User evaluation: how?

Évaluation des utilisateurs: comment?

Helen Purchase
University of Glasgow
January 2012

“To investigate the use of a visual mouse in a text reading task”

“Asking people to read text using a visual mouse and seeing how well they do”

“Seeing if the visual mouse works”

“Is reading a piece of text using a visual mouse more efficient than when using physical mouse?”

“Do users prefer a visual mouse to a physical mouse when drawing graphs?”

“How accurate is the use of a visual mouse when performing fine-grained interaction tasks?”

Tip 1

Express your research aim in terms of a clear, unambiguous research question (with a ? at the end)
Ceiling and Floor effects

- How many nodes can node (17) reach using only three edges?
- What is the length of the shortest path between nodes (11) and (4)? (note: not 'what is the shortest path' as this requires a clumsy method for answering the question)
- How many edges are in the sub-graph formed by nodes (9), (15), (13) and (0)?

Example B

- Let two (highlighted) subsidiaries A and B. At what level (above A and B) is sitting the company controlling both A and B the most directly?
- Let A be a (highlighted) company. In how many different countries (colors) are located the subsidiaries directly controlled by A?
- Let A be a (highlighted) company. What is the length of the shortest path separating it from its headquarters?

Different labelling

The combination of (EO, T, C) is here called a ‘trial’ (T)

Example A (graph layouts):
\[3C \times 4EO \times 3Q = 36 \text{ trials}\]

Example B (DAGMap):
\[3C \times 3EO \times 3Q = 27 \text{ trials}\]
Degrees of ‘reality’

- Abstract data, abstract questions: Example A (graph layout)
- Real data, real questions: Example B (DAGMap)

Fabricated data, real questions

Does this diagram match the textual specification?

The Smalltalk Application Domain

<table>
<thead>
<tr>
<th>First</th>
<th>Second</th>
<th>Third</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>DAGMAP</td>
<td>NODELINK</td>
</tr>
<tr>
<td>Group 2</td>
<td>DAGMAP</td>
<td>COMBINATION</td>
</tr>
<tr>
<td>Group 3</td>
<td>NODELINK</td>
<td>DAGMAP</td>
</tr>
<tr>
<td>Group 4</td>
<td>NODELINK</td>
<td>COMBINATION</td>
</tr>
<tr>
<td>Group 5</td>
<td>COMBINATION</td>
<td>DAGMAP</td>
</tr>
<tr>
<td>Group 6</td>
<td>COMBINATION</td>
<td>NODELINK</td>
</tr>
</tbody>
</table>

Example B
Tip 2
Preparation is the key to success

Interview for problems and changes
Parameters
‘Ceiling’ or ‘floor’ effect
Bugs
Confidence

Example A
A graph is a means of representing connected information. A graph consists of several nodes connected together with edges.

Nodes are represented by rectangles and are numbered. Edges are represented by lines connecting the rectangles.

The shortest path between two nodes is the least numbed of edges that need to be followed to get from one node to the other, taking into account all possible routes through the graph.

Do you have any questions?

Example B

Example C

The aim of this experiment is to investigate the best way to draw relational information represented as a graph. The experiment will take around 1 hour to complete....

Several graph drawings will be displayed individually on the computer screen, each with a question....

All results will be held in strict confidence. If you have any questions, please contact: ...

I agree to voluntarily take part in this experiment:
Signature: __________________ Date: ____________
Look at example graph A and answer the question:
How long is the shortest path between node 9 and node 4?

Answer: The answer is 3.

There are several possible paths from 9 to 4: 9-14-1-17-4; 14-5-4; 6-13-4; 14-23-1-17-5-4; 16-18-17-5-4

The shortest of these paths is of length 3, and it is not possible to get from node 9 to node 4 in less than 3 steps.

Do you have any questions?

Tip 3

Create a procedure document that outlines exactly what the experimenter must do and say.
Tip 4
Only collect, analyse and present data that is meaningful.

1. Count the number of nodes
2. Indicate which drawing you like best
3. Does this experiment help demonstrate that graph drawings are a good way to present relational data?

Example A

Tip 5
Before running the experiment, create some simulated data and perform the analysis on it, and ensure that the data you are about to collect is of the correct form for answering your research question.

Example B

Analysis process

- **STEP 1**: An *Analysis of Variance (ANOVA)* statistical test can be used to determine whether the different conditions have had any effect on the performance.

- **STEP 2**: A *post-hoc pair-wise comparison test* will show where the differences between the conditions lie.
“Which of three different automatic layout algorithms produces the most effective drawings?”

<table>
<thead>
<tr>
<th>Participant</th>
<th>Hierarchical</th>
<th>Orthogonal</th>
<th>Spring</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>22</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>24</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>25</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

The ANOVA formula calculation produces a calculated F value of 3.54.

The critical F value read off a statistical table (with degrees of freedom 2, 48, and a p-value of 0.05) is 3.23.

As the calculated F value is > the critical F value, there are significant differences between the performances resulting from the use of the three conditions.

SPSS produces the same F value (3.54), and also gives a p-value, of 0.37. As 0.37 < 0.05, there are significant differences between the performances resulting from the use of the three conditions.

There were significant differences in performance as represented by the error data according to condition (F=3.54>F(2,48, p=0.05)= 3.23).

A post-hoc Tukey test (p=0.05) revealed where the differences lie: The Hierarchical layout produced better performance than the Orthogonal layout. There were no other significant pairwise differences.
Tip 6
Keep copious notes throughout the project, recording the justification for all design decisions made.

- **Aim:**
  - the different conditions, why it is interesting to investigate them
- **Experimental method:**
  - the tasks, EOs, questionnaires etc.
- **Process:**
  - software, tutorials, participants and the participant experience, etc; outcomes of the pilot experiments, and problems arising

- **Data and analysis:**
  - the results, the outcomes of the statistical analysis, bar charts.
- **Conclusions:**
  - state conclusions, in the context of the overall aim of the experiment.
- **Discussion:**
  - wider implications of the conclusions can be discussed; state limitations

Tip 7
Separate the presentation of data (facts) from conclusions (opinion).

Tip 8
Point out the limitations of the experiment before your reviewer does.

Tip 9
Focus on the surprising outcomes: even “negative” results are worth reporting.
Helen’s nine Tips for HCI experimentation:

1. Express your research aim in terms of a clear, unambiguous research question (with a ? at the end).
2. Preparation is the key to success.
3. Create a procedure document that outlines exactly what the experimenter must do and say.
4. Only collect, analyse and present data that is meaningful.
5. Before running the experiment, create some simulated data and perform the analysis on it, and ensure that the data you are about to collect is of the correct form for answering your research question.
6. Keep copious notes throughout the project, recording the justification for all design decisions made.
7. Separate the presentation of data (facts) from conclusions (opinion).
8. Point out the limitations of the experiment before your reviewer does.
9. Focus on the surprising outcomes: even “negative” results are worth reporting.

Any questions?

Feel free to email me:

hcp@dcs.gla.ac.uk