CCHIT 2011 Usability Testing Guide for Ambulatory EHR’s
Contents

- Part 1
  - Overview of Usability Evaluation Process
  - Background

- Part 2
  - How the Usability Program Fits in the Inspection Process
  - Nuts and Bolts of the Rating Process
  - Scoring Methodology

- Part 3
  - References
Overview of Usability Evaluation Process
Overview of Usability Evaluation Process

- All usability questions were developed by CCHIT in consultation with User Centric, Inc., experts in the field of usability testing
- Comprehensive 2011 Ambulatory EHR applicants are required to participate
- Rating process is integrated into clinical portion of the inspection—approximately 30-40 minutes is devoted to the usability rating process
- All scores compiled and averaged, which results in an overall 5 Star rating system
- CCHIT will share juror ratings on each of the questions, so applicants will have valuable, detailed feedback regarding the perceived usability of their system
- The usability rating will NOT affect the certification outcome
- Applicant has the option to publish the results (star rating only); can reverse this decision at any time by contacting CCHIT
- Applicants have the opportunity to apply for Usability Retest at 90-day intervals for a fee
- Usability Testing Guide available at [www.cchit.org/get_certified](http://www.cchit.org/get_certified)
Background
Background

- CCHIT’s Mission is: To accelerate the adoption of robust, interoperable health information technology by creating a credible, efficient certification process.

- Performing usability evaluation is essential to our mission because highly usable systems will increase adoption and less usable systems could require more training and decrease adoption.

- The program launched this year is the first version. More sophistication may be added to the program in the future.
Usability Defined

- Usability is the effectiveness, efficiency, and satisfaction with which the intended users can achieve their tasks in the intended context of product use.

- Adapted from National Institute of Standards & Technology (2007)

- Effectiveness
  - Accuracy in completing tasks

- Efficiency
  - Time and effort used in accomplishment of tasks

- Satisfaction
  - Subjective response to interacting with an application
Objectives of Usability Evaluation

- Develop a first-step instrument that will reflect the perceived usability of an Ambulatory EHR application as rated by content experts (i.e., jurors). The instrument should be:
  - Reliable
  - Face valid
  - Based on (reasonably) observable characteristics
  - Focused on patient safety & efficiency
  - Traceable to literature
  - Easily learned and executed by jurors
  - Reportable to the level of confidence in the precision of the instrument
Usability Program in the Inspection Process
Usability in the Inspection Process

Orientation to EHR System Navigation by Applicant

Applicant Demonstrates Scenarios; Jurors perform inspection as status quo

Jurors Do ASQ Ratings of Scenarios

Jurors Do PERUSE Ratings

Jurors Do SUS Ratings based on Entire Inspection
Orientation to EHR Screen Layout & Navigation

- At the beginning of the inspection process, the applicant will orient the CCHIT jurors and proctor to the following:
  - The overall screen layout for common screens (e.g. patient summary, prescribing, results review)
  - The ways in which users would navigate within a single screen
  - The ways in which users would navigate to other screens
  - The ways in which users would navigate to the home screen (home screen as defined by the vendor)
  - The ways in which users would move from field to field (e.g. tab, cursor placement)
  - A description of key buttons, icons, links that will be frequently used
- Limited to 5-7 minutes (timed by proctor)
Nuts and Bolts of the Rating Process
Rating Model

- Jurors are given a series of questionnaires to create the rating of usability based on observations.
  - After Scenario Questionnaire (ASQ) – jurors rate perceived efficiency (time and effort), learnability, and confidence after viewing scenarios
    - 4 questions after each scenario – 16 overall
  - Perceived Usability Questionnaire (PERUSE) – jurors rate screen-level design attributes based on reasonably observable characteristics
    - 20 questions divided among each of the scenarios; Jurors are allowed to revisit answers to these questions
  - System Usability Survey (SUS) – jurors rate the assessment of usability, and satisfaction with the application
    - 10 questions after all four scenarios have been demonstrated
After Scenario Questionnaire (ASQ)
After Scenario Questionnaire (ASQ)

- After each Scenario, the Jurors will complete four questions that reflect the Juror’s rating of duration, effort, confidence, and learnability with how well the system performs with respect to the scenario they just observed. Four scenarios and four questions will yield 16 responses overall.

- ASQ questions are responded to on a scale from 1 to 5
  - Strongly disagree
  - Disagree
  - Neutral
  - Agree
  - Strongly agree

- ASQ is adapted from Lewis (1995).
After Scenario Questionnaire

- Four questions asked after each of four scenarios during the Inspection

<table>
<thead>
<tr>
<th>Scenario 1</th>
<th>Usability Scenario ASQ (After Scenario Questionnaire)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Question Number</strong></td>
<td><strong>Question</strong></td>
</tr>
<tr>
<td>1</td>
<td>I am satisfied with the time required to complete this scenario – (e.g., time was not excessive)</td>
</tr>
<tr>
<td>2</td>
<td>I am satisfied with the effort it took to complete the scenario - (e.g., keystrokes, mouse clicks, etc. were not excessive)</td>
</tr>
<tr>
<td>3</td>
<td>I am confident that I could use this system to perform the functions in this scenario on a regular basis</td>
</tr>
<tr>
<td>4</td>
<td>I expect that most people would learn to perform the functions in this scenario very quickly</td>
</tr>
</tbody>
</table>
Perceived Usability (PERUSE)
Perceived Usability (PERUSE)

- PERUSE is a list of 20 statements to evaluate user interface components in areas of clinical interest that are known to affect usability.

- Key categories of observations
  - Content Organization
  - Workflow
  - Feedback and Alerts
  - Navigation
  - Screen Layout
  - Highlighting

- Statements will be rated on a scale from 1 to 4 (no mid-point):
  - Strongly disagree
  - Somewhat disagree
  - Somewhat agree
  - Strongly agree

- PERUSE questions are adapted from guidance in Research-Based Web Design and Usability Guidelines (2006).
PERUSE—Question Set 1
PERUSE Question:
The text is easily distinguishable from the background (That is, the text is easily readable)

Principle:
When users are expected to rapidly read and understand prose text, use black text on a plain, high-contrast, non-patterned background.

References:
- Usability Guidelines- Chapter 11.1
Titles and Labels

PERUSE Question:

Field labels accurately describe what data need to be entered; headings for tables and lists accurately describe data.

Principle:

Display an associated label for each data entry field to help users understand what entries are desired.

References:

- Usability Guidelines- Chapter 9
Controls

PERUSE Question:

It is clear which information entry fields are required and which are not.

Principle:

Distinguish clearly and consistently between required and optional data entry fields.

References:

- Usability Guidelines- Chapter 13.1
Controls

PERUSE Question:

Units (for example: milligrams and pounds) are included with the information entry fields so that users are not required to type them in.

Principle:

When using data entry fields, specify the desired measurement units with the field labels rather than requiring users to enter them.

References:

- Usability Guidelines- Chapter 13.16
PERUSE—Question Set 2
Controls

PERUSE Question:

The meaning of icons and symbols is either obvious or the icons/symbols are properly labeled.

Principle:

Use pictorial images to improve the recognition and recall of data, signs and controls.

References:

- Usability Guidelines- Chapter 13.2
Navigation

PERUSE Question:

The system uses a consistent navigational hierarchy so that users understand where to find information.

Principle:

Clearly differentiate navigation elements from one another and from data elements; group and place them in a consistent and easy to find place – typically to the top and left of the screen.

References:

- Usability Guidelines- Chapter 7.2
Screen Layout

PERUSE Question:

Items are in consistent locations across screens. (For example: Patient Name, Medical Record #, etc.)

Principle:

Put important, clickable items in the same locations, and closer to the top of the page, where their location can be better estimated.

References:

- Usability Guidelines- Chapter 6.2
Consistency

PERUSE Question:

Similar objects and controls are used consistently throughout the system.

(For example: Hyperlinks, radio buttons, field labels - system internally consistent)

Principle:

Users bring experiences from other systems and from other parts of the same system. If controls are not used in a manner consistent with user interface standards in common use or controls are inconsistently applied in the application user performance will be negatively impacted.

References:

- Universal Principles - Consistency
PERUSE Question:

If color-coding is used, there is a “key” indicating how the colors should be interpreted.

Principle:

When using color-coding, ensure that the coding scheme can be quickly and easily understood.

References:

- Usability Guidelines- Chapter 11.9
Highlighting

PERUSE Questions:

Some form of visual highlighting is used to draw attention to important information (For example: exceeds acceptable limits, out of range)

Principle:

Visually distinguish (i.e., highlight) important screen elements that require user attention, particularly when those elements are displayed infrequently. However, not everything should be highlighted – there must a noticeable contrast.

References:

- Usability Guidelines- Chapter 9.5
Screen Layout

PERUSE Question:

The screens do not require users to scroll left or right. (For example: There is no horizontal scrolling).

Principle:

Use an appropriate screen layout to bring to a minimum the need for users to scroll horizontally.

References:

- Usability Guidelines- Chapter 8.1
Content Organization

PERUSE Question:

The system uses tables, graphics, and other visualizations to facilitate the understanding of the information.

Principle:

Design quantitative information display to reduce the time required to understand it.

References:

- Usability Guidelines- Chapter 16.6
Content Organization

PERUSE Question:

The meaning of the displayed information is either obvious or explained in the system and users do not need to refer to external documentation. (For example: if results are out of range, users do not have to look to another source to compare ranges).

Principle:

- Display data in formats that require:
  - little or no conversion by the user (e.g., ml to ounces),
  - no reliance on external reference aids, and
  - no visual rotations or transformations.

References:

- Usability Guidelines- Chapter 2.8
PERUSE - Question Set 4
PERUSE Question:

The system uses Mixed Case lettering rather than ALL CAPS in all sections possible.

Principle:

When users must read a lot of information, use lower-case fonts and appropriate capitalization to ensure the fastest possible reading speed.

References:

- Usability Guidelines- Chapter 11.3
Titles and Labels

PERUSE Question:

Labels, information, and alerts are close to the corresponding information entry fields.

Principle:

Ensure that labels are close to their associated data entry fields so that users will recognize the label as describing the data entry field.

References:

- Usability Guidelines- Chapter 9
Feedback and Alerts

PERUSE Question:

Error messages are clear. They explain the reason why the error occurred and suggest what the users should do next.

Principle:

The system design should keep users from getting unintended results and minimize errors of commission or omission.

References:

- Universal Principles - Errors
Screen Layout

PERUSE Question:

Items such as text fields, checkboxes, radio buttons, menus, tables, etc. are neatly aligned on the screens.

Principle:

Visually align screen elements (i.e., controls) to reduce the number of unique margins on the screen vertically; ensure that elements in the same row are horizontally aligned.

References:

- Usability Guidelines- Chapter 6.7
Content Organization

PERUSE Question:

Critical information that is used frequently (for example: allergies and active meds) is visible on the screen in a consistent location.

Principle:

Improve user performance and eliminate error by having high frequency and critical data in a persistent location across screens.

References:

- Universal Principles – 80/20 Rule
PERUSE Question:

There is an easy and obvious way for users to quickly return to the main screen of the patient record from any point in the system.

Principle:

Enable users to access the “main” screen from any screen in the application.

References:

- Usability Guidelines- Chapter 5.1
Consistency

PERUSE Question:

The system uses paradigms that are familiar and consistent with commonly used applications and Web pages.

Principle:

Applications that allow users to positively borrow learning and mental models of operations from other systems are more quickly learned and better retained than those that are novel or in conflict with knowledge of other applications.

References:

- Universal Principles - Consistency
System Usability Scale (SUS)
System Usability Scale (SUS)

- The SUS consists of ten statements – half positive and half negative - designed to collect ratings of usability and satisfaction
- Statements will be judged on a scale from 1 to 5 (with mid-point)
  - Strongly disagree
  - Disagree
  - Neutral
  - Agree
  - Strongly agree
- SUS questions are derived from Brooke (1996).
## System Usability Scale

<table>
<thead>
<tr>
<th>Question</th>
<th>Juror Rating</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>I found the system unnecessarily complex.</td>
<td></td>
<td>1: Strongly Disagree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2: Disagree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3: Neutral</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4: Agree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5: Strongly Agree</td>
</tr>
<tr>
<td>I think that I would need ongoing support to be able to use this system</td>
<td></td>
<td>1: Strongly Disagree</td>
</tr>
<tr>
<td>effectively and efficiently.</td>
<td></td>
<td>2: Disagree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3: Neutral</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4: Agree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5: Strongly Agree</td>
</tr>
<tr>
<td>I thought there was too much inconsistency in this system.</td>
<td></td>
<td>1: Strongly Disagree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2: Disagree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3: Neutral</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4: Agree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5: Strongly Agree</td>
</tr>
<tr>
<td>I think this system would be very cumbersome to use.</td>
<td></td>
<td>1: Strongly Disagree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2: Disagree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3: Neutral</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4: Agree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5: Strongly Agree</td>
</tr>
<tr>
<td>I would need to learn a lot of things before I could begin to use this</td>
<td></td>
<td>1: Strongly Disagree</td>
</tr>
<tr>
<td>system.</td>
<td></td>
<td>2: Disagree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3: Neutral</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4: Agree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5: Strongly Agree</td>
</tr>
<tr>
<td>I think that I would like to use this system frequently.</td>
<td></td>
<td>1: Strongly Disagree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2: Disagree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3: Neutral</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4: Agree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5: Strongly Agree</td>
</tr>
<tr>
<td>I found the various functions in this system were well integrated.</td>
<td></td>
<td>1: Strongly Disagree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2: Disagree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3: Neutral</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4: Agree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5: Strongly Agree</td>
</tr>
<tr>
<td>I thought the system was easy to use.</td>
<td></td>
<td>1: Strongly Disagree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2: Disagree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3: Neutral</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4: Agree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5: Strongly Agree</td>
</tr>
<tr>
<td>I would imagine that most people would learn to use this system very</td>
<td></td>
<td>1: Strongly Disagree</td>
</tr>
<tr>
<td>quickly.</td>
<td></td>
<td>2: Disagree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3: Neutral</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4: Agree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5: Strongly Agree</td>
</tr>
<tr>
<td>I would feel very confident using this system.</td>
<td></td>
<td>1: Strongly Disagree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2: Disagree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3: Neutral</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4: Agree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5: Strongly Agree</td>
</tr>
</tbody>
</table>
Scoring Methodology
Creating the Star Ratings (1 to 5 Stars)

- Following guidance by Tullis and Albert (2008), a single usability score is derived based on a combination of the parts of the survey.
  - The juror scores are summed for each category, then weighted
  - The weighted scores for each juror are then added together to yield the weighted sum for each juror
  - The weighted sums for each juror are added together and averaged to reach an overall score
  - The overall score is compared to the Range Table to determine the Star Rating

- PERUSE and SUS are weighted more heavily (40% each overall) than ASQ (20% overall) because these scales are based on observable characteristics and overall perceived satisfaction respectively
  - Ratings about tasks are limited overall, but it was deemed valuable to have some evaluation and weight given to screen flow (i.e., ASQ)

- Systems with SUS scores between 60-80 are generally considered average in their usability, below 60 are considered to have poor usability, above 80 are considered to above average in usability (Tullis and Albert, 2008)
Example of Summary Scoring and Usability Rating Assignment

<table>
<thead>
<tr>
<th></th>
<th>Scores</th>
<th>Weighting</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Juror</td>
<td>SUS</td>
<td>ASQ</td>
<td>PERUSE</td>
<td>SUS</td>
</tr>
<tr>
<td>A</td>
<td>50.0</td>
<td>83.3</td>
<td>85.3</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>70.0</td>
<td>68.8</td>
<td>93.3</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>40.0</td>
<td>72.9</td>
<td>58.7</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>SUS</td>
<td>ASQ</td>
<td>PERUSE</td>
<td>SUS</td>
</tr>
<tr>
<td>A</td>
<td>100.0</td>
<td>83.3</td>
<td>170.7</td>
<td>354</td>
</tr>
<tr>
<td>B</td>
<td>140.0</td>
<td>68.8</td>
<td>186.7</td>
<td>395</td>
</tr>
<tr>
<td>C</td>
<td>80.0</td>
<td>72.9</td>
<td>117.3</td>
<td>270</td>
</tr>
</tbody>
</table>

Average Weighted Sum: 340

Stars | Range
--- | ---
1    | 0 100
2    | 101 200
3    | 201 300
4    | 301 400
5    | 401 500

© 2009 | Slide 48 | October 2, 2009
Star Ratings – General Meanings

★

Usability issues were observed on multiple dimensions, and included issues that may prevent users from finding, recognizing, or using key features and completing critical tasks. Use of system functionality that is critical to user efficiency or patient safety may be impeded.

★★

Usability issues were observed on multiple dimensions. Many of the observed issues would have the potential to negatively impact user efficiency or create an inconvenience to the user. A few of the observed issues may prevent users from finding, recognizing, or using key features. Use of system functionality that is critical to patient safety would not be severely hindered.

★★★

Most key features were discoverable and usable. While a number of usability issues were observed, use of system functionality that is critical to patient safety would not be hindered. Some observed issues may have a negative impact on user efficiency, though some would only have the potential to cause low-level user irritation.

★★★★

Most or all key features were discoverable and usable. Most observed usability issues would only have the potential to cause low-level user irritation, though a few may have a negative impact on user efficiency. Use of system functionality that is critical to patient safety would not be hindered.

★★★★★

All key features were discoverable and usable. User efficiency would not be dramatically impeded by any of the observed usability issues. All or most observed usability issues would only have the potential to cause low-level user irritation. Use of system functionality that is critical to user efficiency or patient safety would not be hindered.
Results Reporting to Applicants

- Applicants will receive the detailed Juror scores for each question on the ASQ, PERUSE and SUS questionnaires.

- Applicants will receive their overall Star Rating for their application:
  - Including the summary scoring and calculations.

- Applicants will have the option to publish their Star Rating on the CCHIT website.
References & Recommended Reading
References for Measures

- **Definition**
  

- **Single Usability Score**
  

- **ASQ**
  
References for Measures - Continued

- **PERUSE**

  Note: The PERUSE questions were gathered from a number of resources including the two notable ones below. We provide pointers to spots in these references in particular because they offer clear, concise guidance for the principles laid out in the PERUSE section. They are also tied back to primary sources with a research foundation. However, other volumes, like Galitz (see below) offer similar guidance but they may not be as approachable as these two volumes.


- **SUS**


Selected General References: User-Centered Design and Human Factors


Usability Professionals' Association, Body of Knowledge. http://www.usabilitybok.org/design/p287
