The Effect of **Severity Ratings** on **Software Developers’ Priority** of Usability Inspection Results

I-USED 2008
Pisa, Italy

Asbjørn Følstad
SINTEF ICT
Content

- Background – a small inconsistency
- A small study – aiming to resolve the inconsistency
- A controversial conclusion - withdrawn
Background I:
The importance of impact

Usability report
Loaded with important and less important usability evaluation results

High severity issues?

Redesign!

Low severity issues?

Recycle Bin
Background II: Impact defined

$$\text{Impact} = \frac{\text{Redesign!}}{100} + \text{Recycle Bin} \times 100$$
**Background III:**

**A small inconsistency**

- Sawyer, Flanders, Wixon: Making a difference – the impact of inspections (CHI’96)
  - Ten usability inspection cases at Digital Equipment Corp.
  - No difference in impact between low, medium or high severity issues (72, 71, 72 %)

- Law: Evaluating the downstream utility of user tests and examining the developer effect (IJHCl, 2006)
  - One case of user tests - developing a multilingual digital academic library (v0.85 -> v1)
  - Low severity issues given lower priority than medium and high severity issues (26, 42, 47 %)
Background III:
A small inconsistency

- Sawyer, Flanders, Wixon: Making a difference – the impact of inspections (CHI’96)
- Law: Evaluating the downstream utility of user tests and examining the developer effect (IJHCl, 2006)

Relevant study (UIM), but surprising results

Almost relevant study (not UIM), but comforting results
A small study:

Details

- Usability inspections
  - Group-based expert walkthrough
  - 4-5 usability experts in each group
  - One group for each case

- Three cases of mobile application development
  - Medical personnel in hospitals
  - Politicians and political advisors
  - Parking wardens

- Evaluation reports
  - 167 usability issues across all cases
  - Rated as cosmetic, serious, critical

- Impact measures
  - Collected three months following report handover
  - Change: (High priority)
  - No change (Medium or low priority – or wrong)
### A small study:

#### Results

<table>
<thead>
<tr>
<th>CASE 1</th>
<th>Not class.</th>
<th>Cosmetic</th>
<th>Serious</th>
<th>Critical</th>
</tr>
</thead>
<tbody>
<tr>
<td>No change</td>
<td>6</td>
<td>3</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>With change</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Impact ratio</td>
<td>14 %</td>
<td>40 %</td>
<td>57 %</td>
<td>40 %</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CASE 2</th>
<th>Not class.</th>
<th>Cosmetic</th>
<th>Serious</th>
<th>Critical</th>
</tr>
</thead>
<tbody>
<tr>
<td>No change</td>
<td>9</td>
<td>15</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>With change</td>
<td>4</td>
<td>5</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>Impact ratio</td>
<td>31 %</td>
<td>25 %</td>
<td>56 %</td>
<td>42 %</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CASE 3</th>
<th>Not class.</th>
<th>Cosmetic</th>
<th>Serious</th>
<th>Critical</th>
</tr>
</thead>
<tbody>
<tr>
<td>No change</td>
<td>31</td>
<td>13</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>With change</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Impact ratio</td>
<td>3 %</td>
<td>13 %</td>
<td>7 %</td>
<td>100 %</td>
</tr>
</tbody>
</table>
A small study:

Results

<table>
<thead>
<tr>
<th>ALL CASES</th>
<th>Not class.</th>
<th>Cosmetic</th>
<th>Serious</th>
<th>Critical</th>
</tr>
</thead>
<tbody>
<tr>
<td>No change</td>
<td>46</td>
<td>31</td>
<td>26</td>
<td>16</td>
</tr>
<tr>
<td>With change</td>
<td>6</td>
<td>9</td>
<td>18</td>
<td>12</td>
</tr>
<tr>
<td>Impact ratio</td>
<td>12 %</td>
<td>23 %</td>
<td>41 %</td>
<td>43 %</td>
</tr>
</tbody>
</table>
A small study:

Conclusion

- Low severity usability issues with significantly lower impact than medium and high severity issues
- Sawyer et al. contradicted; Law supported
- Why diverging results between the studies?
  - Herzum (2007): Development stage may affect impact
  - Early phase: Higher impact -> Less difference between low and high severity issues?
  - Later phase: Lower impact -> Larger differences?
- Both Law’s study and the small study target later phase
- Phases targeted in Sawyer et al. not known.
A controversial conclusion: Withdrawn

Discussion point: Was I right in withdrawing this conclusion?

The present study, as the study of Law, indicates that low severity usability issues are of relatively little value to software developers. This should serve as a reminder for HCI professionals to spend evaluation resources on identification and communication of higher severity usability issues.