Heuristic Evaluation

Nielsen's Ten Heuristics

1. Visibility of system status 　Perceptibility

2. Match system and real world 　Mapping

3. User control and freedom 　Tolerance

4. Consistency and standards 　Consistency

5. Error prevention 　Constraints

6. Recognition rather than recall 　Discoverability

7. Flexibility and efficiency of use 　Flexibility

8. Aesthetic and minimalist design 　Simplicity

9. Help users recognize, diagnose, recover from errors 　Feedback

10. Help and documentation 　Documentation

Heuristic Evaluation

- Developed by Jakob Nielsen (1994)

- Systematic inspection of interface for compliance to guidelines

Heuristic Evaluation

- 3-5 inspectors
- usability engineers, end users, double experts…
- inspect interface in isolation (~1–2 hours for simple interfaces)
- compare notes afterwards
  - single evaluator only catches ~35% of usability problems
  - 5 evaluators catch 75%

- Works for paper, prototypes, and working systems

The Process of Heuristic Evaluation

1)Pre-evaluation training

- Provide the evaluator with domain knowledge if needed

2)Evaluation

- Individuals evaluate interface then aggregate results
  - Compare interface elements with heuristics
- Work in 2 passes
  - First pass: get a feel for flow and scope
  - Second pass: focus on specific elements
- Each evaluator produces list of problems
  - Explain why with reference to heuristic or other information
  - Be specific and list each problem separately

The Process of Heuristic Evaluation

3)Severity rating

- Establishes a ranking between problems
  - Cosmetic, minor, major and catastrophic
- First rate individually, then as a group

4)Debriefing

- Discuss outcome with design team
- Suggest potential solutions
- Assess how hard things are to fix

Examples

Typography uses mix of upper/lower case formats and fonts

Violates: Consistency and Standards (4)

Problem: Slows users down

Fix: pick a single format for entire interface

Probably wouldn't be found by user testing

Severity Rating

- Used to allocate resources to fix problems

- Estimates of need for more usability efforts

- Combination of Frequency, Impact and Persistence

- Should be calculated after all evaluations are in

- Should be done independently by all judges

Levels of Severity

1. don't agree that this is a usability problem

2. cosmetic problem

3. minor usability problem

4. major usability problem; important to fix

5. usability catastrophe; imperative to fix

Severity Ratings Example

1. [4. Consistency] [Severity 3]

The interface used the string "Save" on the first screen for saving the user's file, but used the string "Write file" on the second screen. Users may be confused by this different terminology for the same function.

Debriefing

- Conduct with evaluators, observers, and development team members

- Discuss general characteristics of UI

- Suggest improvements to address major usability problems

- Development team rates how hard things are to fix

- Make it a brainstorming session
  - Little criticism until end of session

Heuristic Evaluation: Pros

- Minimalist approach
  - a few guidelines identify many common usability problems
  - easily remembered, easily applied with modest effort
- "Discount" usability engineering
  - end users not required
  - cheap and fast
  - can be done by usability experts as well as end users
  - provides common evaluation template (to evaluate across systems)

# Heuristic Evaluation: Cons

- Principles are general
  - not a simple checklist
  - subtleties involved in use
  - you may actually have the wrong design altogether

HE vs. User Testing

- HE is much faster
    - 1-2 hours each evaluator vs. days-weeks
- HE doesn't require interpreting user's actions
- User testing is far more accurate
    - Takes into account actual users and tasks
    - HE may miss problems & find "false positives"
- Good to alternate between HE & user-based testing
    - Find different problems
    - Don't waste participants

Why Multiple Evaluators?

- Every evaluator doesn't find every problem
- Good evaluators find both easy & hard ones
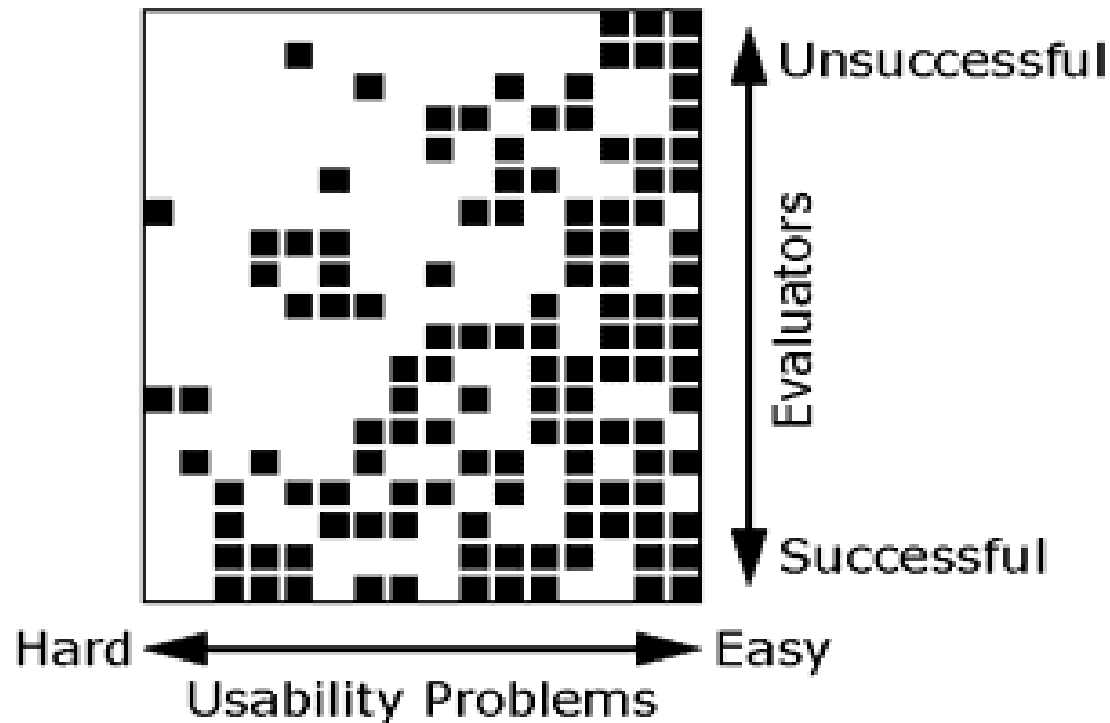
Decreasing Returns
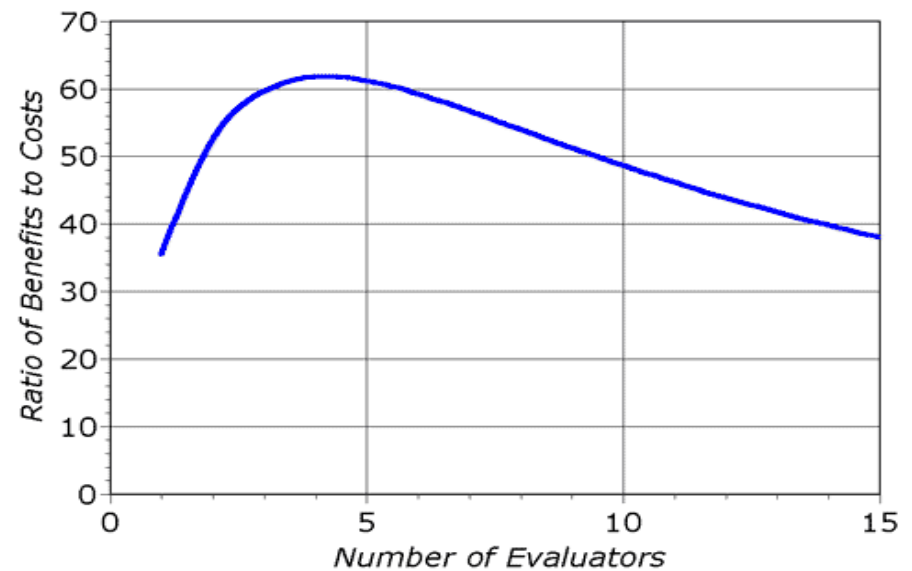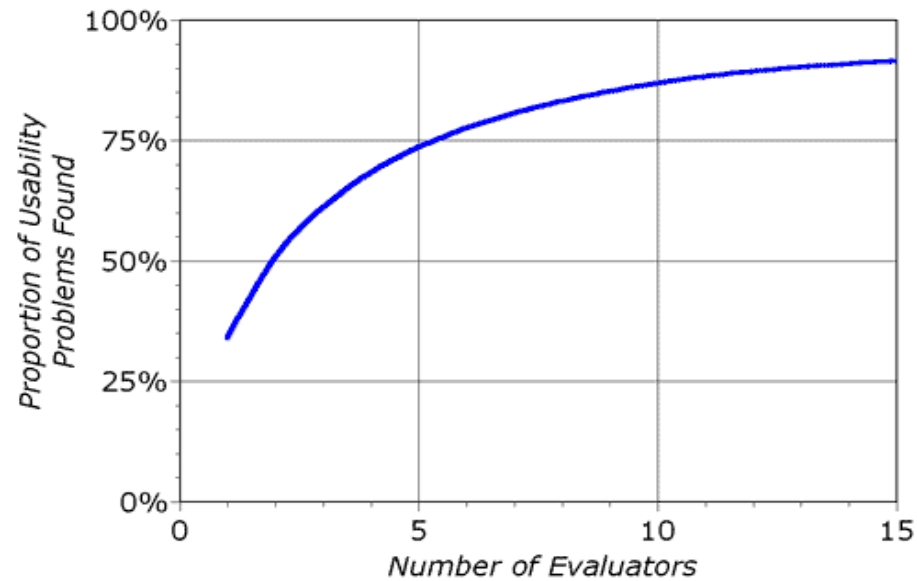
Caveat: graphs are for one specific example!

Number of Evaluators

- Single evaluator achieves poor results
  - Only finds 35% of usability problems
- 5 evaluators find ~ 75% of usability problems
- Why not more evaluators???? 10? 20?
  - Adding evaluators costs more
  - Many evaluators won't find many more problems
- But always depends on market for product:
  - popular products → high support cost for small bugs

Individuals vs teams

- Nielsen
  - recommends individual evaluators inspect the interface alone
- Why?
  - evaluation is not influenced by others
  - independent and unbiased
  - greater variability in the kinds of errors found
  - no overhead required to organize group meetings
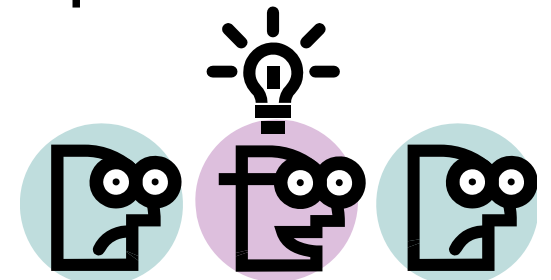
Summary

- Heuristic evaluation is a discount usability method

- Have evaluators go through the UI twice

  - Ask them to see if it complies with heuristics

    - ✓ Note where it doesn't and say why

- Have evaluators independently rate severity

- Combine the findings from 3 to 5 evaluators

- Discuss problems with design team

- Cheaper alternative to user testing

  - Finds different problems, so good to alternate