CS160 Pilot Usability Study Assignment

Grading Guidelines (40 points total)

Introduction (4 points, ¼ page)

We awarded 4 points as follow:

2 points –   introduction to the system being evaluated
2 points –   purpose and rationale behind experiment

Method (9 points, ~ 2 pages)

We awarded 12 points as follow:

1 point –   description of participants (e.g. demographics)
1 point –   brief discussion of how participants were selected

1 points –   brief discussion of equipment used, etc.

1 point –   task description 1 and what you looked for
1 point –   task description 2 and what you looked for
1 point –   task description 3 and what you looked for

3 points –   description of testing procedure, including what you did and how

The point for equipment used was awarded automatically if no special equipment was required. You need to clearly communicate which of the tasks are the easy, medium and difficult ones.

Wizard of Oz is a key element in several groups' pilot usability test, and you should elaborate on how you faked the necessary interaction as part of the test procedure, if applicable.

Test Measures (3 points, ¼ page)

We awarded 3 points as follow:

1 point –   test measures chosen, including the process data (i.e. what is happening in the big picture) and bottom-line data (i.e. time or # of errors) collected
2 points –   rationale for above test measures

Results (12 points, 1 page)

We awarded 12 points as follow:

4 points –   adequate coverage of findings, in terms of bottom-line data
6 points –   adequate coverage of findings, in terms of process data (critical incidents)
2 points –   reasonable severity ratings

We were lenient in grading this section, and accepted most answers as long as they *summarized* the test measures and critical incidents reasonably, with *average* severity ratings given for most critical incidents. We also accepted descriptions of severity ratings (i.e. "major usability problem", "catastrophic problem", etc.) in lieu of the numerical ratings, as well as critical incidents that omitted the severity ratings if the latter could be easily inferred.

But we could not award points for assignments that did not report results pertaining to the test measures. It is incomplete to decide on test measures (both process data and bottom-line data) in the previous section, and not to follow-up in this section by reporting on them.

Discussion (12 points, 1 page)

We awarded 12 points as follow:

6 points –   what you might change for the "real" experiment
6 points –   what you might change in your interface from these results alone