Developing Guidelines for Judging the Cost and Benefit of Fixing Usability Problems

**Expert Analysis Method**

1. **Gather Domain Knowledge**
   - **Kinds of Knowledge Gathered**
     - Interviewing the client, analyzing competitor's products, and observing and interviewing users.
     - Proposed Change: We also identify the relevant, key stakeholders (i.e., the people who ultimately decide which problems get fixed) and what business unit they represent, such as Corporate, Marketing, Sales, or Product Management (personal communication with Steve Fadden, User Experience Architect at PeopleSoft). Also see [4].
     - Interviewing the client, analyzing competitor's products, and observing and interviewing users.
     - Physical models [2], information architectures [11], task flows [3, 5], user profiles, personas [2], and scenarios.

2. **Conduct Heuristic Evaluation**
   - **Number of Evaluators**
     - 3 or more.
     - At least one member from each discipline (Human Sciences, Design, and Engineering) and, when feasible, members from the client organization.
   - **Background of Evaluators**
     - In general, we follow Nielsen [7] and Nielsen and Molich [8].
   - **Procedure**
     - Gathering the information may include:
       - Interviews with the client, analyzing competitor’s products, and observing and interviewing users.
       - The proposed change is to identify the relevant, key stakeholders (i.e., the people who ultimately decide which problems get fixed) and the business unit they represent, such as Corporate, Marketing, Sales, or Product Management.
       - Interviewing the client, analyzing competitor’s products, and observing and interviewing users.
       - Other Rating Criteria
   - **Integration of Domain Knowledge**
     - We use affinity diagramming to group the problems into manageable categories (generally between 10 to 15).
     - Examples of categories that might be created during an affinity diagramming exercise:
       - Antiquated user interface
       - Jargon
       - Confusing navigation schemes
     - Misleading instructions
       - Confusing layout
       - No sense of place
     - Inappropriate messaging
       - Unstable system
     - Inconsistent visual design
       - Violating user expectations

3. **Categorize the Issues**
   - **Affinity Diagramming**
     - We use affinity diagramming to group the problems into manageable categories (generally between 10 to 15).
     - Examples of categories that might be created during an affinity diagramming exercise:
       - Antiquated user interface
       - Jargon
       - Confusing navigation schemes
     - Misleading instructions
       - Confusing layout
       - No sense of place
     - Inappropriate messaging
       - Unstable system
     - Inconsistent visual design
       - Violating user expectations

4. **Prioritize the Issues**
   - **Rank Categories According to Importance**
     - The evaluators rank each category according to how important it is from the users’ perspective to fix the problems in that group.
     - Each evaluator votes on the four or five most problematic categories. Categories that receive the most votes are usually placed at the top of the “Important to fix” list. Eventually, all the categories are placed on the list, with the most important categories at the top and the least important at the bottom.
   - **Rank Categories According to Difficulty**
     - The evaluators rank how difficult it would be from the developers’, and designers’ perspectives to fix the problems.
   - **Other Rating Criteria**
     - The x-axis (horizontal) represents Importance (or benefit)– how important it is to fix the categories from the users’ perspective.
     - That is, how much benefit users will experience if the problems are fixed. It is meant to be viewed as a continuum from lesser importance (1) to greater importance (5).
     - The y-axis (vertical) represents Difficulty (or cost)– how much time, effort, and cost the client must expend to fix the problem. Likewise, it is meant to be viewed as a continuum from lesser difficulty (1) to greater difficulty (5).
     - By mapping both Importance and Difficulty, the Cost-Benefit Chart essentially depicts the Return on Investment (ROI) associated with addressing each category of issues.

5. **Write Recommendations**
   - After we’ve categorized and prioritized the problems, we generate recommendations for fixing them. To help the client understand problems with specific screens, we often include a picture of the screen with our annotations overlaid.
   - Doesn’t discuss writing a report, but does recommend a group-decision session to discuss the outcome of the evaluation and to generate solutions to the usability problems.
   - Determine recommendations to fix the problem. Make sure each recommendation links the heuristic and a design principle. Report all sources, purposes, techniques, procedures, and findings in a format that is easy to read and understand.
   - Discuss the good features of the product. Comment on the product team for allowing others to criticize their work. Create a table that presents summary results and suggested solutions. List references for our statements, methodology, and suggested solutions.
Developing Guidelines for Judging the Cost and Benefit of Fixing Usability Problems

4 Prioritize the Issues (proposed change)

Rate Each Category for Difficulty

Each evaluator rates each category on how difficult it would be from the developers’ and designers’ perspectives to fix the problems in the category. Evaluators rate each category on three basic dimensions (pervasiveness, impact, and resources), as well as an additional dimension (degree of interest in fixing it) for each of the key business units identified in Step 2.

For example, Evaluator A might rate the “Unstable system” category. The mean difficulty score for the Unstable System category is 2.5.

The degree to which the problem is widespread or pervasive:

None | A great deal

The solution requires more research or major restructuring (IA, hardware, system architecture):

A great deal | None

The product (development) team has adequate resources to fix it:

A great deal | None

The key Corporate stakeholders are interested in fixing it:

Extremely | Not at all

The key Product Management stakeholders are interested in fixing it:

Extremely | Not at all

The key Sales stakeholders are interested in fixing it:

Extremely | Not at all

The ratings across the dimensions are averaged. Thus each evaluator generates one Difficulty score for each category.

Rate Each Category for Importance

Evaluators rate each category on 4 basic Importance dimensions (task interference, mental model, frustration, negative perception).

For example, Evaluator A may rate the “Unstable system” category. The mean importance score for “Unstable System” category is 6.0.

Interfere with the users’ ability to accomplish frequent or critical tasks:

Not at all | Extremely

Contribute to users’ feelings of frustration:

Not at all | Extremely

Contribute to users’ negative perception of the company:

Not at all | Extremely

The ratings across the dimensions are averaged (each evaluator generates one Importance score for each category).

The ratings for each category are averaged across evaluators for a total Importance Rating and Difficulty Rating. Each category is then plotted on a Cost-Benefit Chart according to its total, mean rating.

References