Usability Testing Methodology: Effectiveness of Heuristic Evaluation in E-Government Website Development

Ashok Sivaji, Product Quality & Reliability Engineering, MIMOS Berhad
Azween Abdullah, Computer & Information Sciences,
Alan G. Downe, Management & Humanities
Universiti Teknologi PETRONAS

23 May 2011
Usability Inspection – Heuristic Evaluation

• Aim:
  – Study various usability inspection methods
  – Assess the effectiveness of Heuristic Evaluation
  – Comparison between traditional and modified heuristics
  – Assess the effectiveness of the modified heuristics
  – Case Study: To assess effectiveness of modified heuristics in capturing usability problems in E-Government Website
• According to [5],
  – usage of web heuristics is common in many government organizations
  – the USA, UK, German, and Dutch governments have made it compulsory for their websites to comply with their respective heuristics
  – in light of the size of most e-Government websites, regular usability testing on all parts of the website does not seem feasible within time and financial constraints under which most organizations function
1. Also known as expert review or expert evaluations

2. Involves a review of a product and system by few evaluators.

3. The evaluators perform the review according to accepted usability principles known as Heuristics from the body of research, human factors literature and previous professional experience. The viewpoint is that of the specific target population that will use the product.

4. A “double” specialist, that is someone who is an expert in usability principles or human factors as well as an expert in the domain area (such as healthcare, financial service and so on) or in the particular technology employed by the product, can be more effective than one without such knowledge.
• A demerit scoring system was used for the HE scoring.
• In other words, the higher HE score implies poorer usability due to more defects found.
• Criteria
  – frequency of each problem found
  – the severity of the problem (Minor 1 point, Major – 2 points and Critical – 3 points).
• HE Score
  \[
  \sum_{1 \leq s \leq 3} \sum_{1 \leq h \leq 12} P_s N_h
  \]
Original Heuristics - Nielsens

• Nielsen 10 Heuristics [4]

1. Visibility of system status
2. Match between system and the real world
3. User control and freedom
4. Consistency and standards
5. Error prevention
6. Recognition rather than recall
7. Flexibility and efficiency of use
8. Aesthetic and minimalist design
9. Help users recognize, diagnose, and recover from errors
10. Help and documentation
Modified Heuristics

- 12 Modified Heuristics
  1. Compatibility
  2. Consistency
  3. Error Prevention & Correction
  4. Explicitness
  5. Flexibility &
  6. Functionality
  7. Informative Feedback
  8. Language & Content
  9. Navigation
  10. Privacy
  11. User Guidance & Support
  12. Visual Clarity
Case Study – Heuristic Evaluation on E-Government Website

- Applying traditional heuristics
- Applying modified heuristics
- Assess the effectiveness of Heuristic Evaluation
- Effectiveness of the modified heuristics
Results

Defect Captured by Inspection Type

- 157, 30% (Heuristic Evaluation)
- 368, 70% (User Experience Test)

Source: E-Government Project, MIMOS Berhad (April ‘10 – March’11)

1. Indicates that Heuristic Evaluation & User Experience Testing address different types of issues

2. Generally HE captures more issues than UET as HE is more exploratory while UET is scenario based. <Jeffries et al. (1991) had found that heuristic analysis discovered approximately three times more problems than user testing.>

3. Over a period of time, can observe a gradual reduction in HE & UET issues found per iteration
Results

• Number of Evaluators : 2/cycle

• HE Cycle 1:
  – 209 Defects
  – 1st CUT Phase,
  – Code is CUT for the Design

• HE Cycle 2:
  – 139 Defects
  – System Test Phase
  – New Features Introduced

• HE Cycle 3:
  – 20 Defects
  – System Test Phase
Results

<table>
<thead>
<tr>
<th>Heuristics</th>
<th>Cycle 1</th>
<th>Cycle 2</th>
<th>Cycle 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compatibility</td>
<td>8</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Consistency and Standards</td>
<td>68</td>
<td>72</td>
<td>7</td>
</tr>
<tr>
<td>Error Prevention &amp; Correction</td>
<td>2</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>Explicitness</td>
<td>7</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Flexibility &amp; Control</td>
<td>13</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>Functionality</td>
<td>44</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>Informative Feedback</td>
<td>22</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Language &amp; Content</td>
<td>20</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Navigation</td>
<td>21</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>Privacy</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>User Guidance &amp; Support</td>
<td>12</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Visual Clarity</td>
<td>18</td>
<td>12</td>
<td>4</td>
</tr>
</tbody>
</table>

Total Score for HE Cycle 1 235 162 21

- Consistency and standard heuristics had the highest number of defects, ie 68 and 72 respectively.
- This is a typical scenario for a new, large and complex project such an E-Government involving many developers and designers.
- Although large in numbers, these were mainly minor defects. Once the team starts to integrate the modules, it can be seen that the score has been improved significantly in cycle 3 to 7.
Results

<table>
<thead>
<tr>
<th>Heuristics Score for E-Government Website</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heuristics</td>
</tr>
<tr>
<td>Compatibility</td>
</tr>
<tr>
<td>Consistency and Standards</td>
</tr>
<tr>
<td>Error Prevention &amp; Correction</td>
</tr>
<tr>
<td>Explicitness</td>
</tr>
<tr>
<td>Flexibility &amp; Control</td>
</tr>
<tr>
<td>Functionality</td>
</tr>
<tr>
<td>Informative Feedback</td>
</tr>
<tr>
<td>Language &amp; Content</td>
</tr>
<tr>
<td>Navigation</td>
</tr>
<tr>
<td>Privacy</td>
</tr>
<tr>
<td>User Guidance &amp; Support</td>
</tr>
<tr>
<td>Visual Clarity</td>
</tr>
<tr>
<td><strong>Total Score for HE Cycle 1</strong></td>
</tr>
</tbody>
</table>

• The results also shows that more effort from the development and design was emphasized on reducing the functionality, informative feedback, language and content and navigation from cycle 1 to 2.

• For instance there was significant improvement of functionality score from cycle 1 to 2 being 44 to 11.
Secondary Results

• The usability engineer can use the HE scoring table to understand the usability of the E-Government website
  – state of the defects in the website.
• This enables the development and design team to strategize and focus their efforts on the heuristics that impacts the success of the project.
**Traditional vs Modified Heuristics**

- **Traditional** Heuristics from Nielsen manage to capture 83% of the defects.
- Remaining 17% were captured by **Modified** heuristics.
- 17% of defects would have resulted in escapee !!!!
- Results reinforce the need for integrating traditional and more recently developed methods into the field of the project being tested (medical, e-commerce etc). This is consistent with [6] who found unique bugs across systems by incorporating modified heuristics (added 5 heuristics).

### Defects Captured by Modified Heuristics

- 64, 17%
- 304, 83%

Source: E-Government Project, MIMOS Berhad (April ’10 – March’11)
Conclusion

1. Effectiveness of the HE as usability inspection methodology in capturing defects for E-Government website optimization

2. The heuristic scoring provided by the usability engineer drives the project team to strategize their efforts that will have the most impact to the stakeholder and customers at each cycle of deployment.

3. Effectiveness of the modified heuristics in capturing addition defects for E-Government website optimization

4. Importance of integrating traditional and more recently developed heuristics into the domain or field of the project being tested, in this case E-Government
Future Work – Extension of findings

<table>
<thead>
<tr>
<th>Justification</th>
<th>Future Work</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>The heuristic scoring provided by the usability engineer drives the project team to strategize their efforts that will have the most impact to the stakeholder and customers at each cycle of deployment.</td>
</tr>
<tr>
<td></td>
<td>Importance of integrating traditional and more recently developed heuristics into the domain or field of the project being tested, such as E-Health, E-Learning</td>
</tr>
</tbody>
</table>
References


10. Ashok Sivaji, Muhamad Fahmi, Shi–Tzuuan Soo Azween Abdullah, Alan G. Downe, “Importance of Incorporating Fundamental Usability with Social, Trust & Persuasive Elements for E-Commerce Website” accepted for ICBEIA conference on 5-8 June 2011
13. Jeffries and Desurvire, “Usability testing vs. heuristic evaluation: was there a contest?” ACM SIGGCHI, 1992
Reference

18. Laurie Kantner & Stephanie Rosenbaum, “Usability studies of WWW sites: heuristic evaluation vs. laboratory testing”, 1997
22. Virzi, R. Refining the test phase of usability evaluation: how many subjects is enough? Human Factors, 34, p.457
Usability Testing Methodology:
Effectiveness of Heuristic Evaluation in E-Government Website Development

Ashok Sivaji, Product Quality & Reliability Engineering, MIMOS Berhad
ashok.sivaji@mimos.my

Azween Abdullah, Computer & Information Sciences,
Alan G. Downe, Management & Humanities
Universiti Teknologi PETRONAS
http://www.hidi.icitiesmor.org/

23 May 2011
## Usability at SDLC

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Evaluation (non Performance)</th>
<th>Test (Performance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>User Centered Analysis</td>
<td>Heuristic Evaluation <strong>Scope</strong></td>
<td>User Experience Test /</td>
</tr>
<tr>
<td>Task Analysis</td>
<td>Expert Review</td>
<td>Reverse Card Sort</td>
</tr>
<tr>
<td>Card Sorting</td>
<td>Subjective Evaluation</td>
<td>Expectance Test</td>
</tr>
<tr>
<td>Live Site</td>
<td>Preference Test</td>
<td>Performance Test</td>
</tr>
<tr>
<td></td>
<td>Peer Review</td>
<td>Visual Affordance</td>
</tr>
<tr>
<td></td>
<td>Focus Groups</td>
<td>Brand Definition</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Free Exploration</td>
</tr>
</tbody>
</table>
### Comparison of Modified & Traditional

<table>
<thead>
<tr>
<th>Modified</th>
<th>Traditional</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error Prevention &amp; Correction, Language &amp; Content, User Guidance &amp; Support</td>
<td>Help users recognize, diagnose, and recover from errors</td>
<td>Same</td>
</tr>
<tr>
<td>Consistency</td>
<td>Consistency and standards</td>
<td>Same</td>
</tr>
<tr>
<td>Error Prevention &amp; Correction</td>
<td>Error prevention</td>
<td>Same</td>
</tr>
<tr>
<td>Informative Feedback</td>
<td>Visibility of system status</td>
<td>Same</td>
</tr>
<tr>
<td>User Guidance &amp; Support</td>
<td>Help and documentation</td>
<td>Same</td>
</tr>
<tr>
<td>Visual Clarity</td>
<td>Recognition rather than recall</td>
<td>Similar</td>
</tr>
<tr>
<td>Flexibility &amp; Control</td>
<td>Flexibility and efficiency of use</td>
<td>Similar</td>
</tr>
<tr>
<td>Visual Clarity</td>
<td>Aesthetic and minimalist design</td>
<td>Similar</td>
</tr>
<tr>
<td>Flexibility &amp; Control, Error Prevention &amp; Correction</td>
<td>User control and freedom Flexibility and efficiency of use</td>
<td>Similar</td>
</tr>
<tr>
<td>Language &amp; Content</td>
<td>Match between system and the real world</td>
<td>Similar</td>
</tr>
<tr>
<td>Navigation</td>
<td></td>
<td>New</td>
</tr>
<tr>
<td>Privacy</td>
<td></td>
<td>New</td>
</tr>
<tr>
<td>Functionality</td>
<td></td>
<td>New</td>
</tr>
<tr>
<td>Compatibility</td>
<td></td>
<td>New</td>
</tr>
<tr>
<td>Explicitness</td>
<td></td>
<td>New</td>
</tr>
</tbody>
</table>
Iterative Testing in SDLC

- Design Reviews
- Expert Review /Heuristic Evaluation

- REQUIREMENT
- DESIGN
- DEVELOP
- TEST

- Defects
- Defects
- Defects

- Resolution
- Resolution
- Resolution

User experience test
Acknowledgement

• Sponsoring & Review
  – Director of Product Quality & Reliability Engineering, MIMOS Berhad, Redzuan Abdullah

• Heuristic Evaluators
  – Fahmi Mazlan, Fahmi Romli, Fairul Fahrurazi, Mawardi Selawi, Rosnisa Razak, Soo Shi Tzuaan and Syazwani Abdullah