comprehensible	Incomprehensible
Pragmatic Quality 2	Supporting	Obstructing
Pragmatic Quality 3	Simple	Complex
Pragmatic Quality 4	Predictable	Unpredictable
Pragmatic Quality 5	Clear	Confusing
Pragmatic Quality 6	Trustworthy	Shady
Pragmatic Quality 7	Controllable	Uncontrollable
Hedonic Quality 1	Interesting	Boring
Hedonic Quality 2	Costly	Cheap
Hedonic Quality 3	Exciting	Dull
Hedonic Quality 4	Exclusive	Standard
Hedonic Quality 5	Impressive	Nondescript
Hedonic Quality 6	Original	Ordinary
Hedonic Quality 7	Innovative	Conservative
Appeal 1	Pleasant	Unpleasant
Appeal 2	Good	Bad
Appeal 3	Aesthetic	Unaesthetic
Appeal 4	Inviting	Rejecting
Appeal 5	Attractive	Unattractive
Appeal 6	Sympathetic	Unsympathetic
Appeal 7	Motivating	Discouraging
Appeal 8	Desirable	Undesirable

Table 12-3
AttrakDiff emotional impact questionnaire as listed by Hassenzahl, Scho ̈bel, and Trautman (2008), with permission

SOME DATA COLLECTION TECHNIQUES

Scale	Item	English Anchor 1	English Anchor 2
Pragmatic quality	PQ1	People centric	Technical
Pragmatic quality	PQ2	Simple	Complex
Pragmatic quality	PQ3	Practical	Impractical
Pragmatic quality	PQ4	Cumbersome	Facile
Pragmatic quality	PQ5	Predictable	Unpredictable
Pragmatic quality	PQ6	Confusing	Clear
Pragmatic quality	PQ7	Unmanageable	Manageable
Hedonic – identity	HQI1	Isolates	Connects
Hedonic – identity	HQI2	Professional	Unprofessional
Hedonic – identity	HQI3	Stylish	Lacking style
Hedonic – identity	HQI4	Poor quality	High quality
Hedonic – identity	HQI5	Excludes	Draws you in
Hedonic – identity	HQI6	Brings me closer to people	Separates me from people
Hedonic – identity	HQI7	Not presentable	Presentable
Hedonic – stimulation	HQS1	Original	Conventional
Hedonic – stimulation	HQS2	Unimaginative	Creative
Hedonic – stimulation	HQS3	Bold	Cautious
Hedonic – stimulation	HQS4	Innovative	Conservative
Hedonic – stimulation	HQS5	Dull	Absorbing
Hedonic – stimulation	HQS6	Harmless	Challenging
Hedonic – stimulation	HQS7	Novel	Conventional
Attractiveness	ATT1	Pleasant	Unpleasant
Attractiveness	ATT2	Ugly	Pretty
Attractiveness	ATT3	Appealing	Unappealing
Attractiveness	ATT4	Rejecting	Inviting
Attractiveness	ATT5	Good	Bad
Attractiveness	ATT6	Repulsive	Pleasing
Attractiveness	ATT7	Motivating	Discouraging

Table 12-4
A variation of the AttrakDiff emotional impact questionnaire, as listed in Appendix A1 of Schrepp, Held, and Laugwitz (2006), reordered to group related items together, with permission

SOME DATA COLLECTION TECHNIQUES

- Data Collection Techniques to Evaluate Phenomenological Aspects of Interaction
  - Long-term studies required for phenomenological evaluation
  - Goals of phenomenological data collection techniques
  - Diaries in situated longitudinal studies
  - Evaluator triggered reporting for more representative data
  - Periodic questionnaires over time
  - Direct observation and interviews in simulated real usage situations

VARIATIONS IN FORMATIVE EVALUATION RESULTS

- Reasons given by Hertzum and Jacobsen (2003) for the wide variation
  - vague goals (varying evaluation focus)
  - vague evaluation procedures (the methods do not pin down the procedures so each application is a variation and an adaptation)
  - vague problem criteria (it is not clear how to decide when an issue represents a real problem)

Textbook Chapter 13.

RAPID EVALUATION METHODS

INTRODUCTION

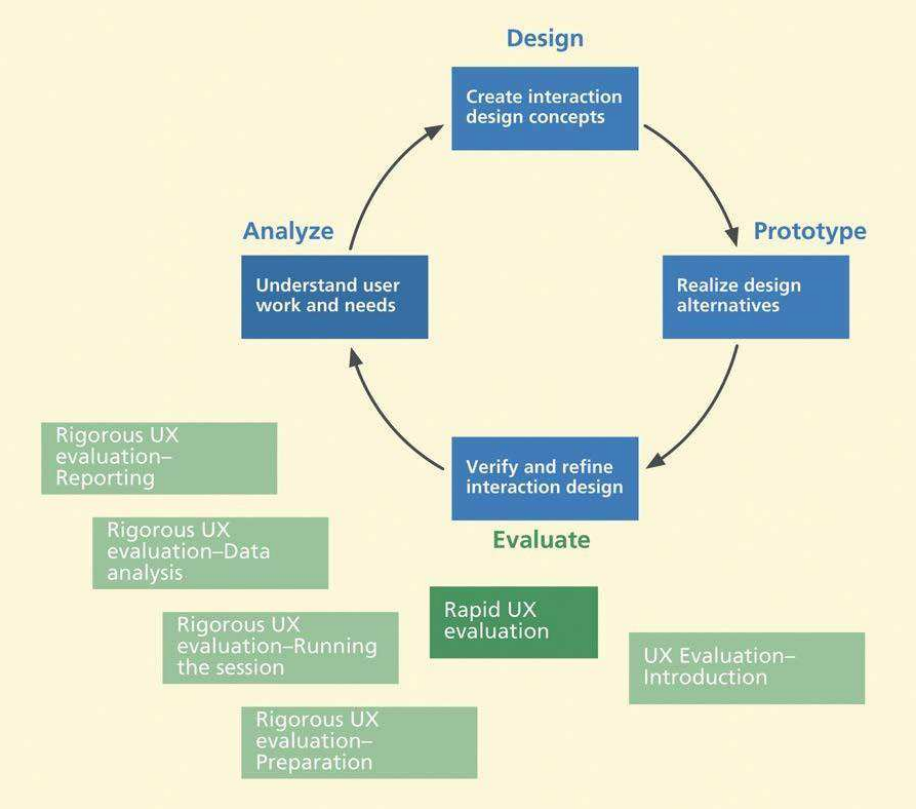

Figure 13-1
You are here, the chapter on rapid evaluation, within the evaluation activity in the context of the overall Wheel lifecycle template.

INTRODUCTION

- The general characteristics of rapid evaluation methods:
  - Rapid evaluation techniques are aimed almost exclusively at finding qualitative data—finding UX problems that are cost-effective to fix.
  - Seldom, if ever, is attention given to quantitative measurements.
  - There is a heavy dependency on practical techniques, such as the "think-aloud" technique.
  - Everything is less formal, with less protocol and fewer rules.
  - There is much more variability in the process, with almost every evaluation "session" being different, tailored to the prevailing conditions.
  - This freedom to adapt to conditions creates more room for spontaneous ingenuity, something experienced practitioners do best.

DESIGN WALKTHROUGHS AND REVIEWS

- You should prepare for a design walkthrough by gathering at least these items:
  - Design representation(s), including storyboards, screen sketches, illustrated scenarios (scenario text interspersed with storyboard frames and/or screen sketches), paper prototypes, and/or higher fidelity prototypes
  - Descriptions of relevant users, work roles, and user classes
  - Usage or design scenarios to drive the walkthrough

UX INSPECTION

- What Is UX Inspection?
- Inspection Is a Valuable Tool in the UX Toolbox
- How Many Inspectors Are Needed?
- What Kind of Inspectors Are Needed?

HEURISTIC EVALUATION, A UX INSPECTION METHOD

- Introduction to Heuristic Evaluation
- How-to-Do-It: HeuristicEvaluation
  - Heuristics
  - The procedure
  - Reporting
  - Variations abound
  - Limitations

Table 13-1

Nielsen's refined heuristics, quoted with permission from www.useit.com

Visibility of System Status
The system should always keep users informed about what is going on through appropriate feedback within reasonable time.
Match Between System and The Real World
The system should speak the users' language, with words, phrases, and concepts familiar to the user rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.
User Control and Freedom
Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.
Consistency and Standards
Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.
Error Prevention
Even better than good error messages is a careful design that prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.
Recognition Rather Than Recall
Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.
Flexibility and Efficiency of Use
Accelerators—unseen by the novice user—may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.
Aesthetic and Minimalist Design
Dialogues should not contain information that is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.
Help Users Recognize, Diagnose, and Recover from Errors
Error messages should be expressed in plain language (no codes), indicate the problem precisely, and suggest a solution constructively.
Help and Documentation
Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

HEURISTIC EVALUATION, A UX INSPECTION METHOD

Table 13-2
Simple HE reporting form, adapted from Brad Myers

Heuristic Evaluation Report

Dated:

MM/DD/YYYY

Prepared By:

NAME:

SIGNATURE:

Evaluation Of:

Name of system being evaluated: XYZ Website

Other information about the system being evaluated:

Problem #: 1

Prototype screen, page, location of problem:

Name of heuristic: Consistency

Reason for reporting as negative or positive: Inconsistent placement of "Add to Cart" buttons: The "Add to Cart" button is below the item in CDW but above in CDW-G.

Scope of problem: Every product page

Severity of problem (high/medium/low): Low—minor, cosmetic problem

Justification for severity rating: Unlikely that users will have trouble with finding or recognizing the button

Suggestions to fix: Move the button on one of the sites to be in the same place as on the other site.

Possible trade-offs (why fix might not work): This may result in an inconsistency with something else, but unknown what that might be.

OUR PRACTICAL APPROACH TO UX INSPECTION

- The Knock on Your Door
- Driven by Experience, Not Heuristics or Guidelines
- Use a Co-Discovery or Team Approach in UX Inspection
- Explore Systematically with a Rich and Comprehensive Usage-Oriented View

OUR PRACTICAL APPROACH TO UX INSPECTION

- Emotional Impact Inspection
  - Is usage fun?
  - Is the visual design attractive (e.g., colors, shapes, layout) and creative?
  - Will the design delight the user visually, aurally, or tactilely?
  - If the target is a product:
    - Is the packaging and product presentation aesthetic?
    - Is the out-of-the-box experience exciting?
    - Does the product feel robust and good to hold?
    - Can the product add to the user's self-esteem?
    - Does the product embody environmental and sustainable practices?
    - Does the product convey the branding of the organization?
    - Does the brand stand for progressive, social, and civic values?
  - Are there opportunities to improve emotional impact in any of the aforementioned areas?

OUR PRACTICAL APPROACH TO UX INSPECTION

- *Most of the questions in a questionnaire for assessing emotional impact are also applicable as inspection questions here. As an example, using attributes from AttrakDiff:*
- Is the system or product interesting?
- Is it exciting?
- Is it innovative?
- Is it engaging?
- Is it motivating?
- Is it desirable?

OUR PRACTICAL APPROACH TO UX INSPECTION

- Use All Your Personalities
- Take Good Notes
- Analyze Your Notes
- Report Your Results

DO UX EVALUATION RITE

- Introduction to the Rapid Iterative Testing and Evaluation (RITE) UX Evaluation Method
- How-to-Do-It: The RITE UX Evaluation Method (Medlock et al. (2002))
  - identifying the characteristics needed in participants
  - deciding on which tasks they will have the participants perform
  - agreeing on critical tasks, the set of tasks that every user must be able to perform
  - constructing a test script based on those tasks
  - deciding how to collect qualitative user behavior data
  - recruiting participants (Chapter 14) and scheduling them to come into the lab
- Variations in RITE Data Collection

QUASI-EMPIRICAL UX EVALUATION

- Introduction to Quasi-Empirical UX Evaluation
- How-to-Do-It: Quasi-Empirical UX Evaluation
  - Prepare
  - Conduct session and collect data
  - Analyze and report results

SPECIALIZED RAPID UX EVALUATION METHODS

- Alpha and Beta Testing and Field Surveys
- Remote UX Evaluation
  - simulating lab-based UX testing using the Internet as a long extension cord to the user (e.g., UserVue by TechSmith)
  - online surveys for getting after-the-fact feedback
  - software instrumentation of click stream and usage event information
  - software plug-ins to capture user self-reporting of UX issues
- Local UX Evaluation
- Automatic UX Evaluation

MORE ABOUT "DISCOUNT" UX ENGINEERING METHODS

- Nielsen and Molich's Original Heuristics(Table 13-3)
- "Discount" Formative UX Evaluation Methods
  - What is a "discount" evaluation method?
  - Do "discount" methods work?
  - Pros and cons as engineering tools
  - Evaluating UX evaluation methods
  - The Comparative Usability Evaluation (CUE) series

MORE ABOUT "DISCOUNT" UX ENGINEERING METHODS

The 10 Original Nielsen and Molich Usablity Inspection Heuristics

- Simple and natural dialogue
  - Good graphic design and use of color
  - Screen layout by gestalt rules of human perception
  - Less is more; avoid extraneous information
- Speak the users' language
  - User-centered terminology, not system or technology centered
  - Use words with standard meanings
  - Vocabulary and meaning from work domain
  - Use mappings and metaphors to support learning
- Minimize user memory load
  - Clear labeling
- Consistency
  - Help avoid errors, especially by novices
- Feedback
  - Make it clear when an error has occurred
  - Show user progress
- Clearly marked exits
  - Provide escape from all dialogue boxes
- Shortcuts
  - Help expert users without penalizing novices
- Good error messages
  - Clear language, not obscure codes
  - Be precise rather than vague or general
  - Be constructive to help solve problem
  - Be polite and not intimidating
- Prevent errors
  - Many potential error situations can be avoided in design
  - Select from lists, where possible, instead of requiring user to type in
  - Avoid modes
- Help and documentation
  - When users want to read the manual, they are usually desperate
  - Be specific with online help

Table 13-3
Original Nielsen and Molich heuristics

MORE ABOUT "DISCOUNT" UX ENGINEERING METHODS

- Yet Somehow Things Work
  - Press on
  - Among the reasons we have to be optimistic in the long run about our UX evaluation methods are:
    - Goals are engineering goals, not scientific goals
    - Iteration helps close the gap
    - Disagreement in the studies was subject to interpretation
    - Evaluation methods can be backed up with UX expertise
  - Practical engineering goals

- Managing risk by mitigating evaluation errors
- Managing the risk of false negatives with iteration
- Managing the risk of false positives with UX expertise
- Look at the bright side of studies.