Severity Ratings for Usability Problems

by Jakob Nielsen

Severity ratings can be used to allocate the most resources to fix the most serious problems and can also provide a rough estimate of the need for additional usability efforts. If the severity ratings indicate that several disastrous usability problems remain in an interface, it will probably be unadvisable to release it. But one might decide to go ahead with the release of a system with several usability problems if they are all judged as being cosmetic in nature.

The severity of a usability problem is a combination of three factors:

- The **frequency** with which the problem occurs: Is it common or rare?
- The **impact** of the problem if it occurs: Will it be easy or difficult for the users to overcome?
- The **persistence** of the problem: Is it a one-time problem that users can overcome once they know about it or will users repeatedly be bothered by the problem?

Finally, of course, one needs to assess the **market impact** of the problem since certain usability problems can have a devastating effect on the popularity of a product, even if they are "objectively" quite easy to overcome. Even though severity has several components, it is common to combine all aspects of severity in a single severity rating as an overall assessment of each usability problem in order to facilitate prioritizing and decision-making.

The following 0 to 4 rating scale can be used to rate the severity of usability problems:

0 = I don't agree that this is a usability problem at all  
1 = Cosmetic problem only: need not be fixed unless extra time is available on project  
2 = Minor usability problem: fixing this should be given low priority  
3 = Major usability problem: important to fix, so should be given high priority  
4 = Usability catastrophe: imperative to fix this before product can be released
Severity Ratings in Heuristic Evaluation

It is difficult to get good severity estimates from the evaluators during a heuristic evaluation session when they are more focused on finding new usability problems. Also, each evaluator will only find a small number of the usability problems, so a set of severity ratings of only the problems found by that evaluator will be incomplete. Instead, severity ratings can be collected by sending a questionnaire to the evaluators after the actual evaluation sessions, listing the complete set of usability problems that have been discovered, and asking them to rate the severity of each problem. Since each evaluator has only identified a subset of the problems included in the list, the problems need to be described in reasonable depth, possibly using screendumps as illustrations. The descriptions can be synthesized by the evaluation observer from the aggregate of comments made by those evaluators who had found each problem (or, if written evaluation reports are used, the descriptions can be synthesized from the descriptions in the reports). These descriptions allow the evaluators to assess the various problems fairly easily even if they have not found them in their own evaluation session. Typically, evaluators need only spend about 30 minutes to provide their severity ratings. It is important to note that each evaluator should provide individual severity ratings independently of the other evaluators.

Often, the evaluators will not have access to the actual system while they are considering the severity of the various usability problems. It is possible that the evaluators can gain additional insights by revisiting parts of the running interface rather than relying on their memory and the written problem descriptions. At the same time, there is no doubt that the evaluators will be slower at arriving at the severity ratings if they are given the option of interacting further with the system. Also, scheduling problems will sometimes make it difficult to provide everybody with computer access at convenient times if special computer resources are needed to run a prototype system or if software distribution is limited due to confidentiality considerations.

My experience indicates that severity ratings from a single evaluator are too unreliable to be trusted. As more evaluators are asked to judge the severity of usability problems, the quality of the mean severity rating increases rapidly, and using the mean of a set of ratings from three evaluators is satisfactory for many practical purposes.