UX Evaluation

SWEN-444

Selected material from The UX Book, Hartson & Pyla
UX Evaluation

Formative Evaluation (engineering)

User Experience Evaluation (engineering)

Summative Evaluation

Informal Summative Evaluation (engineering)

Formal Summative Evaluation (science)
Formative vs. Summative UX Evaluation

• Formative evaluation helps you *form* design
• Summative evaluation helps you *sum* up design
• “When the cook tastes the soup, that’s formative”
• “When the guests taste the soup, that’s summative”
Formative Evaluation

- Diagnostic nature
- Uses qualitative data
- Goal is to identify UX problems and their causes in design and fix them
Summative Evaluation

- **Assess** the level of *user experience quality* due to a design
- Collect *quantitative* (and qualitative) data
- **Goal** is to improve the UX through *re-design* and formative evaluation iteration as necessary
- **Formal** – *comparative benchmark study* based on *rigorous experimental scientific testing* methods aimed at comparing designs (*not our concern*)
- **Informal** – evaluate user *performance* against UX *targets*
  - Repeatable but informal test methods
  - Collect quantitative data
Rigorous Vs. Rapid Evaluation Methods

- **Rapid evaluation** – less formal, fast, less cost
  - Inspections and walkthroughs
  - Heuristic evaluation
  - Analytical – evaluating design attributes
  - Collect only qualitative data
  - Done earlier in the life cycle
  - Risk – “good enough” but not perfect

- **Rigorous evaluation** – formal, planned process
  - Preparation, data collection, analysis, and reporting
  - Empirical – observe users, collect quantitative performance data (plus qualitative data)
  - In the lab or the field
Data Collection Techniques

- Critical incident identification - an event observed within task performance that is …
  - Significant indicator of UX problem
  - Due to effects of design flaws on user
  - Errors but also other cues such as user hesitation or frustration
- Think –Aloud – the user verbalizes their thoughts during the interactive experience
  - Intensions, rationale, perceptions of problems
  - Easy to do but unnatural
- Questionnaires – collect subjective data from users post evaluation
  - Especially good for emotional impact, perceived usefulness
  - Use Likert scale for best results
The System Usability Scale

Extensively used, widely adapted, in the public domain; uses Likert scale 1-5

1. I think that I would like to use this system frequently
2. I found the system unnecessarily complex
3. I thought the system was easy to use
4. I would need technical support to be able to use this system
5. I found functions in this system integrated
6. Too much inconsistency in this system
7. Most people would learn to use this system very quickly
8. I found system very cumbersome to use
9. I felt very confident using the system
10. I needed to learn a lot of things before I could get going
Evaluating Emotional Impact

- Can be “measured” indirectly in terms of its indicators
- “Emotion is a multifaceted phenomenon”
  - Expressed through feelings
  - Verbal and non-verbal languages
  - Facial expressions and other behaviors
- Emotional impact indicators
  - Self-reported via verbal techniques
  - Physiological responses observed, e.g., facial expressions, body language
  - Physiological responses measured, e.g., biometrics
Surprise

Joy

Anger

Sadness

Fear

Contempt

Disgust

(c) David Matsumoto 2008