Heuristic Evaluation

A usability evaluation method in which one or more reviewers, preferably experts, compare a software, documentation, or hardware product to a list of design principles (commonly referred to as heuristics) and identify where the product does not follow those principles.

The original justifying assumptions about the heuristic evaluation method presented by Nielsen (1989) and Nielsen and Molich (1990) were: the method is relatively inexpensive, evaluators do not have to be usability experts, there would be evaluations by “several” evaluators, and that when results are aggregated, the evaluators will find “most” of the issues identified by more expensive methods.

The literature presents mixed messages about the relative advantages and disadvantages of the heuristic review method. Cockton and Woolrych (2002) reviewed “discount” usability methods from the perspective of a cost benefit analysis. They point out that actual user problems result from a complex interaction between the user and the system (Cockton & Woolrych, 2002). They believe that discount methods, including heuristic reviews, are too simple to accurately evaluate this interaction (Cockton & Woolrych, 2002). They concluded that these methods are so prone to error that the potential costs far outweigh the benefits (Cockton & Woolrych, 2002). Cockton and Woolrych (2002) recommend that these methods “should be cleared off the HCI store’s shelves.”

Molich and Dumas (2008) reviewed the results of Comparative Usability Evaluation 4 (CUE-4). In the study, 17 teams of usability professionals evaluated the reservation system for the Hotel Pennsylvania. Nine of the teams performed usability tests and eight teams performed expert reviews . While only one expert review team used a heuristic review as described by Nielsen and Molich (1990), Molich and Dumas’s general conclusions about how expert reviews compared to usability testing are of interest.

* There were no false alarms, which contradicts a common belief that expert reviews will produce more false alarms than usability tests (Molich & Dumas, 2008).
* Expert reviews may be more efficient than usability tests, in terms of number of issues found as a function of resources expended (Molich & Dumas, 2008).
* Expert reviews identify the same proportion of major and minor problems as usability tests (Molich & Dumas, 2008), which contradicts earlier studies.

**Benefits**

* Inexpensive relative to other evaluation methods
* Intuitive, and easy to motivate potential evaluators to use the method
* Advanced planning not required (Nielsen & Molich, 1990).
* Can be used early in the development process (Nielsen & Molich, 1990).
* Faster turnaround time than user testing (Kantner & Rosenbaum, 1997).
* Can serve as a good pretest for user testing so that evaluators can stop problems with the system that might disrupt user test sessions.

**Limitations**

* Expert Reviewers Are Better.As originally proposed by Nielsen and Molich, the evaluators would have knowledge of usability design principles, but were not usability experts (Nielsen & Molich, 1990). However, Nielsen subsequently showed that usability experts would identify more issues than non-experts, and “double experts” – usability experts who also had expertise with the type of interface (or the domain) being evaluated – identified the most issues (Nielsen, 1992). Such double experts may be hard to come by, especially for small domains.
* Must use multiple Reviewers and Aggregate Their Findings. Individual evaluators identify a relatively small number of usability issues (Nielsen & Molich, 1990). Multiple evaluators are recommended since a single expert is likely to find only a small percentage of problems. The results from multiple evaluators must be aggregated. (Nielsen & Molich, 1990).
* Finds Low-Severity Issues at a Higher Rate. Heuristic evaluation may identify more minor issues and fewer major issues than would be identified in a think-aloud usability test (Jeffries and Desurvire, 1992).
* Evaluators Focus On Their Areas of Concern. Biased by the preconceptions of the evaluators (Nielsen & Molich, 1990).

Risk of False Positives. Heuristic evaluations may be prone to reporting false alarms – problems that are reported that are not actual usability problems in application (Jeffries, 1994).

**Outcomes**

• Produce a list of potential usability problems, categorized by severity and priority.  
• Provide supporting screenshots or illustrations of the problem.

**Study Execution**

|  |  |
| --- | --- |
| **​Milestone** | **​Owner** |
| ​Initiate Kick-off Call | HFE |
| ​Provide system access or a representation of the system | ​                           Business Office​ |
| ​Recruit Evaluators | ​HFE |
| ​Conduct usability assessment | ​                                    HFE |
| ​Analyze findings | HFE |
| ​Meet to review findings | ​             HFE, Business Office, Developers |
| ​Determine responses and resolutions | ​​                  Business Office, Developers |
| ​Update Findings | ​                                    HFE |
| ​Repeat evaluation as necessary | HFE​, Developers |
| ​Develop final report | ​                                    HFE |
| [**​Conduct After Action Review**](https://vaww.portal2.va.gov/sites/humanfactors/SitePages/AAR%20Process.aspx) |  |

|  |  |
| --- | --- |
| **Phase of Development​** | |
| ​ | ​Planning, Scoping & Definition |
| ​ | ​Requirements Gathering |
| **✔** | **​Early Design** |
| **​✔** | **​Detailed Design & Development** |
| **​✔** | **​Field Testing** |
| ​ | ​Deployment |
| **​✔** | **​Post-Deployment** |

|  |  |
| --- | --- |
| **Study Characteristics ​** | |
| ​**Timeframe** | 2 to 7 days per iteration |
| **​Level of Effort** | Low |
| ​**Data Collection** | Attitudinal |
| ​**Data Reporting** | Qualitative |

|  |  |
| --- | --- |
| **​Related Methods ​** | |
| **Derived from** | Expert Review |
| **​Complimentary Methods** | [**Formative Usability Test**](https://vaww.portal2.va.gov/sites/humanfactors/HFBoK/SitePages/Formative%20Usability%20Test.aspx)  [**Summative Usability Test**](https://vaww.portal2.va.gov/sites/humanfactors/HFBoK/SitePages/Summative%20Usability%20Test.aspx) |
| **​Similar Methods** | [**Participatory Heuristic Evaluation**](https://vaww.portal2.va.gov/sites/humanfactors/HFBoK/SitePages/Participatory%20Heuristic%20Evaluation.aspx) |
| **​Follow-Up** | All |

|  |
| --- |
| **Referenced on this page...** |
| ​Nielsen, J. (1989). Usability engineering at a discount. In Salvendy, G. and Smith, M.J. (Eds.), Designing and Using Human-Computer Interfaces and Knowledge Based Systems , Elsevier Science Publishers, Amsterdam, 394-401. |
| ​Molich, R., and Nielsen, J. (1990). Improving a human-computer dialogue. Communications of the ACM 33, 3 (March), 338-348. |
| ​Jeffries, R. and H. Desurvire (1992). Usability tes- ting vs. heuristic evaluation: was there a contest? SIGCHI Bulletin 24(4), 39-41. |
| ​Jeffries, R. (1994). Usability problem reports: helping evaluators communicate effectively with developers. In Nielsen, J. and Mack, R. (eds.), Usability inspection methods, John Wiley & Sons, Inc., New York, 273-294. |
| Kantner, L., & Rosenbaum, S. (1997). Usability studies of WWW sites: Heuristic evaluation vs. laboratory testing. Proceedings of the 15th Annual International Conference of Computer Documentation, Crossroads in Communication (pp. 153-160)  Salt Lake City, Utah. |
| Woolrych, A., & Cockton, G. (2002). Testing a conjecture based on the DR-AR model of usability inspection method effectiveness. Proceedings of HCI 2002, Volume 2. Retrieved from: [**http://osiris.sunderland.ac.uk/~cs0gco/drar.doc**](http://osiris.sunderland.ac.uk/~cs0gco/drar.doc) |
| ​​Molich, R. and Dumas, J. (2008). Comparative Usability Evaluation (CUE-4). Behaviour & Information Technology, Vol. 27, issue 3. |