NISTIR 7741

NIST Guide to the Processes Approach for Improving the Usability of Electronic Health Records

Robert M. Schumacher
User Centric. Inc,

Svetlana Z. Lowry
Information Access Division
Information Technology Laboratory
National Institute of Standards and Technology

U.S. Department of Commerce
Gary Locke, Secretary

National Institute of Standards and Technology
Patrick D. Gallagher, Director

NISTIR 7741

NIST Guide to the Processes Approach for Improving the Usability of Electronic Health Records

Robert M. Schumacher
User Centric. Inc,

Svetlana Z. Lowry
Information Access Division
Information Technology Laboratory
National Institute of Standards and Technology

November 2010

U.S. Department of Commerce
Gary Locke, Secretary

National Institute of Standards and Technology
Patrick D. Gallagher, Director

NIST Guide to the Processes Approach for Improving the Usability of Electronic Health Records

NIST Guide to the Processes Approach for Improving the Usability of Electronic Health Records 1

NIST Guide to the Processes Approach for Improving the Usability of Electronic Health Records 2

1. Executive Summary 5

2. Intended Purpose of Document 9

3. Intended Audiences 10

4. Usability Defined 10

4.1. Effectiveness 11

4.2. Efficiency 12

4.3. Satisfaction 13

4.4. What Usability is Not 13

4.5. Comparing Usability and Utility 15

5. Motivation 15

5.1. Current State of Usability in Health IT 15

5.2. Benefits of Implementing UCD Process 21

6. User-Centered Design Process in EHRs 22

6.1. Understand User Needs, Workflows and Work Environments 24

6.2. Engage Users Early and Often 25

6.3. Set User Performance Objectives 27

6.4. Design the User Interface from Known Human Behavior Principles and from Familiar User Interface Patterns 29

6.5. Conduct Usability Tests to Measure how Well the Interface Meets the Needs 30

6.6. Adapt the Design and Iteratively Test with Users until Performance Objectives are Met 30

6.7. An Example of UCD Methods 31

7. Unique Features of Developing Usable EHRs 33

8. Organizational Maturity for Usability 37

8.1. Data as a Basis for Decision 38

8.2. Management Support 38

8.3. Design Team 39

8.4. Resources 40

9. Usability Testing for EHRs 42
9.1. What, When and Why of Usability Testing 42
9.2. Essentials of Testing 45
9.2.1. Test Planning 45
9.2.2. Sample Test Plan 47
9.2.3. Test Metrics 49
9.2.4. Test Application 53
9.2.5. Selecting Participants 53
9.2.6. Usability Testing Script / Moderator's Guide 57
9.2.7. Running the Usability Test 57
9.2.8. Doing the analysis 60
9.2.9. Reporting Findings: The Customized Common Industry Format (CIF) for EHRs 61
10. Appendix A: Guide to Completing the Customized Common Industry Format Template for Electronic Health Record Usability Testing 61

1. Executive Summary

Usability has been a topic of considerable interest in the health IT community. Gans et al. (2005) provided evidence that some frequently cited reasons for lack of adoption of Electronic Health Records (EHRs) – security, privacy, and systems integration – are outranked by productivity and usability concerns. Similar themes emerge in reports from AHRQ (Agency for Healthcare Research and Quality) (2009a, 2009b, 2010) and from the Health Information Management Systems Society (2009)[1]. Confirming this, DesRoches et al. (2008) write that "Improving the usability of electronic health records may be critical to the continued successful diffusion of the technology" (p. 57). The AHRQ report recommends that user–centered design and testing processes be implemented.

This document provides NIST guidance for those developing electronic health record (EHR) applications who need to know more about processes of user-centered design (UCD). An established UCD process ensures that designed EHRs are efficient, effective, and satisfying to the user. Following the guidance

[1] Belden J, Grayson R, Barnes J. *Defining and Testing EMR Usability: Principles and Proposed Methods of EMR Usability Evaluation and Rating*. Healthcare Information Management and Systems Society Electronic Health Record Usability Task Force. Available at: http://www.himss.org/content/files/HIMSS_DefiningandTestingEMRUsability.pdf Accessed June 2009.

Armijo D, McDonnell C, Werner K. (2009a) *Electronic Health Record Usability: Interface Design Considerations.* AHRQ Publication No. 09(10)-0091-2-EF. Rockville, MD: Agency for Healthcare Research and Quality. October 2009.

Armijo D, McDonnell C, Werner K. (2009b). *Electronic Health Record Usability: Evaluation and Use Case Framework.* AHRQ Publication No. 09(10)-0091-1-EF. Rockville, MD: Agency for Healthcare Research and Quality. October 2009.

McDonnell C, Werner K., Wendell, L. Electronic Health Record Usability: Vendor Practices and Perspectives. AHRQ Publication No. 09(10)-0091-3-EF. Rockville, MD: Agency for Healthcare Research and Quality. May 2010.

in this document will greatly increase the likelihood of achieving the goal of building usable user interfaces and a better user experience.

One of the main purposes of this guide is to provide practical guidance on methods relating to UCD and usability testing. Specific information is provided, in the context of EHRs, on to how to recruit participants, develop the moderator's guide, write tasks, conduct usability tests, record the data, develop the report, and, ultimately, improve the application.

The intended audiences of this document are those with a role in determining the features and functions contained in the EHR and how those are represented in the user interface. The most important users of this document are those who have authority and responsibility for delivering on the quality of the application: e.g., a product manager or application manager in an organizational leadership position. This document will also be valuable to programmers, systems analysts, business analysts, clinical experts (e.g., doctors and nurses), medical informaticists, information architects, and visual designers.

Usability is defined by the International Organization for Standardization (ISO) as "…the effectiveness, efficiency, and satisfaction with which the intended users can achieve their tasks in the intended context of product use."[2] This definition establishes a framework for setting usability goals and specific evaluation measures. For a given application, measures of these attributes enable comparing the application's progress over time as well as the comparison of competitor applications.

UCD serves to engineer improved usability and human performance into a system or device. UCD models have the following principles:

- Understand user needs, workflows and work environments

[2] ISO 9241-11 (1998): Ergonomic requirements for office work with visual display terminals (VDTs) – Guidance on usability.

- Engage users early and often
- Set user performance objectives
- Design the user interface from known human behavior principles and familiar user interface models
- Conduct usability tests to measure how well the interface meets user needs
- Adapt the design and iteratively test with users until performance objectives are met

UCD is an iterative process that serves to continually improve the application. For each iteration, critical points and issues are uncovered which can be improved upon and implemented in subsequent releases.

Established UCD processes are followed by organizations that have a culture of usability. The degree to which the process of constructing usable experiences is systematized can be evaluated using a Usability Maturity Model3. The purpose of such a model is both diagnostic and prescriptive. Within the Earthy (1999) model there are six levels that describe an organization's embrace of usability and user-centered design:

- 0 – Incomplete: Not able to carry out process
- 1 – Performed: Individuals carry out process
- 2 – Managed: Quality, time and resource requirements for process known and controlled

[3] Earthy, J.V.,1998. Usability Maturity Model: Human-Centeredness Scale. IE2016 INUSE Deliverable D5.1.4s.

Earthy, J. "*Usability Maturity Model: Processes*." Project Report, Lloyd's Register of Shipping, London, 1999.

ISO/IEC (2000) 18529 Human-centered Lifecycle Process Descriptions, ISO/IEC TR 18529: 2000 (E).

- 3 – Established: Process carried out as specified by organization, resources are defined
- 4 – Predictable: Performance of process within predicted resource and quality limits
- 5 – Optimizing: Organization can reliably tailor process to particular requirements

Each of these levels is cumulative and can be characterized by their use of data as a basis of decision making, management support for usability, who is involved on the design team, and resources applied.

Usability testing is a core component of user-centered design. The point of doing a usability test is to improve the EHR whether that means its workflow, navigation, screen layout, interaction, visual design, etc. One should test early in the design/development process ('formative' testing) continuously through to the final stages of development ('summative' testing). These two types of testing should be seen as ends of a continuum. They share the same goal (improve the user interface), but require different techniques and are driven by different measures.

As more of the issues are discovered and corrected through formative testing, more controlled (i.e., summative) studies across broader sections of the user interface, with time and error recording, should be done. Summative testing is important when establishing baselines, measuring the application against benchmark or competing applications, and/or with the goal of ensuring the application is ready for launch. Summative testing may also be necessary when the vendor must provide formal evidence of testing.

The process of usability testing involves detailed planning, including a complete description of recruitment, moderator's guide, methods, tasks and measures.

The measures typically collected include task success, task times, errors, subjective satisfaction, and verbal report.

2. Intended Purpose of Document

This document provides NIST guidance for those developing electronic health record (EHR) applications who need to know more about processes of user-centered design (UCD). UCD ensures that designed EHRs are efficient, effective, and satisfying to the user. Following the guidance in this document will greatly increase the likelihood of achieving the goal of building a usable user interface and a better user experience.

With this document and the *Customized Common Industry Format Template for Electronic Health Record Usability Testing*, the readers will be able to successfully implement a UCD process.

A well-designed user experience derived from good UCD processes delivers a positive return on investment.[4]

In addition, many in the vendor community consider "ease of use" to be a differentiator. Following a UCD process does not mean all user interfaces will look the same, nor does it prescribe the components for an EHR or how those components are put together.

As shown below, UCD is a methodology to improve system usability by (a) applying user research, (b) including users in the development process, (c) measuring and factoring in user behavior, and (d) iterating design-test until all user performance objectives are achieved.

---

[4] See Bias and Mayhew, 1994, *Cost-Justifying Usability*, Academic Press and Bias and Mayhew, 2005, *Cost-Justifying Usability: An Update for the Internet Age*, Morgan Kaufmann Publishers, San Francisco, CA

3. Intended Audiences

The intended audiences of this document are those with a role in determining the features and functions contained in the EHR and/or how those features and functions are represented in the user interface. The most important users of this document are those who have authority and responsibility for delivering on the quality of the application: e.g., a product manager or application manager in an organizational leadership position. This document will also be valuable to programmers, systems analysts, business analysts, clinical experts (e.g., doctors and nurses), medical informaticists, information architects, and visual designers.

This document may be used by vendors as well as by in-house development groups in the hospital and health system community. Vendors strive to deliver an application with an outstanding user experience "out of the box." However, each hospital or health system may modify the "out of the box" application to suit its own needs. In-house groups must ensure user performance is optimized. UCD processes are designed to ensure that the EHR continues to support the user at the highest levels of performance.

Many of the processes described here have enterprise-wide implications. Proper implementation will require breakthrough leadership and strong commitment from senior management.

Application of the UCD process will be a significant contributing factor for passing the criteria of the usability evaluation process of EHRs. This usability evaluation process is anticipated to be developed in the near future.

4. Usability Defined

International standards bodies (ISO 9241-11) define usability as follows:[5]

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

This definition establishes a framework for setting usability goals. These goals are operationally defined by specifying measures of and target values for the attributes: effectiveness, efficiency, and satisfaction. Measures of these attributes enable comparing an application's progress over time as well as the comparison of multiple applications. The next three sections discuss each attribute and what they mean for EHRs.

4.1. Effectiveness

Effectiveness in usability is defined as the degree to which an interface facilitates users in accomplishing their tasks and goals. In general, effectiveness looks at the number of participants who are able to complete a task in a reasonable amount of time. In EHRs, the effectiveness of a system may be how well users can achieve specific goals such as "create a new patient appointment" or "create an order for radiology."

Three main measures of effectiveness include the percentage of tasks accomplished (success rate), the percentage of tasks achieved per unit of time, and the path or manner in which the tasks were achieved. Success must be defined in specific terms for each task. For example, a task may be counted as a success when the user receives a confirmation message saying that the new patient appointment has been successfully created for the specified date and time. A task may not be considered successful if these end points are not reached. Path deviations are instances when the user performs actions that may

[5] ISO 9241-11 (1998): Ergonomic requirements for office work with visual display terminals (VDTs) – Guidance on usability.

be unanticipated by the system designer and may be antithetical to his/her progress in achieving task goals.

Some path deviations may be considered errors. A strict definition of an error is if the user is unable to complete the task or believes he/she completed the task but did so incorrectly. Ordering Drug A rather than Drug B or writing an order in a free text field rather than using structured data fields are both examples of error. A task is not successfully completed if there are errors. Other senses of "error" are better classed as sub-optimal performance on a task; in other words, the user accomplishes the goal, but does so in a roundabout way. For example, the user ultimately accomplishes the task but does so by visiting irrelevant screens, using wrong fields, etc. Consider a user whose goal is to send a prescription to the local pharmacy but the prescription is sent to the wrong pharmacy. In this case, the task is in error. If the wrong pharmacy is detected and corrected within the same session and within the allotted time, the task was successfully completed but sub-optimally. The total number of true errors and sub-optimal behaviors is a negative indication of the effectiveness of an application for a particular combination of user, task, and context. Increased errors result in an application that is less effective, which is typically measured by number of errors and deviations in actual performance from optimal performance (expressed as a ratio of actual steps / optimal steps).

4.2. Efficiency

Efficiency is measured by the length of time required to complete a task. Efficiency can be measured in absolute terms (e.g., 14 seconds) or relative to performance with the same task on other systems or on previous versions of the same system. Efficiency for a task might be compared to a competing application (e.g., ranking applications on efficiency), an absolute standard (e.g., return on investment depends on task times 60 seconds or under), or based on a

measured or estimated value for expert performance (e.g., a fully trained expert should be able to perform the task in 90 seconds 90% of the time).

4.3. Satisfaction

Satisfaction consists of a set of subjective measures regarding a user's perception of usability and evaluation. One standardized rating scale, the Software Usability Measurement Inventory (SUMI), has subscales for efficiency, effect, helpfulness, control and learnability.[6] SUMI and other instruments, such as the System Usability Scale (SUS) have been experimentally validated as good instruments for measuring usability and are publicly available. The SUS has proved to provide fairly robust measures of subjective usability. Brooke (1996)[7] characterized the SUS as a "low-cost usability scale that can be used for global assessments of systems usability." Lewis and Sauro (2009)[8] and others have elaborated on the SUS over the years. Computation of the SUS score can be found in Brooke's paper or in Tullis and Albert (2008).[9]

4.4. What Usability is Not

If usability is defined as above, it can be instructive to distance usability from some common misunderstandings.

- Usability is not User Acceptance Testing (UAT). UAT involves taking use cases or procedures for how the system was designed to perform

[6] Kirakowski, J., & Corbett, M. (1993). SUMI: the Software Usability Measurement Inventory. *British Journal of Educational Technology*, *24*(3), 210-212. doi:10.1111/j.1467-8535.1993.tb00076.x

[7] Brooke, J.: SUS: A "quick and dirty" usability scale. In: Jordan, P. W., Thomas, B., Weerdmeester, B. A., McClelland (eds.) *Usability Evaluation in Industry* pp. 189--194. Taylor & Francis, London, UK (1996). SUS is copyrighted to Digital Equipment Corporation, 1986. Brooke's original paper can be found online at: http://www.usabilitynet.org/trump/documents/Suschapt.doc .

[8] Lewis, J R & Sauro, J. (2009) "The Factor Structure Of The System Usability Scale." in *Proceedings of the Human Computer Interaction International Conference (HCII 2009), San Diego CA, USA*

[9] Tullis, T. & Albert, W. (2008). Measuring the User Experience. Burlington, MA: Morgan Kaufman. Also see www.measuringusability.com

and ensuring that someone who follows the procedure gets the intended result. That is, UAT examines whether the system is capable of performing all specified functions but not necessarily how well the system supports users in performing those functions. Usability testing measures how well the system works (in human performance measures) when *users actually use* the system without the tight procedures dictated by test cases. Intended use and actual use can be two very different things, especially in the case of EHRs where actual users, environments, and situations may differ from defined use cases.

- Usability is not done in focus groups. Focus groups are a valid method for formative usability. Typically, focus groups center on attitude and opinion. Inviting a group of physicians to sit on a focus group to watch a demo of an EHR tells nothing of the overall usefulness or usability of the application. Certainly, some actionable feedback can be obtained, but not at the same level as watching representative users perform realistic tasks. Usability is fundamentally about behavior. In order to assess behavior, users must interact with the system.

- Usability is not graphic or visual design. Graphic design and visual elements can add value to the overall appeal of the system. However, simply applying a pleasing aesthetic to a poorly constructed application does not improve the functionality, the workflow, or the usability.

- Usability is not market research. Market research is, by definition, about markets. Usability goes under the general category of user research.[10] As such, usability is more focused on describing individual behavior and performance and driving systemic improvements back into design than about describing markets and market opportunities in general.

[10] Schumacher, R. (Ed) (2009). Handbook of Global User Research. Burlington, MA: Morgan Kaufman.

4.5. Comparing Usability and Utility

Usability is not usefulness or 'utility'. Utility refers to the existence (or absence) of feature or function necessary to carry out a specific task (e.g., does the EHR have the capability of recording smoking status?). Utility does not reflect whether a feature or function is usable, simply that it is there. Usability is the ease with which those functions can be carried out.

While utility and usability are not truly independent, as a practical matter they can be considered as such. Some functions that are very useful may be very difficult to use. Others that are barely useful may be quite easy to use. The goal for the designer is to make all functions usable, but particularly those that are most useful.

When utility and usability are confused, it is often difficult to untangle the underlying user concerns. For example, if clinical users describe an interface as having 'too many clicks' to perform a task, does this mean that the feature is useful? Or does the number of clicks indicate that the feature is unusable? This can only be evaluated in context.

In the domain of EHRs, usability has, up to this point, been secondary to utility. Ensuring that an EHR has all the features and functions has taken precedence over usability. As the number of features increases, the complexity also increases, demanding more attention to usability. The aim of this document is to expand attention to include usability.

5. Motivation

5.1. Current State of Usability in Health IT

Usability has been a topic of considerable interest in the health IT community. Many authors have commented that the state of affairs for usability of most EHRs is not a positive one.[11]

While the estimates vary, EHRs with at least basic functionality are used by approximately 13% of physicians and 10% of hospitals nationwide.[12] These percentages are relatively low, despite industry and government promotion of EHRs as a means of controlling costs and improving patient care. Government officials have based their estimates of savings and safety improvements on assumptions of increased adoption rates of EHRs by the end of 2014. With the advent of 'Meaningful Use' guidelines,[13] the governmental, technical and industry impetus for adoption is high, which will continue to drive EHRs into the hands of medical providers.

The promise of EHRs' ability to transform medical practice – improve patient safety, enhance efficiency, and improve billing – has been around for some time. Among the most frequently cited drawbacks are cost of implementation, privacy and security. Overcoming these factors alone is not sufficient for successful implementation of any EHR system. To understand why the adoption rate of EHRs has been low, Gans et al. (2005)[14] surveyed experts at nearly 3000 group practices nationwide. As shown in Table 1, Gans et al. identified 15 barriers to EHR adoption.

---

[11] Blumenthal, D. (2009). Stimulating the Adoption of Health Information Technology. N Engl J Med 2009; n engl j med 360;15 april 9, 2009

[12] DesRoches CM, Campbell EG, Rao, SR, et al. Electronic health records in ambulatory care — a national survey of physicians. N Engl J Med 2008;359:50-60.
Hagen, S. (2008). Estimating the Effects of Health IT on Health Care Costs (Congressional Budget Office). Retrieved January 30, 2009 http://www.himss.org/advocacy/d/2008PublicPolicyForumHandouts/StuartHagen.pdf
Jha, DesRoches, Campbell, et al. Use of Electronic Health Records in U.S. Hospitals. N Engl J Med 2009;360:1628-38.

[13] See description of Meaningful Use at: http://healthit.hhs.gov/portal/server.pt?open=512&objID=2996&mode=2

[14] Gans, D., Kralewski, J., Hammons, T., & Dowd, B. (September/October 2005). Medical groups' adoption
of electronic health records and information systems. *Health Affairs*, *24*(5), 1323-1333. Retrieved November 17, 2008, from http://content.healthaffairs.org/cgi/content/full/24/5/1323

Barriers To Implementing Electronic Health Records (EHRs), 2005

Barrier	Mean rating		
	Practices with EHRs	Practices without EHRs	All practices responding
Lack of support from practice physicians	3.32	3.15	3.18
Lack of capital resources to invest in an EHR	3.31	3.58	3.54
Concern about physicians' ability to input into the EHR	3.18	3.40	3.37
Concern about loss of productivity during transition to EHR	3.04	3.24	3.21
Inability to easily input historic medical record data into EHR	2.97	3.24	3.20
Available EHR software does not meet the practice's needs	2.77	2.81	2.81
Insufficient return on investment from EHR system	2.74	3.15	3.09
Lack of support from practice clinical staff	2.73	2.43	2.48
Insufficient time to select, contract, install, implement EHR	2.70	2.88	2.86
Lack of support from practice nonphysician providers	2.68	2.31	2.37
Inability to integrate EHR with practice billing/claims system	2.67	2.90	2.87
Practice staff does not have skills or training to use EHR	2.65	2.62	2.63
Inability to evaluate, compare, and select appropriate EHR	2.60	2.86	2.82
Lack of support from practice administration	2.43	2.06	2.12
Security and privacy concerns	2.31	2.34	2.34

SOURCE: The information in this exhibit is derived from the authors' own analyses.

NOTE: Based on a five-point scale ranging from 1 (not a problem) to 5 (makes implementation very difficult).

Table 1. Barriers to EHR adoption from Gans et al. (2005)

This table shows that well-known factors like security and cost are key, but another theme – usability – floats near the top. Usability is rarely mentioned by name as a barrier to EHR adoption by respondents at these group practices; yet, two of the top five barriers to implementation are related to the usability of EHRs (item 3, concern about physicians' ability to input into the EHR, and item 4, concern about lack of productivity during transition to EHR). While implementation costs are important barriers to practitioners, some other popularly cited reasons for lack of adoption such as security, privacy, and systems integration are outranked by usability and productivity concerns. Similar themes emerge in a widely circulated white paper from the Health Information Management Systems Society (2009) and reports from AHRQ (2009a, 2009b,

2010).[15] Confirming this, DesRoches et al. (2008, p. 57) write that "*Improving the usability of electronic health records may be critical to the continued successful diffusion of the technology."*

Usability issues are also a factor in some cases of EHR implementation failure. In a survey conducted by Linder et al., (2006)[16], primary care physicians were asked to list reasons they did not use the EHRs available to them. Thirty-five percent of those physicians listed specific EHR usability issues, the most common of which were:

- problems with screen navigation,
- no access to secondary functions, and
- concerns that data will be lost.

Anecdotal support for usability and EHR failure comes from Cedars-Sinai Medical Center in Los Angeles. The hospital developed a $34 million Computerized Physician Order Entry system, but only sought the input of a few physicians before launching it hospital-wide in late 2002 without thorough

---

[15] Belden J, Grayson R, Barnes J. *Defining and Testing EMR Usability: Principles and Proposed Methods of EMR Usability Evaluation and Rating*. Healthcare Information Management and Systems Society Electronic Health Record Usability Task Force. Available at: http://www.himss.org/content/files/HIMSS_DefiningandTestingEMRUsability.pdf Accessed June 2009.

Armijo D, McDonnell C, Werner K. (2009a) *Electronic Health Record Usability: Interface Design Considerations.* AHRQ Publication No. 09(10)-0091-2-EF. Rockville, MD: Agency for Healthcare Research and Quality. October 2009.

Armijo D, McDonnell C, Werner K. (2009b). *Electronic Health Record Usability: Evaluation and Use Case Framework.* AHRQ Publication No. 09(10)-0091-1-EF. Rockville, MD: Agency for Healthcare Research and Quality. October 2009.

McDonnell C, Werner K., Wendell, L. Electronic Health Record Usability: Vendor Practices and Perspectives. AHRQ Publication No. 09(10)-0091-3-EF. Rockville, MD: Agency for Healthcare Research and Quality. May 2010.

[16] Linder, J. A., Schnipper, J. L., Tsurikova, R., Melnikas, A. J., Volk, L. A., & Middleton, B. (2006). Barriers to electronic health record use during patient visits. *AMIA 2006 Symposium Proceedings*, 499-503. Retrieved November 17, 2008, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1839290

training.[17] Physicians used to writing a few notes by hand were now required to go through nearly a dozen screens and respond to numerous alerts for even common orders. Usability issues with the "clunky and slow" interface caused more than 400 doctors to demand its removal within three months of its launch.[18]

There are also examples of injury inadvertently caused by a physician arising from poorly designed interfaces to EHRs. One example of a usability failure was a display that did not clearly indicate stop orders for treatment, leading to reported cases of unnecessary drug doses. The Associated Press (2009)[19] reported that "patients at VA health centers were given incorrect doses of drugs, had needed treatments delayed and may have been exposed to other medical errors due to the glitches that showed faulty displays of their electronic health records."

A recent report from AHRQ[20] summarized key findings of interviews conducted with nine EHR vendors as to their perspectives on usability. These findings are presented in Table 2 (emphasis added). The conclusions of the AHRQ report address these findings as follows and are presented in Table 3 (emphasis added):

[17] Connolly, C. (2005, March 21). Cedars-Sinai Doctors Cling to Pen and Paper. *Washington Post*.
Retrieved November 24, 2008, from http://www.washingtonpost.com/wp-dyn/a rticles/A52384-2005Mar20.html

[18] Ornstein, C. (2003, January 22). Hospital heeds doctors, suspends use of software. *Los Angeles Times*. Retrieved November 24, 2008, from http://articles.latimes.com/2003/jan/22/local/me-cedars22

[19] Associated Press (2009, January 15). Lawmaker to investigate software glitches at VA. Retrieved from http://www.google.com/hostednews/ap/article/ALeqM5hzWcaC_f76P1tpPibAn0aRA83TLQD95NMFM02

[20] McDonnell C, Werner K., Wendell, L. Electronic Health Record Usability: Vendor Practices and Perspectives. AHRQ Publication No. 09(10)-0091-3-EF. Rockville, MD: Agency for Healthcare Research and Quality. May 2010.

All vendors expressed a deep commitment to the development and provision of usable EHR product(s) to the market.

Although vendors described an array of usability engineering processes and the use of end users throughout the product life cycle, **practices such as formal usability testing, the use of user-centered design processes, and specific resource personnel with expertise in usability engineering are not common**.

Specific best practices and standards of design, testing, and monitoring of the usability of EHR products are not readily available. Vendors reported use of general (software) and proprietary industry guidelines and best practices to support usability. Reported perspectives on critical issues such as allowable level of customization by customers varied dramatically.

Many vendors did not initially address potential negative impacts of their products as a priority design issue. **Vendors reported a variety of formal and informal processes for identifying, tracking, and addressing patient safety issues related to the usability of their products**.

Most vendors reported that they collect, but do not share, lists of incidents related to usability as a subset of user-reported "bugs" and product-enhancement requests. While all vendors described a process, **procedures to classify and report usability issues of EHR products are not standardized across the industry**.

No vendors reported placing specific contractual restrictions on disclosures by system users of patient safety incidents that were potentially related to their products.

Disagreement exists among vendors as to the ideal method for ensuring usability standards and best practices are evaluated and communicated across the industry as well as to customers. **Many view the inclusion of usability as part of product certification as part of a larger "game" for staying competitive, but also as potentially too complex or something that will "stifle innovation" in this area**.

Because **nearly all vendors view usability as their chief competitive differentiator**, collaboration among vendors with regard to usability is almost nonexistent.

To overcome competitive pressures, **many vendors expressed interest in an independent body guiding the development of voluntary usability standards for EHRs**. This body could build on existing models of vendor collaboration, which are currently focused predominantly on issues of interoperability.

Table 2. Summary of Findings from AHRQ Report on Vendor Perspectives on Usability (AHRQ, 2010 – emphasis added)

Encourage vendors to address key shortcomings that exist in current processes and practices related to the usability of their products. Most critical among these are lack of adherence to formal user-design processes and a lack of diversity in end users involved in the testing and evaluation process.
Include in the design and testing process, and collect feedback from, a variety of end-user contingents throughout the product life cycle. Potentially undersampled populations include end users from nonacademic backgrounds with limited past experience with health information technology and those with disabilities.
Support an independent body for vendor collaboration and standards development to overcome market forces that discourage collaboration, development of best practices, and standards harmonization in this area.
Develop standards and best practices in use of customization during EHR deployment.
Encourage formal usability testing early in the design and development phase as a best practice, and discourage dependence on post deployment review supporting usability assessments.
Support research and development of tools that evaluate and report EHR ease of learning, effectiveness, and satisfaction both qualitatively and quantitatively.
Increase research and development of best practices supporting designing for patient safety.
Design certification programs for EHR usability in a way that focuses on objective and important aspects of system usability.

Table 3. Summary of Conclusions from AHRQ Report on Vendor Perspectives on Usability (AHRQ, 2010 – emphasis added)

If we accept that usability is fundamental to EHR adoption, the essence of the findings from the AHRQ report is that little is being done systematically to improve usability. Not surprisingly, the AHRQ report recommends that user–centered design and testing processes be implemented. Though these processes are necessary, it must be understood that implementing them is not easy, nor is it without some cost. But that cost must be balance against the benefits. In the next section, we touch on the benefits of a UCD process.

5.2. Benefits of Implementing UCD Process

There are costs associated with UCD; however return on investment is much more substantial. Products with high usability have:[21]

- Greater commercial success in the marketplace
- Fewer reported defects

[21] Bias & Mayhew, op cit.

- Lower technical and user support costs
- Lower training costs
- Shorter development time due to fewer change orders

These and other benefits can be balanced against the costs. While there can be a tendency to 'build it now, fix usability later' the 'fix usability later' mentality often fails to materialize because costs to change increase as the product matures.

6. User-Centered Design Process in EHRs

User-centered design is a bedrock principle for creating usable systems and devices. One of the most common reasons why systems are poorly designed is that designers and developers fail to engage users in appropriate ways at appropriate times. At its core UCD is a process that relies on systematic understanding of users and their environments, and iterative design and testing based on user performance objectives. (Details on usability testing are provided in Section 9.)

UCD has been shown to be effective in many fields. In aviation, for example, this method has been used to develop cockpit navigation displays for low-visibility surface operations.[22] By taking the limitations and capabilities of the flight crew into account, navigation errors have decreased by almost 100%. The adoption of UCD has also been shown to be effective in the design of personal computers. When working on a redesign of the laptop computer, a UCD process was employed. Users were asked to offer feedback about the current model and to offer input about ways to improve the current design. User-centered design was

[22] Hooey, B. L., D. C. Foyle, and A. D. Andre. "Integration of Cockpit Displays for Surface Operations: The Final Stage of Human-Centered Design Approach." *SAE Transactions, Journal of Aerospace* 109 (2000): 1053.

successful in increasing market share, brand equity, and customer satisfaction.[23] In fact, user-centered design has been elevated to an ISO standard.[24]

UCD serves to engineer improved human performance into a system or device, and has been crystallizing for several decades as a design philosophy.[25] While there is no singular model of UCD, the instantiations embody the following principles:

- Understand user needs, workflows and work environments
- Engage users early and often
- Set user performance objectives
- Design the user interface from known human behavior principles and familiar user interface models
- Conduct usability tests to measure how well the interface meets user needs
- Adapt the design and iteratively test with users until performance objectives are met

As an iterative process, UCD is a cycle that serves to continually improve the application. For each iteration, critical points and issues are uncovered which can be improved upon and implemented in subsequent releases. An illustration of the UCD process is included in Figure 1.

[23] David A. Sawin, Kazuhiko Yamazaki, Atsushi Kumaki: Putting the "D" in UCD: User-Centered Design in the ThinkPad Experience Development. Int. J. Hum. Comput. Interaction 14(3-4): 307-334 (2002)

[24] ISO FDIS 9241-210:2009. Ergonomics of human system interaction - Part 210: Human-centred design for interactive systems (formerly known as 13407). International Organization for Standardization (ISO). Switzerland.

25 See Norman, D. A., & Draper, S. W. (Eds.) (1986). User centered system design: New perspectives on human-computer interaction. Hillsdale, NJ: Lawrence Erlbaum Associates, Gould, John D., Stephen J. Boies, and Clayton Lewis. "Making Usable, Useful, Productivity: Enhancing Computer Applications." *Communications of the ACM* (January 1991): 72-86., and Vredenburg, K. Isensee, S., and Y Righi, C. (2002). User-Centered Design: An Integrated Approach. Upper Saddle River, NJ: Prentice Hall.

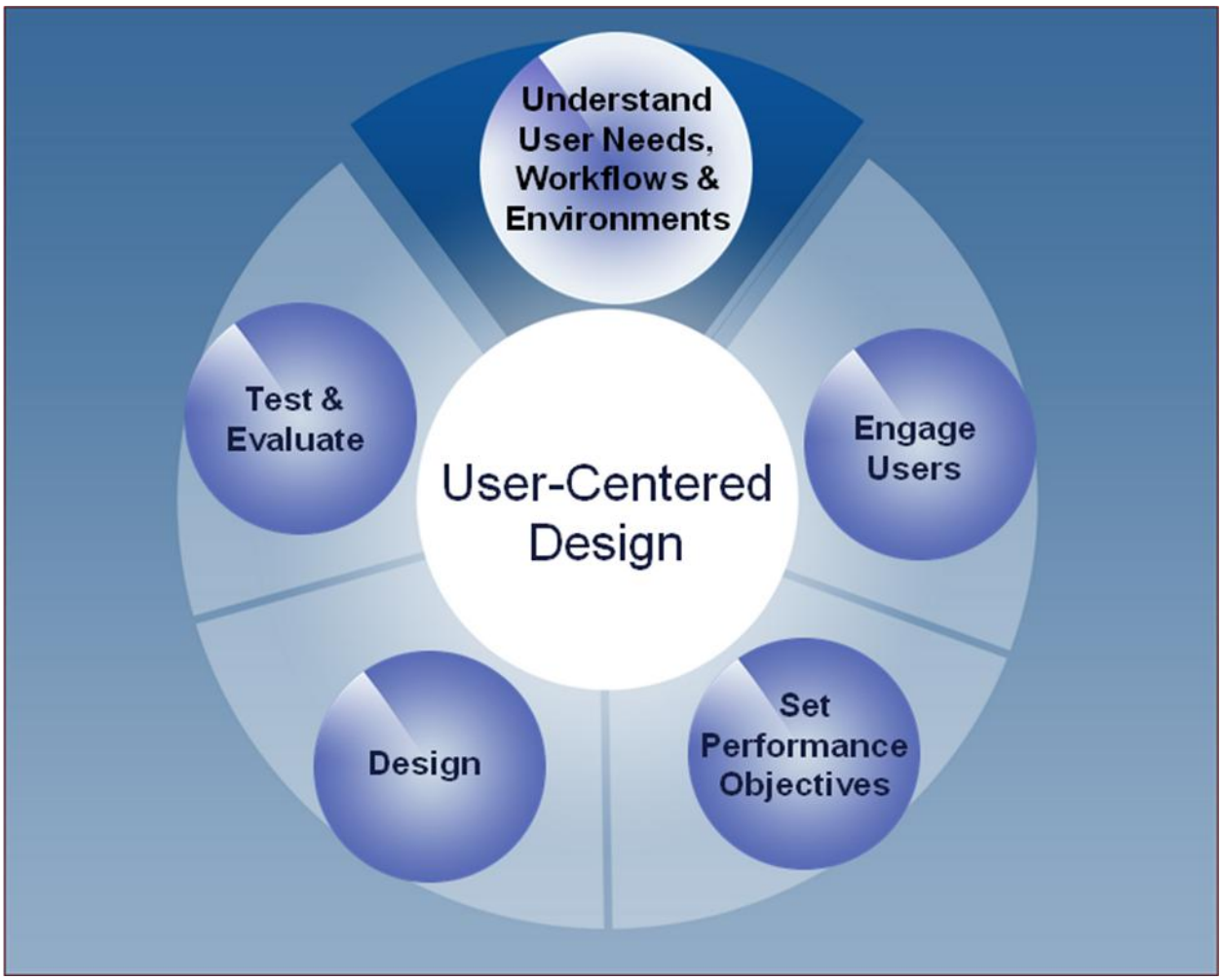


Figure 1. UCD Process involves understanding the background of the users and the work environment, getting the users involved early in the design, setting objectives for the application, doing detailed design, testing, evaluating, and changing. The process cycles can go through many iterations at each step and can cycle repeatedly from understanding through testing.

6.1. Understand User Needs, Workflows and Work Environments

In order to design any interface, the purpose of that interface must first be described. Additionally, the many, varied user goals and needs in the larger environment must be considered. For example, a screen within an EHR is designed with the purpose of ordering prescriptions. The user of this EHR may also need to locate drug information, interpret drug-drug interactions, and correct errors. Thus, the design of the EHR requires documentation of not only the high-

level goal (i.e., place prescription order) but all of the sub-goals and associated work flows. Detailing all these flows, even on something as seemingly straightforward as a recording smoking status, will help the designer create a more usable device.

One must also consider the role of the environment. A system or device is never used in a vacuum. The EHR might be used in a quiet office, overcrowded noisy clinic, or in ways never considered by the designer. While the administrator has little or no control over where a device will be used, it is important that early in its development the administrator gathers sufficient information to understand the typical as well as the out-of-the-ordinary environments in which the system will be used.

In short, design begins by understanding who the users are, their needs, typical and atypical workflows, and the context in which the system will be used.

6.2. Engage Users Early and Often

It is generally accepted that involving end users in EHR design is a good idea. The question is often *when* and *how* to get users involved. One starting point, with respect to users' EHR goals and objectives, is to gather information through qualitative means: ethnographic studies, contextual interviews, patterns of use of existing systems, etc. The outcome of these qualitative methods should include:

- how is the EHR meant to fit within the practice or system,
- where will the EHR be used,
- how often,
- by whom,
- what is the intended user's experience is with similar devices,
- what level of complexity is appropriate,
- barriers to adoption, etc.

The study should be focused around examples of problems that the interface is intended to solve or outcomes that the users' envision. Putting the user goals in

context of the environment brings the design requirements into sharper focus, and should give the user interface team the information to create early prototypes. Getting a picture of the users, the context, and the environment will provide the foundation for the design effort.

In some design groups, the design team simply begins designing screens prior to spending time listening to and observing users. This can result in a user interface design that gets solidified before all of the thinking has a chance to get realized in the system. In the UCD process, design begins in the field by learning about users, their goals, and their environments long before any screens are built.

It is important to note that users are rarely designers. User interface design is a skill that requires specialized knowledge and experience. Asking users to design something is asking for expertise beyond the capability of most users. However, users are very good at reacting to designs. That is, users are usually better at providing feedback about specific cases rather than dealing in abstractions.

Early design prototypes should be presented to a small number of users (e.g., fewer than ten) for their reaction (e.g., early formative usability testing).[26] These incremental design-test cycles are repeated with users until a solid vision is achieved. Users are engaged early during user research and preliminary design but also at each subsequent step such as design reviews and (obviously) during usability testing.

[26] Formative testing is done early in the design cycle with a small group of users to identify and prioritize major problems. Verbal reporting is a critical data component though having participants think aloud can affect performance. Summative testing is done later in the design cycle with larger and more diverse groups of users and gathers real human performance data (i.e., less/no reliance on verbal report). Formative and summative usability testing should be seen as complementary and should be brought in at the appropriate times. There are a variety of books cited earlier and resources on the Internet describing the nuances of formative and summative usability testing (for instance see: ISO 9241-11).

6.3. Set User Performance Objectives[27]

One result from qualitative work, developing application objectives, and early usability testing, is the formation of user performance objectives. Objective measures of success are important to decide when the interface has reached the appropriate maturity to be released. Typically, most user performance objectives are related to effectiveness (e.g., optimal and error-free performance), efficiency (e.g., speed), and satisfaction and subjective assessment. In the UCD process, key performance measures are operationalized for target user groups such that core tasks (e.g., time to update blood pressure on a particular screen) are given target values (e.g., less than 20 seconds by 90% of first-time users). Table 4 shows an example of this.

[27] A rigorous treatment of this can be found at: National Institute of Standards and Technology-NIST (2007).*Common industry specification for usability –requirements NISTIR 7432.* Retrieved from http://zing.ncsl.nist.gov/iusr/documents/CISU-RIR7432.pdf. See Section 6 “Usability requirements specification” in particular.

	Measuring Usability Relative to Goals		
Task	**Effectiveness**	**Efficiency**	**Satisfaction**
Record Patient Demographics (First-time users)	**Goal**: 90% correct on first trial irrespective of path **Test Cycle 1**: 100% correct **Test Cycle 2**: 95% correct	**Goal**: Successful completion in 90 seconds by 80% of users **Test Cycle 1**: 120 Seconds by 80% **Test Cycle 2**: 80 seconds by 95% of participants	**Goal:** All participants give task difficulty ratings of "Easy" or "Very Easy" **Test Cycle 1:** 5 of 8 rated "Easy" or "Very Easy"
Record Patient Demographics (Experienced users)	**Goal**: 100% correct using optimal path on first trial. **Test Cycle 2:** 100% correct; 80% used optimal path	**Goal**: Successful completion in 45 seconds by 95% of users **Test Cycle 2**: 40 seconds by 95% of participants	**Goal:** All participants give task difficulty ratings of "Easy" or "Very Easy" **Test Cycle 1:** 5 of 8 rated "Easy" or "Very Easy"
Provide clinical summary of visit	**Goal**: 100% correct using optimal path on first trial. **Test Cycle 1:** 100% correct; 100% used optimal path	**Goal**: Successful completion in 45 seconds by 95% of users **Test Cycle 1**: 40 seconds by 95% of participants	**Goal:** All participants give task difficulty ratings of "Easy" or "Very Easy" **Test Cycle 1:** 8 of 8 rated "Easy" or "Very Easy"
Prescribe medication (with drug-drug interaction)	**Goal**: 75% correct using optimal path on first trial. **Test Cycle 1:** 100% correct; 80% used optimal path	**Goal**: Successful completion in 45 seconds by 95% of users **Test Cycle 1**: 40 seconds by 95% of participants	**Goal:** All participants give task difficulty ratings of "Easy" or "Very Easy" **Test Cycle 1:** 2 of 8 rated "Easy" or "Very Easy"
User Group: Primary Care Physicians			

Table 4. Measuring of User Performance on Key Tasks

There are many attributes of this table to point out. First, the tasks should be those that are either high frequency, have high importance, or are difficult functions to perform. Tasks should also be meaningful to the design team for commercial or market reasons. Second, the performance measures should be expressed in terms that are operational. That is, efficiency must be operationalized to a dependent variable (e.g., time) that can be reliably measured and will serve as the basis of comparison as well as identify where improvements have been made or further work needs to be done. Third, user groups and numbers of participants need to be specified. Learning criteria can also be indicated in the table by adding criteria for early learners versus experienced users. Not all tasks need to be tested on each test cycle (see "Record Patient

Demographics"). Last, the results from multiple rounds of testing can be recorded in the table. Keeping systematic records of tasks, goals and results will ensure that the application serves the needs of the users.

The development team agrees to the performance objectives and continues to measure them through usability testing at points in the development process. One of the tenets often adhered to is that the interface is not released until all goals have been reached simultaneously. Ultimately this will ensure that the interface is not sub-optimal on one measure for the sake of others.

6.4. Design the User Interface from Known Human Behavior Principles and from Familiar User Interface Patterns

Successful EHR design requires not only an understanding of users' needs, but also knowledge of the human factors design principles that ensure effective, efficient and satisfying usage. The human factors field has a body of knowledge derived from cognitive psychology that supports UCD.[28] The design guidance is built from evidence-based knowledge of human performance, from known best practices, and from commonly used interface patterns. Many of these practices are detailed in design documents and design standards.[29] Having good

[28] There are many. Here are a few:

- Card, S., Moran, T.P., & Newell, A. (1983).The psychology of human computer interaction. New York: Lawrence Erlbaum Associates.
- Wickens and Hollands (2000). Engineering Psychology and Human Performance. Upper Saddle River, NJ: Prentice Hall.
- Wickens, C., Liu, Y., and Gordon, S. (1997). An Introduction to Human Factors Engineering. Upper Saddle River, NJ: Prentice Hall.

[29] A few examples include:

- U.S. Department of Health and Human Services. (2006). Research-Based Web Design & Usability Guidelines. Available at: www.usability.gov/pdfs/guidelines.html.
- Shneiderman, B. (1998). Designing the user interface. Third edition. Addison-Wesley. (First edition published 1987).
- Galitz, W. (2007). The essential guide to user interface design: An introduction to GUI design principles and techniques (3rd Edition). Indianapolis, IN: Wiley Publishing.
- Lidwell, W., Holden, K., and Butler, J. (2003). Universal principles of design. Rockport Publishers.

There are many platform style guides that center around user interface design practices.

knowledge of the users' tasks, environments, and human performance, skilled designers can make usable designs from the outset.

6.5. Conduct Usability Tests to Measure how Well the Interface Meets the Needs

Understanding the objectives is not enough. The designs must be tested with representative users under realistic contexts. The purpose of usability testing is to identify problems for repair during subsequent development and to measure user performance to ensure that the objectives are achieved. Early in the design cycle, testing is often informal with a small number of users (called 'formative' testing). However, as development formalizes the interface, these informal methods become less effective. Formal and summative methods, described in subsequent sections, must be used to ensure that the applications meet the user needs with respect to the performance objectives laid out in the beginning.

6.6. Adapt the Design and Iteratively Test with Users until Performance Objectives are Met

The UCD process ends when the user performance objectives are met through reproducible summative usability testing. The iterative design-test cycle requires the designer to objectively record and review the results, make changes accordingly, and then retest. If the design is built from a good foundation of user research and known user interface standards and conventions, the number of test iterations can be quite minimal.

Usability testing must be built into the overall planning of application development. In many cases, much of the UCD process is conducted before a single line of application code is written.

UCD has been rightfully perceived as being more evolutionary than revolutionary in improving performance. UCD is not likely to produce incredible breakthroughs

in human performance. However, often what is needed is solid incremental improvement, and this is where the UCD process is very beneficial.

6.7. An Example of UCD Methods

Figure 2 portrays a typical UCD methodology. The details of the process were covered in the previous sections, but on a practical level it is instructive to illustrate the process.

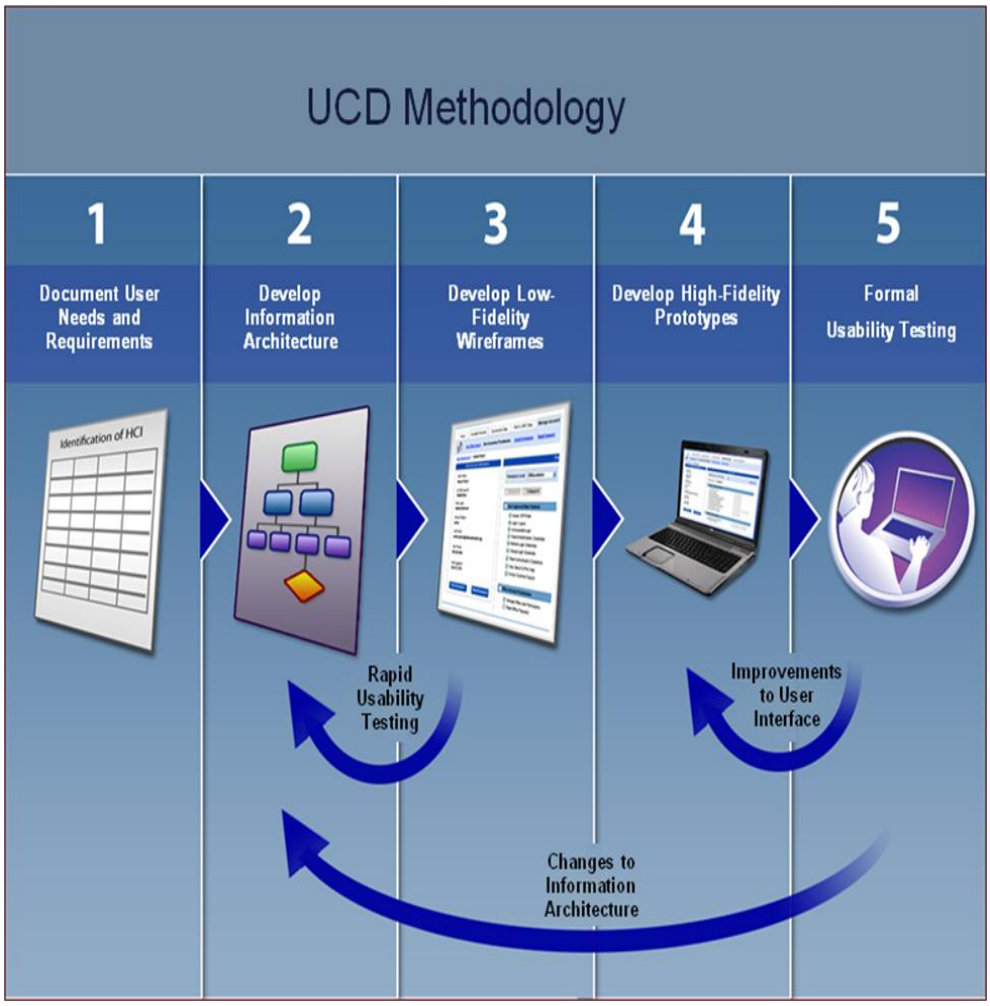


Figure 2. A User Centered Design Methodology

Design of any system begins with understanding and documenting the users' needs and requirements and how they are to be fulfilled (in abstract) by the application (Step 1). For example, the application must be able to display the

history of present illness, show all prior encounter notes, and allow for writing of electronic orders for blood work.

Once these requirements have been documented, the requirements are mapped into a set of workflows (Step 2). Note that because this is user-centered and not system-centered the requirements and process flows are at the user interface layer, not at a system or data layer. To be clear, the user interface is more than what the user sees. The user interface involves the screens (of course), but it also involves the flow across screens, the actions the user needs to take, the alert and error boxes, etc. Defining the process or work flows are critical to the later development of the screen interface. A completed workflow analysis is used to create an information architecture. For instance, one process flow would demonstrate each of the data requirements and data flows (from the users' perspective) for a doctor writing an order for a blood test. The flow would include all starting conditions and assumptions, each step and action along the way, and all feedback the user would receive.

Once the work flows are documented, early prototypes of the design can be developed (Step 3). These early prototypes translate the work flow diagrams and data needs to screen flows and screen images that are typically either hand drawn or low-fidelity images. Users will often have a difficult time reacting to workflow diagrams, but will have no trouble providing feedback on the paper prototypes. The arc in Figure 6.3 between steps 3 and 2 represents informal, rapid usability testing. The outcome of that testing may reveal certain flaws in the workflow analysis creating the need for adjustment. Usability testing at this stage is informal, often quite rapid, and involves few users. The parts of the application that deal with ordering blood work are sketched using a drawing program or even by hand. Potential users (who have had no exposure to the application) would be given the task (e.g., "Using these screens please complete

an order for blood work up for Patient X"). Feedback would then be used to make changes to the architecture and/or the screens.

The iterative design-test cycle continues until enough confidence is gained to make a higher fidelity version – this time one that is executable (Step 4) and ready for formal usability testing (Step 5). The paper 'order screens' would be turned into a fully functioning working prototype. Again, users would perform tasks; however, this time performance data would be recorded and formally documented. Once testing has occurred, the outcome of that testing is evaluated and often changes are required either to the user interface (Step 4) or to the information architecture and design (Step 2). Adjustments are made until the performance criteria are met.

This method of iterative design and testing relies on increasing the fidelity of the prototypes as the design improves, and increasing the number of users involved during each iteration. Iterative testing is foundational to UCD. In the beginning, more informal, formative (i.e., "discount") usability tests are needed to enable rapid turnaround of findings to the development team (e.g., Nielsen, 1989).[30] As the interface matures, more formal, summative methods provide greater coverage of tasks and include more diverse user groups.

7. Unique Features of Developing Usable EHRs

In the broad spectrum of user interface design, one must consider user groups, tasks/goals, and contexts of use.

The characteristics of the user groups make design and the design process of an EHR more challenging than that of other applications. In general, EHR users are highly educated and highly skilled: physicians in general and specialty practice,

[30] Nielsen, J. 1989. Usability Engineering at a Discount. *Designing and Using Human-Computer Interfaces and Knowledge Based Systems*, edited by G. Salvendy and M.J. Smiths. Amsterdam: Elsevier, 391-401.

residents, hospitalists, nurse practitioners, nurses, lab technicians, administrative staff, etc. The design process must embrace the full range of skills, knowledge, and experience that exists today in clinical practice. Specifically, during design and development, efforts must be made to include users from all relevant user populations. When developing EHRs, it is important to ensure a sample of the population that exhibits a range of like characteristics, such as technology sophistication and acceptance of technology. Usability for groups at one end of the technology sophistication dimension often is much different than for those at the other. In short, ensuring that the sampling process covers characteristics of relevant users is paramount to developing a usable EHR. Testing with too few people or narrow populations will not expose a sufficient number of problems.[31]

EHRs should be built with Section 508 accessibility requirements in mind to ensure that diverse user populations can use the application, including those with disabilities (see www.section508.gov). In the UCD process, efforts must be made to characterize the diversity of user groups and ensure that these groups are included during the design phase and during testing.

The tasks of EHR users are highly specialized. Clinical applications have the burden to ensure patient safety, avoid errors, maximize efficiency, offer support at point of care, *and* maintain a legally defensible record. This is a very tall order for any application. What makes this problem harder is that there are significant differences between tasks done in ambulatory settings and tasks done in the hospital setting – not to mention among tasks done in various specialty groups or hospital departments. All this leads to a very complex application environment. What is important in the UCD process is that the tasks used as benchmarks reflect reality. It is important to choose tasks that are high frequency and tasks

[31] See Lewis, J. R. (2006). Sample sizes for usability tests: mostly math, not magic. interactions 13, 6 (Nov. 2006), 29-33. And Turner, C. W., Lewis, J. R., & Nielsen, J. (2006). Determining usability test sample size. In W. Karwowski (ed.), International Encyclopedia of Ergonomics and Human Factors (pp. 3084-3088). Boca Raton, FL: CRC Press.

that are of high importance or have high patient safety implications. Moreover, it is important to test enough of these tasks to ensure that the problems have been uncovered and resolved.

An often neglected area in user interface design is the context of use. Unlike applications that are built for traditional office environments, EHRs are used collaboratively in work groups, on-the-go, and in very stressful situations. Designers' perceptions of context could affect user interface designs tremendously. For example, a designer might make an implicit assumption that a certain screen will be used by a nurse at the nursing station. In fact, the screen might be used by sleep deprived nurses at a slow terminal in the patient room with many distractions. Understanding context is critical to design. Usability testing in context should be strongly considered as an additional step to ensuring usability.

Additionally, context should be seen from the perspective of the user. Many health care professionals are users of technology and many are daily users of EHRs. Designing for people who are current users means that we have to understand their current mental models and methods for performing certain tasks. How one system displays details of flow sheets may be entirely different from how another system does it. These differences can create difficulty for users and increase the time it takes to perform even simple tasks. Complementary or similar user models or patterns in turn facilitate performance. To the extent that users can maximize their existing knowledge by positive transfer of training (that is, the new system behaves and looks like a known system) user performance will benefit. In short, designers should keep in mind that the clinical environment has many different systems and user interface designs should endeavor to minimize initial learning.

Despite the enormous diversity of users, tasks and contexts, the UCD process guides user interface design efforts and provides a foundation for ensuring increased usability.

8. Organizational Maturity for Usability

Every EHR has a user experience, whether the user judges it as good or bad. These experiences are created by a team of people: product managers, designers, programmers, content experts, etc. Good UCD processes are best delivered by organizations that have a culture of usability. The degree to which the process of constructing usable experiences is systematized can be evaluated using a Usability Maturity Model (UMM)[32]. Similar to other maturity models that describe an organization's processes and capabilities in software or product development, UMM can be used to assess and guide an organization in its endeavor to produce usable software.

The purpose of such a model is both diagnostic and prescriptive. Within the Earthy (1999) model there are six levels that describe an organization's embrace of usability and user-centered design. These levels, as shown in Table 5 are cumulative and are characterized by several attributes.

[32] Earthy, J.V.,1998. Usability Maturity Model: Human-Centeredness Scale. IE2016 INUSE Deliverable D5.1.4s.

Earthy, J. "*Usability Maturity Model: Processes*." Project Report, Lloyd's Register of Shipping, London, 1999.

ISO/IEC (2000) 18529 Human-centered Lifecycle Process Descriptions, ISO/IEC TR 18529: 2000 (E).

Level	Description
0 – Incomplete	Not able to carry out process
1 – Performed	Individuals carry out process
2 – Managed	Quality, time and resource requirements for process known and controlled
3 – Established	Process carried out as specified by organization, resources are defined
4 – Predictable	Performance of process within predicted resource and quality limits)
5 – Optimizing	Organization can reliably tailor process to particular requirements

Table 5. Usability Maturity Model (from Earthy, 1999)

Full treatment of usability maturity models is beyond the scope of this document; however, usability maturity within organizations can be characterized by four dimensions. These dimensions include 1) using data as a basis for decisions, 2) management support, 3) the design team, and resources. Each dimension, as described in the following sections, has characteristics that will enable the organization to produce more usable applications.

8.1. Data as a Basis for Decision

Intuition and idiosyncratic experience as sources of design guidelines are rarely sufficient to create outstanding experiences in applications as complex as EHRs. As complexity increases, the need to rely on user performance data provided from representative users in realistic contexts increases. Setting benchmark behavioral objectives based on goals that are user and/or market defined is the ideal. It is easily *claimed* that an EHR is usable. However, such claims can and should be empirically tested given the right assumptions and the correct set of tasks. There is little that defines maturity more than reliance on user performance data to make improvements to the user interface and user experience.

8.2. Management Support

"'Usability," as discussed at the outset, is not a well-defined term. Some organizations do not have a culture of usability. In organizations without a culture of usability there is little more than token thought given to 'ease of use.'

Even those within the organization who “own” the user interface may not appreciate the need for data to drive design decisions and improvements.

Perhaps because of success of an *ad hoc* test or the influence of skilled individuals, the organization matures and some layers of management begin to take usability seriously. Usability activities (e.g., usability tests) then may become more formal and more regular.

At the most mature levels, senior management of the company has a passion for the usability of the application. These senior managers understand the relationship between product quality and usability, as well as the relationship between brand identity and usability. These managers understand that the user interface to the EHR is the lens through which customers make critical purchase judgments about their application; in fact, the UI (User Interface) *is* their product. They also recognize that creating these user experiences requires a commitment to the necessary human resources, and to the UCD process. UCD is tightly woven into product design, development, support, sales, etc. The most mature organizations will have a senior level officer (sometimes called the Chief Experience Officer) who is accountable and responsible for the experiences customers have with the company's products and services.

8.3. Design Team

It is assumed, in some less mature organizations, that programmers write the code and that the user interface is simply part of the code. But user interface design requires an in-depth understanding of the users, tasks, and context of use, which programmers often lack. Understanding this, many organizations rely on the expertise of staff clinical experts or outside clinical advisers to help design their EHR. Clinicians understand the contexts and use cases, and they themselves are users. Engaging clinicians is important and necessary, but it is

often insufficient. Additionally, the number of clinical users is limited, and the way they are engaged (e.g., focus groups) can be inefficient at obtaining good data.

Organizations that best support user-centered design practices round out their design teams by employing experienced experts in human factors engineering and usability. These experts have knowledge of human capabilities with technology combined with expertise in user interface design (i.e., information architecture and wire framing). They also engage visual designers, who augment and extend the information architecture and the screen designs to improve performance. Usability experts and visual designers together with the business, software, and clinical experts are all part of a cross-functional team that lays a foundation for successful design.

8.4. Resources

Commitment to usability requires resources -- human, time, and financial. In the design team dimension, it was pointed out that there are multiple members of the design team necessary to produce quality design. A UCD process does take time; however, the time is usually made up in the development phase. Less mature organizations may be impatient with the time required by the UCD process and feel rushed to start developing due to hard deadlines. But UCD done well produces high quality user interface specifications that actually reduce time and reduce rework because 'bugs' have been worked out before user interface development begins.[33] Finally, mature organizations will dedicate sufficient financial resources toward having the right staff, facilities, and testing budgets to achieve the goal of a quality user experience.

[33] Bias and Mayhew. Cost justifying user-centered design.

Some (e.g., Nielsen, 2006)[34] argue that to go from no focus on usability to a fully mature organization can take many years, and an organization cannot skip stages along the way. Whether this is true or not, understanding where an organization is in its maturity helps to understand the next steps and the rate of change.

[34] Nielsen, J. (May 1, 2006). Corporate Usability Maturity: Stages 5-8. http://www.useit.com/alertbox/process_maturity.html

9. Usability Testing for EHRs

Earlier sections of this document have discussed usability testing without describing specific procedures. This section is intended to provide a primer on usability testing particularly when testing EHRs. This section is not intended to be the only resource used or address every nuance. For more detailed background, the reader can find several books on the subject of usability testing, such as Rubin and Chisnell (2008) and Dumas and Redish (1999). [35]

9.1. What, When and Why of Usability Testing

Usability testing is a core component of user-centered design. The point of doing a usability test is to improve the EHR whether that means its workflow, navigation, screen layout, interaction, navigation, visual design, etc. It is not necessary to test the whole application at once except at the end of the design cycle prior to application launch. In fact, much like in software unit testing, usability tests should take place for portions of the user interface at very early stages. Early testing can be done with low-fidelity wireframes or paper prototypes of the application. During early design the designers should take proposed designs for (say) collecting and recording vital signs and test these. Later, as each portion of the application functionality is developed and tested, these units should be tested together up to the point of doing several use cases for an entire patient encounter. At the very end, usability testing of the whole application is conducted.

---

[35] Dumas, Joseph S., and Janice C. Redish. *A Practical Guide to Usability Testing*. London: Intellect Books, 1999. (ISBN: 1841500208)

Rubin, J. & Chisnell, D. (2008). *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests (Second Edition).* Hoboken, NJ: Wiley.

At a fundamental level, usability testing is simple: representative users do realistic tasks under typical conditions. During the testing various measures of behavior are collected by the usability administrator. These measures and feedback are used to make corrections in the application. This is the ideal. It is obviously more complex than this because there are a lot of details, and rarely are conditions ideal.

One should test early in the design/development process ('formative' testing) and continuously through to the final stages of development ('summative' testing). Table 6 displays the main characteristics of formative and summative testing. These two types of testing should be seen as ends of a continuum. They share the same goal (improve the user interface), but require different techniques and are driven by different measures. Formative testing finds major user interface bugs. It is rapid, iterative, informal, low cost, and qualitative. Often it goes under the title "discount usability".

Formative Testing	Summative Testing
• Earlier and throughout the application life cycle when looking for major, high-level usability issues • Rapid • Diagnostic • Iterative • Used for 'bug' fixes • Qualitative	• Later in application life cycle when 'hard' data are needed • Formal • Deliberative • Used for verification of user performance • Quantitative

Table 6. Characteristics of Formative and Summative Testing

Formative testing can also be done as a diagnostic tool with a fully functioning deployed system. Suppose users at a hospital are having a problem with an installed EHR. Users are complaining of "too many clicks" when placing a medication order. Conducting a usability test with representative users, observing their interactions, and listening to their feedback should reveal the

locus of many of their complaints. Perhaps the problem is in the type of controls that are used (e.g., drop down lists vs. radio buttons) or in the location of controls or poor labeling or excessive alerting. In any event, it takes skilled administrators to observe and then to translate these observations into actionable recommendations to resolve the complaints.

As more of the issues are discovered and corrected through formative testing, more controlled (i.e., summative) studies across broader sections of the user interface, with time and error recording, should be done. Summative testing measures the application against benchmark or baseline performance, competing applications, and/or with the goal of ensuring the application is ready for launch. Summative testing may be necessary when the vendor must provide formal evidence of testing.

An AHRQ EHR study of vendor usability practices and perspectives revealed that many of the vendors interviewed restricted their use of formal usability testing methods to the final design phase (McDonnell et al., 2010). In fact, as discussed above in the Usability Maturity Model, usability testing should be a continuous part of the software development life cycle. Reserving testing until the end can be frustrating and counterproductive. A fully developed system has substantial organizational inertia built in, and the willingness to make substantial change is much lower because the cost is much higher.

Begin testing once the design team has begun to flesh out designs; it is never too early to test with representative users with realistic use cases.[36] Test early and test often.

When approaching usability testing of EHRs, it is important to consider the complexities that are inherent to EHRs. A few complexities are listed below.

[36] Steve Krug writes in 'Don't Make Me Think:' *Testing one user is a 100% better than testing none.* To be clear, this is *testing* not getting opinion or comments.

- First, the users are intelligent, highly trained, busy, healthcare workers. Recruiting these participants is not easy but it is necessary.
- Second, users of EHRs are usually experiencing substantial mental workload when they use the applications.
- Third, EHRs can rarely be tested under realistic conditions. Healthcare workers use EHRs in clinical settings. Most testing is done in a lab or office environment and does not have the atmosphere (or stress) of the real environment. The best guidance for this is to do formative testing in the lab or office environment. For summative testing, best efforts should be made to test in facilities that are similar to the actual setting.
- Fourth, EHRs are complex applications that are often tightly integrated with other systems. Moreover, they are often customized from their 'out of the box' forms. Performance that works well in lab testing may not work as well in field testing.

9.2. Essentials of Testing

This section covers several of the practical matters of usability testing.

9.2.1. Test Planning

The primary concern in testing is to make operational decisions about the objectives of the test, how to test the objectives, what data to record, and what is currently usable in the application and what needs improvement .

As the first activity, test team members should meet to ensure that the objectives, plans, methodology, resources, and timetables are properly aligned. Proper planning is essential. Effort spent early will improve the quality of the study, its results, and might prevent the need for another study. The primary goals of the meeting will be to determine specific objectives for the test and understand the key questions and areas of interest As a practical matter, a lot of test planning is

good project management. If the application is not in the right form or pointing to the right database or the test room is unavailable, the test cannot proceed properly.

Usability administrators create a test plan according to which the study will be conducted. The plan is discussed with the key stakeholders and project team members and, if necessary, revised based on their feedback. In the next section, a test plan is described and annotated.

One of the first and main sections of the test plan outlines the **test objectives**. Objectives are what stakeholders are hoping to learn and what decisions will be made based on the outcomes. Rubin (1994)[37] provides examples of clearly worded user test objectives, such as:

- Can end users perform common tasks within established benchmarks (certain amount of time, errors, etc.)?
- Does the application contain major usability flaws that prevent completion of the most common tasks?
- Is the new release harder to use than its predecessor?
- Is the response time a cause of user frustration or errors?
- Do the screens reflect the end user's conceptual model?

During planning, the project lead must also assess available resources and capabilities, assign responsibilities, and develop a timeline for activities.

Resources that must be decided upon include:

- Budget: What financial resources are required for this study?
- Team members: Who has the skills and availability to be on the project?
    - The research may include a project lead, administrator (moderator), note taker, analyst, and report writer as a few examples.
    - One person may take on several different roles.
- Location: Will you have access to a usability lab to conduct testing or will the test occur on-location?

[37] Rubin, J. (1994). Handbook of usability testing: How to plan, design, and conduct effective tests. New York: John Wiley.

- Location is dependent on several factors such as accessibility of the team to potential participants, budget, time, and test application mobility (in other words, is the interface only available within a certain environment).

Project planning will include the creation of a timeline that will include major activities, when they occur, and who will perform them. The timeline is a valuable tool for project management though it will likely need editing and updating throughout the study.

9.2.2. Sample Test Plan

The test plan document should be comprehensive and concise. In general, usability test plans should include the following sections:[38]

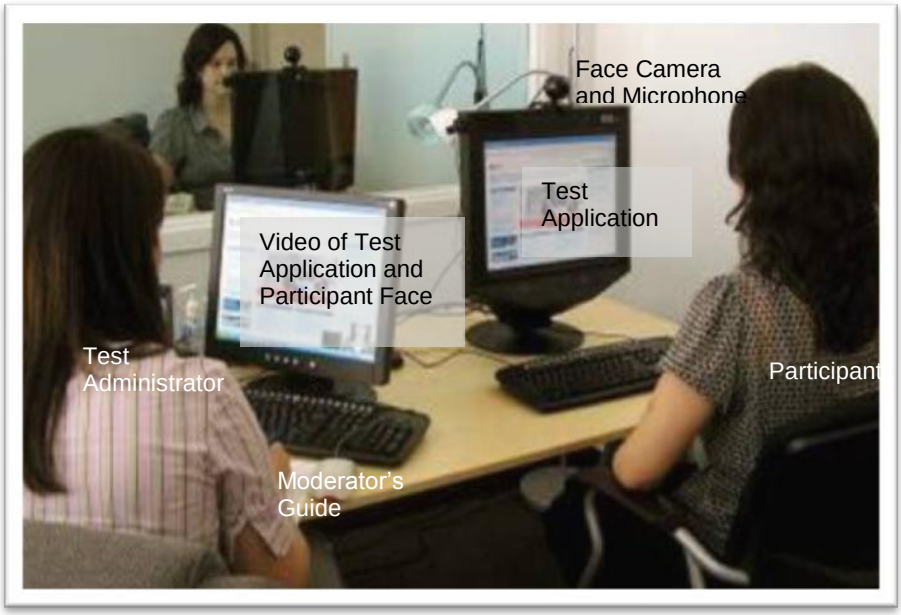
1. Test Objectives
    - List the specific questions that the study has been designed to answer.
    - For example:
        - *Objective 1: To evaluate sample EHR application in terms of the user experience that it creates. Specifically:*
            - *Do users understand the navigation of the EHR and how the information is structured?*
            - *Can users complete key tasks (e.g., use patient chart to find lab result) or do they require assistance?*
            - *Do users understand the content in the EHR?*
            - *Do users feel that the content meets their needs?*
        - *Objective 2: To identify the EHR's key strengths and assess how it could be further improved to better meet the needs and expectations of users:*
            - *Identify areas for improvement*
            - *Prioritize these areas*
            - *Provide actionable recommendations*
2. Test Application

[38] See Bojko's chapter on preparing for usability testing (Chapter 3) Schumacher, R. (Ed) (2009), Handbook of Global User Research. Burlington, MA: Morgan Kaufman.

- Describe the application including version number where appropriate.
- Also include minimum requirements for computers that need to host the application, if necessary.
- For example:
  - *The application to be tested is sample EHR application, version 1.0.*
  - *The test application may be run on laptop personal computer running a standard Internet browser.*

3. Performance and Satisfaction Metrics
   - Describe the types of data that will be collected and analyzed.
   - For example:
     - *Qualitative measures:*
       - *Usability issues observed*
       - *User comments*
     - *Quantitative measures:*
       - *Success/failure for task completion*
       - *Ease of use ratings for each task*
       - *Overall ease of use rating for the application*
       - *Task time*
4. Method
   - Describe the methodology, including the participants, study design, tasks, and procedure.
   - For example:
     - *Participants*
       - *14 participants will be recruited*
       - *Approximately half physicians, half nurses*
       - *Each have at least 1 year of experience using EHRs*
     - *Tasks*
       - *Find information in Patient Summary screen*
       - *Use patient chart to find lab results*
       - *Check vital signs*
       - *Look for interactions and allergies*
       - *Add notes to patient chart*

        - *Request lab*
    - *Procedure*
        - *Introduction*
        - *Warm-up Questions*
        - *Tasks*
        - *Ratings Questions*
- Test Environment / Equipment
    - Provide pictures or sketches of the lab setup, denoting where the administrator and participant should sit, and where the cameras should be set up.
    - For example:



- Analysis
    - Describe any analyses to be performed on the collected data.
- Timeline
    - The project schedule, including test and deliverable dates

9.2.3.Test Metrics

Human behavior can be reliably measured. In general, usability metrics are observable, quantifiable (even if it is simply a count of the number of times

participants do a particular action), and represent some aspect of the user experience. Usability testing measures three main components of usability[39]:

- Effectiveness (are the participants able to complete a task error-free?)
- Efficiency (how much time is required to complete a task?)
- Satisfaction (what is the degree to which participants perceive the application to be usable?)

A number of specific measures can be chosen to assess an application's effectiveness, efficiency, and satisfaction. The following list includes sample measures grouped into four categories: performance metrics, issues-based metrics, self reported metrics, and behavioral metrics. Measures should be chosen based on the test objectives and type of testing. For example, summative testing often requires several quantitative measures of participant performance such as task success, time on task, and number of errors. Formative tests, however, may benefit from open-ended responses, ease of use ratings, and issues-based metrics such as the number of usability issues identified by participants.

In general, measures from several of the following categories may be necessary to assess effectiveness, efficiency, and satisfaction:

Performance metrics (quantitative measures, often used for summative testing)

- Task success (whether the participant can complete the task in the allotted time without assistance from the administrator)
- Time on task (the length of time in seconds required for participants to complete the task)
- Errors (may include clicking on an incorrect menu item, incorrect link, or interacting incorrectly with an on-screen control)
- Efficiency (e.g. task time, and errors)
- Learnability (metrics above taken from trials within same session, with breaks between sessions)

[39] Tullis, T., & Albert, W. (2008). Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics. Morgan Kaufmann.

Issues-based metrics (primarily used for formative testing)

- Usability issues (# issues found, % of participants who found an issue)
- Severity ratings (rating assigned to usability issues that reflects the impact of each issue on the user's satisfaction and ability to complete tasks)

Self-reported metrics (both quantitative and qualitative measures which provide insights about participant satisfaction)

- Post-task ratings (may be a Likert scale from 1 to 5 where, for example, 1 is very difficult and 5 is very easy).
- Post-session ratings (may be a Likert scale from 1 to 5 where, for example, 1 is very dissatisfied and 5 is very satisfied).
- System Usability Scale (Participants rate statements on a five-point scale where "1 - Strongly Disagree" to "5 - Strongly Agree." The SUS survey yields a single number that represents a composite measure of the overall usability of the system).
- NASA TLX (subjective workload assessment tool that asks users to rate how demanding particular tasks are to perform).
- Specific attribute questions (e.g,. Rate: Strongly Agree to Strongly Disagree for the question "Overall, I am satisfied with how well this application supported this task").
- Semantic differentials (e.g., "This task was "Easy" "Difficult"").
- Answers to open-ended questions (For example, "What did you find to be the most difficult or frustrating aspect of this application?").

Behavioral metrics (often add context to performance, issues-based, and self-report metrics)

- Verbal (positive / negative) comments and non-verbal behavior. For example, a long pause or a confused look might indicate lack of understanding.

Table 7 details measures commonly used in usability testing to evaluate effectiveness, efficiency, and satisfaction:[40]

Measures	Rationale and Scoring
Effectiveness: Task Success	A task is counted as a "Success" if the participant was able to achieve the correct outcome, without assistance, within the time allotted on a

[40] From: Customized Common Industry Format Template for Electronic Health Record Usability Testing

	per task basis. The total number of successes are calculated for each task and then divided by the total number of times that task was attempted. The results are provided as a percentage.
Effectiveness: Task Failures	If the participant abandons the task, does not reach the correct answer or performs it incorrectly, or reaches the end of the allotted time before successful completion, the task is counted as a "Failure." No task times are taken for errors. The total number of errors is calculated for each task and then divided by the total number of times that task was attempted. Not all deviations would be counted as errors.[41] This should also be expressed as the mean number of failed tasks per participant. On a qualitative level, an enumeration of errors and error types should be collected.
Efficiency: Task Deviations	The participant's path (i.e., steps) through the application is recorded. Deviations occur if the participant, for example, visits an incorrect screen, clicks on an incorrect menu item, follows an incorrect link, or interacts incorrectly with an on-screen control. This path is compared to the optimal path. The number of steps in the observed path is divided by the number of optimal steps to provide a ratio of path deviation. Deviations do not necessarily mean failure – simply a less efficient method through the interface. It is strongly recommended that task deviations be reported. Optimal paths (i.e., procedural steps) should be recorded when constructing tasks.
Efficiency: Task Time	Each task is timed from when the administrator says "Begin" until the participant says "Done." If he or she fails to say "Done," the time is stopped when the participant stopped performing the task. Only task times for tasks that are successfully completed are included in the average task time analysis. Average time per task is calculated for each task. Variance measures (standard deviation and standard error) are also calculated. Task times are recorded for successes. Observed task times divided by the optimal time for each task is a measure of optimal efficiency. Optimal task performance time, as benchmarked by expert performance under realistic conditions, is recorded when constructing tasks. Target task times used for task times in the Moderator's Guide must be operationally defined by taking multiple measures of optimal performance and multiplying by some factor (e.g., 1.25) that allows some time buffer because the participants are presumably not trained to expert performance. Thus, if expert, optimal performance on a task was 100 seconds then allotted task time performance would be 125 seconds. This ratio should be aggregated across tasks and reported with mean and variance scores.
Satisfaction: Task Rating	Participant's subjective impression of the ease of use of the application is measured by administering both a simple post-task question as well as a post-session questionnaire. After each task, the

[41] Errors have to be operationally defined by the test team prior to testing.

	participant is asked to rate "Overall, this task was:" on a scale of 1 (Very Difficult) to 5 (Very Easy). These data are averaged across participants. [42] Common convention is that average ratings for systems judged easy to use should be 3.3 or above. To measure participants' confidence in and likeability of the application overall, the testing team can administer the System Usability Scale (SUS) post-test questionnaire. Questions include, "I think I would like to use this system frequently," "I thought the system was easy to use," and "I would imagine that most people would learn to use this system very quickly."

Table 7. Details of how observed data will be scored.

9.2.4.Test Application

The test application itself may represent various stages of completeness, from paper mock-ups, to semi-functional prototypes, to full designs. Full designs may better approximate the full functionality and task flows of the interface; however prototypes can be valuable for testing. Prototypes can be static, paper-based, or interactive. These are often used to incorporate user feedback into the design early in the process. They may also allow exploration of a few concepts before choosing one. Prototypes provide some visual reference for participants and make the interface more concrete and tangible.

It is important for usability administrators to review the application thoroughly and become familiar with the interface, understanding its capabilities and constraints. There may be additional components of the application that must be explored such as the packaging, instruction manual, accompanying materials, and application training.

9.2.5.Selecting Participants

[42] See Tedesco and Tullis (2006) for a comparison of post-task ratings for usability tests. Tedesco, D. & Tullis, T. (2006) A comparison of methods for eliciting post-task subjective ratings in usability testing. *Usability Professionals association Conference*, June 12 – 16, Broomfield, CO.

Based on the target user groups, the usability administrator can create a Screening Questionnaire for recruiting study participants. The questionnaire will identify individuals who meet defined criteria (e.g., hospitalists). The Screening Questionnaire also defines targeted breakdowns in terms of demographics and/or user profiles (e.g., age, gender, income, etc.). Individuals are recruited based on their responses. Rubin[43] provides a comprehensive list of guidelines for developing screening questionnaires, and, in addition to his notes, a good screening questionnaire is:

- Clear for the test team, recruiters, and respondents
    - Target criteria, quotas, and termination points should be easy to identify.
        - Summarize this information at the beginning of the document to provide a snapshot of what is required.
        - Clearly identify which criteria are must-have priorities, versus nice-to-have characteristics.
        - For each question, restate applicable termination criteria and quotas.
        - When multiple profiles are being recruited, ensure that they are mutually exclusive and question branches are easy to follow.
    - Questions should facilitate quick responses, requiring minimal thought.
        - The more prospective participants "think about it," the more likely they are to over think and not give an honest response.
        - Respondents should not have to ask "what do you mean?"
        - Multiple-choice responses expedite analysis and should have clear, mutually exclusive options.
        - Open-ended questions are sometimes necessary to elicit honest responses and assess the articulateness of the potential participant.
        - Avoid leading questions, which allow the respondent to determine the desired response and respond accordingly.

[43] Rubin, J. (1994). Handbook of usability testing: How to plan, design, and conduct effective tests. New York: John Wiley

        - Yes/no questions are particularly obvious and should be avoided unless absolutely necessary.
        - Where appropriate, add distracter responses to the multiple choice set (e.g., medical specialties).
        - Ask open-ended questions (e.g., have respondents state the activities they perform rather than having the recruiter read a list).
    - Overall, questions should be easy enough that respondents can accurately remember details required to answer.
- Not overly restrictive
    - Overly restrictive screeners require more recruitment time and are less likely to fulfill quotas.
    - To reduce recruiting costs, strive for the least restrictive screener that will get the respondents needed to satisfy test objectives.
    - Targeting users of specific features or devices are challenging recruits, especially if adoption rates are low.
    - Probable incidence rates should be considered when creating multiple choice responses. Ranges (e.g., frequency of use) should not be too narrowly defined.
- Efficient
    - Allow sufficient time to recruit the appropriate user groups (i.e., difficult groups take longer to recruit); two weeks is sufficient for most groups.
    - Ask the minimum number of questions to effectively screen respondents.
        - Aim for a maximum of 10 questions that are relevant to the EHR clinical context.
    - Obtain participants' home, work, and (especially) mobile phone numbers so they may be reached on the day of testing and prior.
    - Test day logistics should be outlined and provided to qualified respondents.
        - Provide the facility address, phone number, and driving directions.
        - Require a state-issued ID.
        - If respondents need corrective lenses, require that they bring them to the session.

        - If applicable, state that respondents will be required to sign a consent form and that their session will be videotaped.

Poorly constructed screeners pose several risks. Most egregious is that participants may not be representative of actual users. As a result, the data will likely be compromised, incomplete, and misleading. Participation rates may also be adversely affected by ineffective screeners, which make the recruitment difficult to fill and fail to set appropriate participant expectations.

How many participants are needed in the study? The goal is to learn what is necessary while using the minimum number of participants. How many to test depends on:

- What questions are being answered?
    - Are you looking to discover critical usability issues prior to further development of the design? (fewer participants)
    - Are you looking definitively to know whether the application is ready for launch? (more participants)
    - Are you looking to generalize the results to the population? (more participants)
- Is it a formative (requires fewer participants) or summative (requires more participants) test?
    - Formative tests tend to be small and iterative. The goal is to find bugs, fix them, and test again.
    - Summative tests are larger and look for statistical significance. Testing against benchmarks or previous test results typically involve this level of analysis.
    - Free online sample size calculator from Creative Research Systems: http://www.surveysystem.com/sscalc.htm
- Are there multiple user groups? (e.g., physicians and nurses)
    - Will they be doing the same or different tasks?
- Are there too many tasks for one person to complete in a session?
    - As application and task complexity increase, so does the amount of session time required.
    - The number of planned and ad-hoc probes also adds to session time.

- Divide tasks between participants. (Tasks 1-10 with half of the participants, Tasks 11-20 with the other half).

Many other factors may influence the sample size, such as timelines, budget, access to users, etc. In general, a few participants are better than no participants.

9.2.6. Usability Testing Script / Moderator's Guide

The usability administrator creates a moderator's guide with collaboration from the design/test team and formulates specific sets of questions and tasks that will be used during the usability test sessions. The moderator's guide can be a semi-structured interview script to aid in task administration and data collection.

The tasks to be included in the moderator's guide may have been defined during project meetings following a task analysis. The author of the moderator's guide must understand the workflow or process flow that users undertake with the test application. Functions are then selected for usability testing according to several criteria: frequency of use, task criticality and complexity. They may also be selected based on user feedback, difficult design areas, risk and liability, effects on revenue, and compliance issues.

A good moderator's guide includes tasks and questions that are in perfect alignment with the test objectives.[44] Objectives without tasks/questions can lead to a lack of necessary data. Conversely, tasks/questions not addressing any objective are a waste of time and create unnecessary work when recording and analyzing data.

9.2.7. Running the Usability Test

The technical setup of a test can range from very high- to very low-tech. There are two main elements that must be included in test setup: (a) a platform from

[44] Bojko, A. (2009) "Chapter 3: Preparation" in Schumacher, R. (Ed), Handbook of Global User Research. Burlington, MA: Morgan Kaufman.

which participants can access and experience the interface, and (b) means for the administrator, notetaker(s), and/or observer(s) to observe the participant actions, behaviors, and comments.

For EHR and other computer-based application testing, the equipment will likely consist of:

- One computer or laptop, capable of running the test application, for the participant to use.
- Some way for the administrator to view the participant's computer.
    - Typically, a second computer or some other means for projection (monitor, TV, projector) can show the participant's screen.
    - The test team and any stakeholders are able to see exactly what the participant is viewing and can observe the participant's actions.
    - Software is available that can record the participant's screen along with other camera shots of interest.
    - Direct observation through a one-way mirror is also possible in certain facilities. However, one-way mirrors have limited benefits when testing computer-based applications.
- Often, a webcam/ small video camera is used to gather an audio and video recording of participant reactions and feedback.
    - Participant verbal commentary and facial expressions can provide many insights and data about the test application.
- Note-taking tools may be electronic (e.g., structured spreadsheet or online survey collector), or hand-written (e.g., in the available spaces in the moderator's guide or on notes stuck to an enlarged screen print).

A pilot test of the study evaluates the readiness of the technical setup and the effectiveness of the moderator's guide. The ideal pilot participant is someone outside the project team who, if possible, fits the participant profile. The following aspects of the test can be evaluated during the pilot:

- Do the instructions flow well and make sense to the pilot participant?
- Is there enough time in the session to complete the necessary tasks and gather the necessary data?
    - If certain tasks are taking longer than they should and the session is going to go longer than the allotted time, the stakeholders may need to reassess the priority of the tasks.

    - Tasks most important for achieving objectives may be placed toward the beginning of the session. Less important tasks should be placed at the end and will be completed if time allows, or be cut.
- Is the administrator and/or notetaker able to record data sufficiently with the note taking tools that were created?
- Do the test recordings clearly reproduce the participant's actions and verbal comments?

The technical setup and moderator's guide will likely need to be changed based on the results of the pilot study. Once the pilot session runs as intended, real participants may be tested.

Usability tests often consist of the following activities performed by the administrator:

- Provide a brief introduction about the study, advise the participant of his or her rights and asks the participant to read and sign an informed consent form.
- Ask background questions about the participant to both put the participant at ease and verify the recruit criteria.
- Administer tasks to be completed using the EHR.
- Note participant's behaviors and comments.
    - Observe problems and strengths of the EHR, keep a list.
- Record any data such as task time, errors, navigational path, and success.
- Gather qualitative feedback from an in-depth interview.

Moderation and interviewing skills are critical to the success of user research. A good administrator:

- Is expert in user centered design and research
    - The administrator must be familiar with testing protocols and the research and business objectives specific to the current study.
- Builds a rapport with participants and is a good communicator
    - A good administrator is a warm, empathic "people person" who actively listens to participants.

    - Long and repetitive test sessions require an administrator with a long attention span, stamina, and genuine interest in each participant.
- Avoids biasing test participants
    - A good administrator is not the designer of the interface.
    - The administrator should not ask leading questions or encourage either positive or negative responses.
        - Avoid using adjectives or adverbs ("Was that task easy?" "Could you do that easily?" vs. "How was that task?" "Did you find it easy or difficult?")
- Is adaptable
    - Moderators must quickly learn new interfaces and innovate methodologies to meet specific research objectives.
    - Flexibility and comfort with ambiguity are essential skills when dealing with rapidly changing interfaces and a wide variety of participants.
    - While mired in the details of each session, a good administrator must be able to step back to see the "big picture" as trends emerge.

9.2.8. Doing the analysis

Depending on the methodology and types of data gathered, analyses vary. For studies involving quantitative data such as time, errors, and success, descriptive statistics or other statistical analyses may be performed. Descriptive statistics commonly used in usability testing include the mean (e.g., mean task time), counts (e.g., number of times a usability issue is mentioned or encountered), and percentages (e.g., percentage of successful outcomes for a particular task). Qualitative data, such as participant comments and actions, may be summarized, tabulated, or analyzed using content analysis to look for trends. Findings can then be organized by scope and severity.

If possible, track trends and unexpected findings encountered during testing. It is important to maintain an open mind to results that occur, but organizing the results can save time and effort in the analysis and reporting phase. Thinking

about results helps to ensure that findings are meeting objectives and will also help to prevent surprises once the results are delivered.

The test administrator can examine these quantitative results as well as qualitative ones to assess and form conclusions about the overall usability of the application.

9.2.9. Reporting Findings: The Customized Common Industry Format (CIF) for EHRs

Depending on the needs of the project, user test results can be reported in many ways. The audience for the findings and recommendations should determine the form that these take. In general, methods and results should be presented with enough detail that the study can be reproduced and results can be compared. The recommended format is the companion document to this: *Customized Common Industry Format Template for Electronic Health Record Usability Testing.*[45] The Customized CIF for EHRs provides a template for reporting EHR usability test results conforming to this document.

The results of a usability test should also be used to create recommendations for any design changes. Recommendations are most beneficial when they clearly illustrate actionable changes that should be made to the interface based on specific participant performance. Once the design team implements recommendations, the interface will be ready for its next round of iterative testing.

10. Appendix A: Guide to Completing the Customized Common Industry Format Template for Electronic Health Record Usability Testing

See companion document NISTIR 7742.

[45] Derived from *ISO/IEC 25062, Software engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Common Industry Format (CIF) for usability test reports*.

The modifications here allow for the reporting of qualitative findings (i.e., formative) but strongly recommend and encourage the collection of quantitative measures of user performance.