Three activities:
1. **specify** usability goals early in the development process (the **Usability Specification**)
   - need goals that can be **objectively** measured, ex.: time to complete task, number of errors
2. **measure** the usability goals throughout development
3. **modify design/implementation** as needed for goals

Two uses of Usability Specifications (Hix and Hartson):
1. **Formative evaluation**: evaluation of the UI design during SW development
   - HCI approach: Star life-cycle model (iterative design philosophy) instead of Waterfall model
   - Goal to discover problems in UI design before it is too late to fix them
   - Performed several times during lifecycle
   - May use rapid prototype for testing UI
   - May be performed in a Usability Laboratory
2. **Summative evaluation**: evaluation of UI after completed
   - Used in field testing or beta testing or to compare products
   - Usually performed only once (may be too late!)
Usability specification table: (Hix & Hartson)

- **usability attribute** (e.g. learnability):
- **measuring instrument**: method for measuring this usability attribute - two types:
  - **objective**: measurement while user performing task
    - specify **user class** and **task**
    - ex. *new user* performs task of *renaming file*
    - states **what** the user will do, but not **how**
  - **subjective**: measure user opinion (ex. using questionnaire)

- **measurement**: value to be measured
  - ex. time to complete or number of errors
  - in terms of mean (average) values

- **comparison values**: establish bounds
  - **current level**: old system or current system
  - **worst case**: lowest acceptable level (not worst that could happen)
  - **target level**: success (what you would like)
  - **best case**: realistic state-of-the-art upper limit (attainable, not a wild dream)

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Instrument</th>
<th>Measurement</th>
<th>Current (Unix shell)</th>
<th>Worst</th>
<th>Target</th>
<th>Best</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learnability</td>
<td>Intermittent user copies file in currently displayed folder at 8th level to folder at 5th level</td>
<td>Time to successfully complete task</td>
<td>5.0 sec.</td>
<td>10 sec.</td>
<td>2.0 sec.</td>
<td>1.0 sec.</td>
</tr>
<tr>
<td>Learnability</td>
<td>Same as above</td>
<td>Number of errors before succeeding</td>
<td>0.5 (on average)</td>
<td>2.0</td>
<td>0.3</td>
<td>0.1</td>
</tr>
</tbody>
</table>
Heuristics for meaningful usability specifications:

- make sure that measurement instrument is stated precisely enough to reproduce test each time

- make sure that it is possible and practical to objectively measure the value to be compared to current, worst, target, and best

- current level: if no automated system currently available, then measure current non-automated approach (e.g. paper and pen) or non-skilled user's performance

- worst acceptable: near current level if current level not acceptable to users

- target level: usually set higher than current level, also can look at competitor's systems

- best possible: set higher than target, based upon ideal conditions and expert usage, e.g. measure developer's performance under ideal conditions
**Detailed example - 1**

**System**: web-based airplane flight reservation system

**User group**: first-time users who are familiar with other e-commerce web sites (including competitor's)

**Attribute**: Familiarity (Guessability)

**Instrument**: Reserve seat on cheapest flight satisfying these criteria: depart: Greensboro; arrive: Orlando; roundtrip; depart date: Oct 10, p.m.; return date: Oct 12, p.m.; max stops: 2; aircraft type: jet; restrictions: no refund

**Measurement**: number of screens visited before booking flight

**Current**: 10 (average number of screens for test group of new users on competitor's system)

**Worst**: 10 (same as current for new users)

**Target**: 5 (half current level)

**Best**: 2 (one to find/book flight and one to confirm)
Detailed example - 2

System: web-based airplane flight reservation system

User group: first-time users who are familiar with other e-commerce web sites (including competitor's)

Attribute: Flexibility

Instrument: Reserve seat on cheapest flight satisfying these criteria: depart: Greensboro; arrive: Orlando; roundtrip; depart date: Oct 10, p.m.; return date: Oct 12, p.m.; max stops: 2; aircraft type: jet; restrictions: no refund; price: under $100 if possible, otherwise under $150
  • note: no flights satisfy all these requirements, so user must modify price requirement to succeed

Measurement: user satisfaction on 5 point scale (5=I love it, 4=I like it, 3=I would use this system again, 2=not if I can help it, 1=never again!)
  • Current: 3 (average score for test group of new users on competitor's system)
  • Worst: 4 (we want are worst to be better than competitor's)
  • Target: 4
  • Best: 5
more examples of possible Usability Specification measurement values:
Dix et al. Human-Computer Interaction 2nd ed, Prentice Hall, 1998, ch. 5, Table 5.3 & 5.5

• Time or number of commands to complete a task
• Percent of task completed [per unit time]
• Time spent in errors, Number of errors, Ratio of successes to failures
• Percent of time spent using help
• Number of repetitions of failed commands
• Number of time interface misleads the user
• Number of times user expresses frustration or satisfaction, ratio of positive to negative comments
• Time to learn task, percentage of tasks learned per unit time
Star Life-Cycle Model

- Requirements & Usability Analysis
- User/Task/Functional Analysis
- Usability Evaluation
- Prototyping
- UI Design

Note: may begin in any stage and may repeat as often as necessary