EHR Usability Toolkit: A Background Report on Usability and Electronic Health Records

Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • www.ahrq.gov

HEALTH IT

EHR Usability Toolkit: A Background Report on Usability and Electronic Health Records

Authors:
Constance M. Johnson, Ph.D.
Douglas Johnston, MTS
P. Kenyon Crowley, M.B.A., M.S., C.P.H.I.M.S.
Helen Culbertson, M.B.A.
Helga E. Rippen, M.D., Ph.D., M.P.H.
David J. Damico, M.A., C.H.F.P.
Catherine Plaisant, Ph.D.

Task Order Officer:

Teresa Zayas-Cabán, AHRQ

Suggested Citation:

Contents

Tables

Appendixes

Executive Summary

The reports funded by the Agency for Healthcare Research and Quality (AHRQ) and developed under the "Use of Dense Display of Data and Principles in Primary Care Health Information Technology (IT) Systems" project identified key shortcomings in the usability of certified electronic health record (EHR) products and the ways that health care organizations implement them.[1,2] Notably, these shortcomings included a lack of standard practices, best-practice sharing, and systematic processes. Key recommendations from these reports include establishment of usability as a core aspect of EHR certification and development of objective criteria that reflect best practices in EHR usability.

To begin exploration of improving EHR usability, AHRQ contracted with Westat to develop, test, refine, and disseminate a Web-based toolkit for primary care providers to assess how well the usability of health IT systems support them in delivering safe, effective, and efficient care. This toolkit would also provide objective yet practical means to assess whether the usability of EHRs support such primary care functions as proper diagnosis, identification of high-risk patients, tracking of health parameters over time, and population health management. Finally, the toolkit should assess how well EHRs adhere to critical "best practice" usability principles as well as help to identify flaws in usability that might adversely impact the quality of care. Thus, three important objectives of this project are the following:

1. Develop and refine a toolkit for primary care providers that supports rapid yet meaningful usability evaluations of EHRs.

2. Disseminate the toolkit as a means to increase attention on the importance of EHR usability, to promote use of evidence-based usability evaluation methods, and to stimulate collaboration among entities developing and implementing EHR systems.

3. Inform EHR accreditation efforts on usability, including those by the National Institute of Standards and Technology.

To support these objectives, the project identified and convened a panel of experts from the fields of usability testing and assessment, biomedical and clinical informatics, human factors, cognitive psychology, human-computer interaction, tool development, and end-users of EHRs. Members of the expert panel (listed in Appendix A of this report) were called upon for their experience with existing toolkits and resources in the area and with measures and metrics for measuring usability of EHR systems, and to provide feedback on the design, testing, and dissemination of the toolkit.

This report provides the basis for recommendations that will guide the development of an initial Web-based toolkit that supports primary care providers by identifying usability issues in their current EHRs. To ensure our recommendations are based on the most current evidence, we have reviewed and analyzed the literature on usability issues in health care information systems and on usability evaluation methods.

The toolkit will reflect the complex attributes of primary care settings—such as diversity of EHR users, complexity of shared tasks, differences in clinical encounters, and the variety of patients and conditions. The toolkit will also need to address usability issues often encountered by primary care EHR users including challenges in information processing, workflow interruption, and increases in cognitive burden stemming from poorly designed interfaces and EHR functions. Central to our approach, then, is ensuring that the attributes of primary care settings and the evidence of usability issues in EHRs inform selection of usability evaluation methods for inclusion in the Web-based toolkit.

The authors searched the literature to identify the range of usability evaluation methods that could be used to identify problems in EHR systems. These methods include heuristic evaluation, cognitive walkthrough, remote evaluation, laboratory testing, and usability questionnaires. In addition, the authors reviewed other human factors methods that can be used to evaluate information systems. These are predictive modeling, risk assessment, and workload evaluation methods. These human factors methods are included because these methods can help to identify usability problems within a system. Based on the scope of the project, guidance from AHRQ and the Technical Expert Panel, and factors identified through consideration of EHR capabilities, usability problems, and attributes of primary care settings, high-level criteria for selecting methods for inclusion in the initial toolkit were developed. Two important criteria include the efficiency of a particular method (i.e., how easy it is to use and how quickly it can be applied) and the ability for primary care practices to administer a method independently from human factors experts. When assessing the advantages and disadvantages of existing usability evaluation methods with these and other criteria, it was determined that usability questionnaires would be the most practical, rapid, and useful methods for the initial toolkit.

To meet the goals of this project, the report recommends the development of a Web-based toolkit consisting of three main components: assessment, education, and tailoring. The assessment component will consist of core usability questions most relevant to EHRs. The educational component will provide a description of the usability problem and why it is relevant to an EHR. It may also include links to resources to help toolkit users understand usability and a summary sheet that could be shared with vendors. The tailoring component will provide a summary that providers can share with vendors and EHR certifying bodies to complement their user-centered design programs.

Chapter 1. Introduction

Background

The current health care system in the United States delivers suboptimal quality[3] and lacks a robust health care information infrastructure. The Committee on Quality of Health Care in America concluded that information technology (IT), through the widespread automation of clinical records, must perform a pivotal role in the redesign of the U.S. health care system.[3] To move forward on providing this infrastructure, the Federal Government supported the national adoption of electronic health records (EHRs) because of the promise that the adoption and use of health IT can increase health system efficiency, improve care quality, and reduce medical errors. Most recently, the Health Information Technology for Economic and Clinical Health (HITECH) Act, enacted as part of the American Recovery and Reinvestment Act of 2009, provides financial incentives and technical assistance to support health care providers in implementing and using EHRs and beginning in 2015 will penalize those who do not—as a means to achieve widespread adoption of health IT. Although health IT offers the potential to change the way health care is structured and delivered, the success of universal implementation will depend heavily on how well these systems support clinicians and other stakeholders in performing their work. EHR usability and accessibility are crucial to effective health IT adoption and to realizing value from these systems.

Primary care practices are a key component of the U.S. health care system, providing acute, chronic, and preventive care services. Clinicians in these practices continually process complex data, information, and knowledge to support a range of activities across diagnosis, care planning, treatment, and health management. As such, primary care practices are an important area of focus for national EHR adoption efforts. Although EHRs offer the potential to provide complete, accurate, and timely data and decision support, these systems have not been adopted widely in primary care settings.[4] In addition to financial, organizational, and technological barriers,[5] poor user interfaces (UIs) have been cited as a major obstacle to the acceptance and use of health care information systems.[6-9]

Deficiencies in user interface design primarily stem from a lack of understanding of the cognitive needs of the clinicians, and from a failure to fully account for common problems in human-computer interaction.[8] Information displays that do not match the user's workflow or mental models, or do not appropriately represent data and information, can lead to inefficient care such as missing information important to diagnoses or ordering of unnecessary tests.[8,10,11] To assist with clinical reasoning, problem solving, and decisionmaking, primary care clinicians need a succinct representation of complex clinical data. Health IT applications—especially EHRs—must be designed, developed, and evaluated with serious consideration of the characteristics of the users, their tasks, and their environments. Yet, systems that discount important user characteristics, workflows, tasks, cognition, preferences, and usability issues continue to be designed and implemented, resulting in systems that are inefficient, ineffective, or unusable,[8,12,13] and that may become independent sources of error.[14-16]

The application of well-documented usability principles and guidelines to EHR interface design and evaluation will help create systems that clinicians can use easily, safely, accurately, and efficiently. While the importance of human factors issues such as usability is widely recognized,[17-21] the application of usability principles and guidelines remains limited. Recent public and private sector initiatives have produced resources to inform the design and assessment of usable EHR interfaces, but there are no universally accepted software tools that support usability problem reporting and analysis.[22]

Project Purpose

Funded by the Agency for Healthcare Research and Quality (AHRQ), the *EHR Usability Toolkit* project focuses on developing, testing, refining, and disseminating practical tools for assessing EHR usability. The goal of this toolkit is to provide a practical and useful way to assess whether the usability of EHRs supports such primary care functions as proper diagnosis, identification of high-risk patients, tracking of patient health parameters over time, and population health management. In addition, this toolkit seeks to assess how well system implementations adhere to critical "best practice" usability principles as well as to identify flaws in usability that might adversely impact the quality or safety of care. A panel of experts in cognitive science, health IT, primary care, and human factors engineering has been created to provide guidance throughout project phases, including the development of this report and toolkit design, testing, and dissemination. A list of the expert panelists who are serving on the Technical Expert Panel can be found in Appendix A.

Building on prior AHRQ work,[1,2] and referencing other public[23] and private[24] sector usability initiatives, this project will rely upon the best available evidence to produce an effective toolkit to support primary care providers in assessing the usability of their current EHRs. The intended users of this toolkit are primary care providers (including physicians, physician assistants, and nurse practitioners in the primary care setting taking care of patients, whose goals and tasks may vary depending on their role and responsibilities. Although nonclinicians may use the system, the focus is narrowed to ensure adequate attention to the main users of the EHR in the primary care setting, the providers. Moreover, the toolkit may also be valuable to vendors, health IT organizations, and EHR certification organizations to test systems and as a guideline for the types of tests they could potentially use to test their systems.

Toolkit Definition and Goals

The EHR usability toolkit is being constructed with multiple objectives in mind. First it is envisioned to be a collection of evaluation tool(s), feedback mechanisms, and educational content developed to accomplish the following goals: (1) allow the toolkit user (clinician) to appraise how easily and effectively the EHR system supports key clinical tasks, (2) evaluate how well usability of the EHR system supports workflow, (3) assess how well the EHR system adheres to "best practice" usability principles, and (4) identify flaws in information and interface design that are likely to constrain use and/or adversely impact the quality or safety of care.

A secondary goal of the toolkit is to provide usability evaluation methods for those developing ambulatory care EHRs. By assessing their systems with the methods provided in the toolkit, vendors can identify global usability issues within their EHRs.

Finally, the toolkit may inform a methodology and framework that will serve as a core component of National Institute of Standards and Technology's (NIST) program to accredit EHR system usability.

Report Organization

This background report provides the basis for recommendations and requirements for developing an EHR usability toolkit. The report summarizes findings from a literature review on general usability issues in information systems, specific usability issues in EHRs, and usability evaluation methods. By comparing common usability issues against existing usability evaluation methods, this report provides recommendations for which methods should be included in the toolkit.

This report is organized as follows. Chapter 2 defines EHR systems, their core functions, and discusses usability factors important to effective use of information systems. Chapter 3 reviews primary care settings. Chapter 4 reviews usability issues in EHRs. Chapter 5 provides a review of the methods and tools to measure usability, while Chapter 6 discusses which of these methods are most appropriate for inclusion in the toolkit. Chapter 7 reviews and assesses specific instruments associated with the methods identified in Chapter 6, and conclusions are presented in Chapter 8.

Chapter 2. Defining EHRs and Usability

To support the review of usability problems and evaluation methods, this chapter defines EHR systems, their core capabilities, and the roles these systems serve with clinicians. It also defines usability and how usability can be operationally measured through its core attributes: useful, usable, and satisfying.

EHRs Defined

Computerized patient records have been characterized in many ways, most commonly as either electronic medical records (EMRs) or electronic health records (EHRs). The primary difference between the two is interoperability, or the ability for systems to share clinical data and functions electronically across organizations. EMRs are used within one organization whereas EHRs are used across more than one health care organization.[25,26] For the purposes of this report, we will use EHR to encompass all electronic patient records since the U.S. health care system is moving towards interoperability of all records.

The definition of the EHR has varied over time due to the complexity of summarizing all of the numerous and diverse components inclusively. For example, the International Organization for Standardization (ISO) definition of the EHR is "a repository of information regarding the health of a subject of care, in a computer processable form."[27] However, the Healthcare Information and Management Systems Society's (HIMSS) definition is more inclusive and comprehensive. It defines an EHR as "a longitudinal electronic record of patient health information generated by one or more encounters in any care delivery setting. Included in this information are patient demographics, progress notes, problems, medications, vital signs, past medical history, immunizations, laboratory data and radiology reports. The EHR automates and streamlines the clinician's workflow. The EHR has the ability to generate a complete record of a clinical patient encounter—as well as support other care-related activities directly or indirectly via interface—including evidence-based decision support, quality management, and outcomes reporting." [28]

Many health care professionals—physicians, nurses, radiologists, pharmacists, laboratory technicians, and radiographers—use various components of EHRs. Additionally, EHR data may also be used by patients through patient portals, or linked to their personal health record (PHR).[29] EHRs are used in numerous care settings such as hospitals, ambulatory care, nursing homes, community care, home health agencies, dental care, and pharmacies; however, for the purposes of this report, we are focusing on ambulatory care settings, specifically primary care.

EHRs are, in many ways, collections of applications and features that support different users across a range of clinical and administrative tasks. To help assess the range of usability issues associated with EHRs, it is important to understand these systems both in terms of their *capabilities*—the range and types of EHR features—as well as their *uses*—the role that EHRs serve in supporting clinical and administrative tasks. For this project, we define EHRs as discussed below.

EHR Functionalities and Uses

The Institute of Medicine identified eight key functionalities of an EHR for the care and management of patients.[30]

1. **Health information and data management.** The EHR may record and display a wide variety of health data, which may be in structured (coded) or unstructured form. Key types of data include patient demographics, problems, medications, allergies, test results, lists of procedures, clinical notes, and treatment plans.

2. **Order entry management.** Computerized provider order entry (CPOE) enables care providers to order medications, laboratory and radiology tests, procedures, and other order types electronically.

3. **Results management.** When interfaced with ancillary systems, EHRs may enable electronic access to results of laboratory, radiology, microbiology, and pathology tests as well as findings from consults.

4. **Clinical decision support (CDS).** The report on *A Roadmap for National Action on Clinical Decision Support* defines CDS as providing "clinicians, staff, patients, or other individuals with knowledge and person-specific information, intelligently filtered or presented at appropriate times, to enhance health and health care. It encompasses a variety of tools and interventions such as computerized alerts and reminders, clinical guidelines, order sets, patient data reports and dashboards, documentation templates, diagnostic support, and clinical workflow tools." [31] The most common type of CDS tools are alerts and reminders, which include warnings of drug interactions, allergies, and contra-indications; warnings of out-of-range test results; and reminders for guideline-based interventions or screenings.[32]

5. **Electronic communication and coordination.** EHRs may facilitate electronic information sharing through interfacing with other EHRs and health information exchanges. EHRs may support care coordination through data sharing, as well as through providing access to secure communication tools such as patient and clinical messaging.

6. **Patient support.** EHRs may provide clinicians with access to electronic patient education material, and connect to local and remote (home) monitoring systems. EHRs may also provide patients with access to their personal health information though a portal or extract data for use in PHR platforms.

7. **Reporting and population health management.** By interfacing with registries and public health systems, EHRs may be able to support reporting of communicable diseases and disease outbreaks to public health agencies. By supporting the aggregation, reporting, and analysis of data across patients, EHRs help to identify and manage populations of patients. This includes the ability to provide reporting of quality and meaningful use measures.

8. **Administrative processes.** EHR systems may be able to capture patient demographic data and support optimal coding of charges for services. EHR systems may link to billing systems to help streamline the claims submission process.

In real-world systems, these functionalities are represented in many different features and functions. For a comprehensive review of published studies on the functionality of the EHRs, see the AHRQ report *Incorporating Health Information Technology Into Workflow Redesign*.[33]

EHRs are *used* across a range of clinical and administrative processes, including patient care delivery, care management, financial and other administrative processes, and patient self-management.[30] During clinical encounters, EHRs support such general tasks as "review patient history, conduct patient assessment, determine clinical decision, develop treatment plan, order additional services, prescribe medications, and document visits."[2(p3)] Armijo and colleagues additionally define the use of EHRs as aiding four different cognitive processes involved in these general tasks. As a *memory aid*, EHRs reduce "the need to rely on memory alone for information required to complete a task." As a *computational aid*, these systems reduce "the need to mentally group, compare, or analyze information."[2] As a *decision support aid*, EHRs enhance "the ability to integrate information from multiple sources to make evidence-based decisions."[2] Finally, as a *collaboration aid*, these systems enhance "the ability to communicate information and findings to other providers and patients."[2(p2)]

EHRs are multifaceted applications that are usually integrated with other systems or modules; as such, EHRs are not static systems, but are typically updated, customized, and configured to support a range of users and tasks across clinical and administrative processes. In this regard, EHRs may offer a few core functionalities: what is termed "EHR-Lite," consisting of a few basic features such as health information and results reporting to more advanced systems integrating all of the functions described above. This level of design is amenable to mobile platforms, such as smartphones, and will not be considered in this toolkit as it does not represent a comprehensive EHR. Nonetheless, defining EHRs in terms of their *functionalities* and *uses* will help structure the identification and analysis of usability problems (according to a specific function) and the usability methods and instruments appropriate for use in the toolkit.

EHR Usability Defined

Usability is important in promoting both the widespread adoption and "meaningful use" of EHRs as prescribed in the HITECH Act. While utility refers to the existence (or absence) of a system capability or feature necessary to carry out a task, usability describes how easy it is for users to accurately and efficiently accomplish a task while using a system. Missing critical functionalities, poor reliability of the software, or inadequate match between interface features

and user tasks in general will have a strong impact on users' ability to conduct their work, independently from the usability of the available system features.

The National Center for Cognitive Informatics and Decision Making in Healthcare (NCCD) funded by the Office of the National Coordinator for Health IT under the Strategic Health IT Advanced Research Projects (SHARP) Program, defines usability based on its TURF (Task, User, Representation, and Function) framework as "how useful, usable, and satisfying a system is for the intended users to accomplish goals in the work domain by performing certain sequences of tasks."[34(p3)] This definition establishes a framework for setting usability goals and specific evaluation measures. Within the TURF framework there are three dimensions and measures of usability. These are the following:

- *Useful* refers to "how well the system supports the work domain where the users accomplish the goals for their work, independent of how the system is implemented." [34(p4)] A system is useful if it contains the domain functions necessary and only those functions that are essential for the work. Useful is measured by the "percentage of domain functions in the EHR vs. all domain functions in the work domain and percentage of domain functions over all functions (domain and nondomain) in the EHR."[34(p4)]

- *Usable* refers to whether a system "is easy to learn, easy to use, and error-tolerant."[34(p4)] Usable may be measured by learnability, efficiency, and error tolerance.

  - *Learnability* is how quickly a new or novice user learns or relearns the user interface to conduct basic tasks. Learnability is dependent on the consistency of the interface and the ability of the interface to allow exploratory learning by including undo or cancel functions.[24] It can be measured by the time it takes to learn a new task.

  - *Efficiency* is defined as the speed with which a user can complete a task or accomplish a goal. It is typically measured by the length of time required to complete a task, task success, number of keystrokes, and number of screens visited.[24,35] Efficiency may also be measured by objective measures of mental effort, such as the percentage of mental steps over all steps (mental and physical).

  - *Error tolerance* refers to the ability of the system to help users avoid and recover from error. Examples of error measurement include frequency of errors and recovery rate of errors.

- *Satisfying* consists of a set of subjective measures regarding a user's perception of a system's usefulness and impact and how likable a system is.[24,35] The main measures include instruments and interviews that may measure the users' perception of a system.[34]

One proposed advantage of using the TURF framework usability dimensions is that all three dimensions (useful, usable, satisfying) can be measured systematically, and in the case of the first two dimensions, there are objective and empirically proven measures available.

There are a wide range of usability factors that may be extended or grouped under the usability factors above. For example, the degree of simplicity and degree of naturalness,[*] respectively, may be viewed as extensions to the factor of learnability. Further, factors may be extensible and flexible depending on the system's use environment. For example, the concept of usable may be extended to a measurement of "interruptibility" in the primary care setting, which measures how well a user can operate the system in an environment beset by frequent task shifting and disruptions to the flow of activities. We consider that the usability factors outlined above are sufficient for identifying the majority of fundamental usability problems found in EHRs in primary care settings. Thus each of these factors should be identified and quantified within the methods offered in the toolkit.

Summary. There are eight core functionalities of EHRs that are relevant to the design of the toolkit. For the purposes of developing the toolkit, usability is defined in terms of three main attributes: useful, usable, and satisfaction, which have several subdomains and facilitate measurement. The main usability factors described above will help frame which general usability factors need to be assessed within a usability toolkit. Since the toolkit is focused on the use of EHRs in primary care settings, the next chapter discusses important attributes of primary care settings that should inform selection of methods and analyses for inclusion in the toolkit. Examples of current usability problems specific to EHR functions are also discussed in Chapter 4. These examples show how pervasive usability problems are in currently implemented EHRs.

---

[*]Naturalness makes reference to how automatically "familiar" and easy to use the application feels to the user, and how consistent the design and screen flows correlate to the user's tasks and expectations.[85]

Chapter 3. Primary Care Attributes

This chapter reviews some of the important aspects of primary care practices that should inform selection of usability methods and toolkit development, including the types of users, their tasks, and their environments. We additionally provide examples from the literature regarding common usability issues with existing EHR systems.

Primary Care Attributes

Prior work has noted that primary care settings and EHR users within these settings have attributes important to assessing EHR usability.[20,24,35-38] These include the following:

- **Diverse users.** In primary care settings, EHR users include a range of clinical and administrative staff, such as physicians, nurses, physician assistants, advanced nurse practitioners, medical assistants, and office staff. For all these users, the ability to quickly gather, select, and interpret information is critical to effective decisionmaking. These users also possess a range of skills, knowledge, experiences, and responsibilities; as such, they may have different mental models and methods for performing tasks. Well-designed EHRs and tests for assessing their usability must consider the range of users, their information needs, and corresponding mental models.

- **Varied encounters and patients.** Primary care practices treat a wide range of acute and chronic conditions and offer preventive care services. Further, within a given population, primary care providers treat patients at different levels of disease severity and with varying comorbidities (for example, type 2 diabetes and heart disease). From the simple to the complex, from the urgent to the longitudinal, from the preventive to the therapeutic, primary care practices and their information systems must support interventions for a wide range of clinical situations. Accordingly, a prior AHRQ report has suggested that the evaluation of EHR usability in primary care settings focus on four areas: acute care, chronic care, preventive care, and care of patients with undifferentiated symptoms.[2]

- **Complex tasks.** Common tasks in primary care are frequently complex. For example, in order to "refill a medication,"[24] primary care providers need to consider the following:

  1. Past data points (e.g., medication history, last visit date, relevant lab values, last clinic note).

  2. Future data points (e.g., next lab or visit date).

  3. Medical evidence personalized for the patient (e.g., what is the cholesterol goal for this patient, how often do labs need to be checked on this medication).

4. Contextual relevance of #1-3: where is the patient in the life cycle of this medication (e.g., did he or she just start it or has the patient been on it for 5 years, reason for refill).

5. Task of formally creating/approving the refill.

6. Review of cost and formulary coverage that may affect medication options and cost: communicating with their assistant or the pharmacy if needed.

Furthermore, Armijo et al. outlines seven key example tasks that an EHR in a primary care setting needs to support,[2] and we offer an eighth task (followup activities). A well-structured usability assessment should consider these common tasks, which include the following:

1. Review patient history

2. Conduct patient assessment

3. Determine clinical decision

4. Develop treatment plan

5. Order additional services

6. Prescribe medication

7. Document visit

8. Followup activities

- **Shared workflows.** Across primary care practices, multiple clinicians are often involved in the care of a single patient. At different points during a single encounter, nurses, physician assistants, and physicians are interacting with patients either separately or together in assessing problems, making diagnoses, determining therapeutic options, and communicating care plans. EHRs must support the collaboration of clinicians in jointly completing a task, or set of related tasks, in caring for patients. Similarly, assessments of EHRs need to account for the shared nature of therapeutic and preventive care.

- **High pressure, interruptive environments.** Primary care clinicians are typically conducting tasks under significant time pressure and in environments that include multiple demands for their attention. The ability of an EHR system to meet the demands of this environment will affect efficiency and effectiveness, which are defined as the ability to complete a task quickly and accurately, respectively, and are considered two of the most important human factors for EHR users.[24]

- **Risk averse, low tolerance for errors.** As with all care providers, primary care clinicians seek to maximize benefits to patients while minimizing risks to their health—and to the clinician's own legal liability—from different therapies and

procedures. Moreover, primary care clinicians have little tolerance for errors. Systems that are difficult to use and do not reliably aid clinicians may lead to errors in documentation or suboptimal judgment—resulting in unnecessary risks to patients and increased liabilities to providers.

The diversity of users, complexity of needs, varieties of patients, approaches to task completion, accompanying time pressures, and low tolerance for errors are important considerations in determining what evaluation methods are most appropriate and relevant for the toolkit. Primary care attributes, then, will inform the development of selection criteria used in our assessment of usability methods in later chapters of this report. Whatever method(s) are selected, they must be able to assess the three usability factors outlined in the prior chapter.

In addition to selection of usability methods, these primary care attributes are also important to consider in the design of the toolkit itself. Ideally, an effective toolkit will account for these attributes and be tailored to meet clinician's needs for fast, accurate, clear, and actionable results from usability assessments of their current EHRs.

Summary. Primary care settings feature a diversity of EHR users with complex needs, different approaches to task completion, and low tolerance for errors. They serve a variety of patients with a range of conditions in a highly interruptive, time-pressured environment. These attributes are important to consider when evaluating usability methods and designing the toolkit. Chapter 4 delves into examples of specific EHR usability issues encountered in primary care as well as other health care settings.

Chapter 4. Usability Issues in EHRs

We identified several studies from inpatient and outpatient settings that consistently reported a few core EHR usability issues. EHRs, and specific features within these systems, such as order entry and results reporting, have been shown to exhibit the following usability issues: poor organization and display of information, interference with practice workflow, increase in cognitive burden, and poor functional design.

Usability Issues in EHRs

Poor organization and display of information. To perform tasks efficiently and correctly, clinician EHR users need clear, concise, and easily accessible information, and well-integrated system functions. The limited screen space and poor interface design of some systems add to clinician's cognitive workload and increase the risk of errors.[14,20] For example, Smith et al. describes three case studies where poor screen design and layout increased prescribing errors in e-prescribing systems.[39] In a tertiary teaching hospital, Koppel et al. found that some order entry system displays did not provide a coherent view of patients' medications, separated ordering functions, or had inflexible ordering formats. These issues were associated with errors such as double dosing, incompatible orders, or wrong orders.[16]

Interference with practice workflow. Research indicates the importance of EHRs matching the workflow of its users, making information easy to find and allowing for quick navigation within and across functions.[12] Yet a chief complaint about EHR systems is that they do not integrate well with user workflows; health care providers often cite the need to change their workflow as one of the main reasons for not adopting an EHR.[32] Research on clinical documentation systems in an inpatient setting shows that unless there is compatibility between the work structure imposed by the system and routines that clinicians typically follow, there is likely to be resistance to the technology.[40] According to Khajouei's systematic review,[41] lists that require scrolling do not match the workflow pattern of physicians. Presumably, scrolling through long lists is time consuming. These long lists also often have options grouped too close together, making it very easy for the user to incorrectly choose the option above or below the correct options.[41]

Further, the lack of integration between systems may inhibit optimal workflows.[42] An example of this is when a system recommends an action but does not enable a user to easily and quickly complete that action.[43] Specifically, a drug-drug interaction alert might also offer suggestions for other drugs and feature a direct link to the order entry system for ordering the suggested medication(s).[44] Further, the lack of integration with administrative systems, such as billing and scheduling, creates additional steps in the practice workflow.[45]

Increases in cognitive burden. Clinicians are typically conducting tasks under significant time pressure and in settings that include multiple demands for their attention. Clinicians must also process a massive amount of information while remaining accurate and efficient. When combined, time pressures, conflicting demands, and information burden can lead to cognitive overload.[24] The likelihood of errors increases when EHR systems do not adequately support the

mental models of physicians.[46] For instance, an EHR's failure to cluster test results in meaningful ways increases a clinician's cognitive burden; grouping tests related to a certain diagnosis is a way to help alleviate this problem.[47,48] A lack of visual cues for lab results that are out of normal range is another issue that can tax clinician's cognitive resources.[49] Several studies have identified practices that they believe will improve their management of test results, such as highlighting abnormal results by placing them at the top of a list and grouping meaningful items together, so that all tests related to a certain diagnosis can be seen together.[47,48]

The problem of cognitive burden also extends to problems with naming conventions. A lack of standard clinical nomenclatures, or poor implementation of these nomenclatures within EHRs, makes information search and retrieval difficult across systems.[50]

Additionally, physicians experience an increase in mental workload from alerts. Low specificity and high frequency of alerts are causes of alert fatigue,[37] and clinicians may override useful alerts due to this fatigue.[51] A lack of visual cues and feedback, especially for severe alerts, can compound mental workload and alert fatigue.[49]

Poor design of system functions. The challenges EHRs pose to clinician's workflow may stem from the design of EHR functions. Poor documentation capability at the point of care affects the quality of information to support health care decisions as well as other key activities including compliance, external reporting (such as population health and surveillance activities), and billing.[52] Excessive use of defaults, templates, and copying may affect data quality and increase potential for fraud and abuse.[52] Several studies document provider dissatisfaction with the clinical note format in their EHR systems.[53] For instance, many EHRs only provide discrete fields that force providers to choose one symptom over another by clicking a checkbox, and constrain free-form notation to describing or annotating that symptom.[53,54] A comprehensive analysis of a large body of handwritten clinical notes showed that forcing a physician to choose one option to describe a given symptom was not adequate for matching the workflow of the physician.[55] Moreover, limiting options may prevent a physician from adequately documenting a patient's condition and may potentially lead to adverse outcomes. A time-motion study of a clinical documentation system in an inpatient setting[40] found that clinicians requested the software vendor to add free-text data capture to their EHR. Although this may be problematic for those trying to develop and implement more robust CDS interventions, there is a need for a balance when structured data are preferred to enable other EHR functionality, and it may be that better representations of fields or options are needed to capture relevant clinical information.

Management of test results is another area of particular concern for health care providers.[48] Errors in managing test results can create significant harm to patients.[56] Finally, well-designed ordering capabilities are important to clinicians. Certain features, such as auto-fill of medications, are more likely to improve usability by increasing efficiency and thereby increasing provider use of the system.[57]

Usability Issues by EHR Functionality

In addition to these general usability problems found in the EHR, we cluster these findings by EHR functionality as seen in Table 1. This categorization of these problems found in the literature allows the development of tools within the toolkit to focus on specific usability issues found in specific areas of the EHR and additionally on the tasks and workflow of the users.

Table 1. Common usability issues by EHR functionality

EHR Capability	Examples from the literature
Across all functionalities	- Poor organization of the information display that fails to balance the information needs of the physician and the presentation of information; it may include overly cluttered screens or screens that do not present all the necessary information.[14,17,20,24,36,46,55,58] - Poor documentation functionality at the point of care, which affects the quality of information to support health care decisions as well as other key uses including compliance, external reporting (such as population health and surveillance activities), and revenue cycle management.[52] - Excessive layers of screens and click-through increase burden and disrupt the workflow.[41,59] - Practice workflows that are not aligned to support efficient and effective use of the technology.[14,43,60-63]
Health Information and Data Management	- Lack of options for structured data entry may increase workload and affect data quality.[14,52,64] - Lack of options for free text entry (i.e., overly structured notes) may remove important context.[39,52-54,65] - Excessive use of defaults, templates, and copying may affect data quality and increase potential for fraud and abuse.[52]
Order Entry Management	- Prescribing systems that are not integrated with other parts of the EHR system, for example, with preferred formulary lists, pharmacy database, and clinical decision support systems.[44,58,66,67] - Lab order systems that fail to cluster laboratory tests into groups that match preferred cognitive patterns.[41,68]
Results Management	- Failure to present results information to match the physician's cognitive pattern; for example, the system only lists test results alphabetically, whereas physicians may prefer chronological or clustered by clinical relevance or severity listing of results.[24,36,48,69,70] - Failure to warn if a patient has missed a test.[48,70] - Use of lists that require scrolling.[41]
Clinical Decision Support	- A lack of customization of alerts for sensitivity or severity leads to poor implementation, limited use of decision support, and alert fatigue.[37,51,52,57,71-75] - Decision support systems that do not allow the filtering of certain alerts that the clinician may find inconvenient or not clinically useful based on their practice setting (e.g., certain drug-drug interactions, or excluding certain drug-allergy cross-sensitivity rules).[76] - Alerts that are not accompanied with appropriate visual cues, such as based on severity.[49,51,77] - The lack of integration between dual systems in order to complete a recommended action; for example, the decision support system recommends an alternate drug dose, but does not integrate with the e-prescription system to provide for easy selection and ordering of the recommended drug.[39,42,43,54,58,78] - Ill-timed alert during encounter, not useful or inappropriate alerts.[46] - Not useful or inappropriate alerts.[46]

Table 1. Common usability issues by EHR functionality (continued)

EHR Capability	Examples from the literature
Electronic Communication and Coordination	• Lack of standard nomenclatures across systems can make search, retrieval, and comprehension difficult.[50,79,80]
Patient Support	• Lack of ability to provide patient education materials in plain English.[48,81-84]
Administrative Processes	• Lack of integration with administrative systems, such as billing and scheduling.[45,59]

We identified usability issues specific to EHR functionalities. These usability issues will be important to factor into the development of selection criteria for including or excluding methods from the toolkit. Ideally, methods and instruments for usability evaluation should be assessed and selected based on whether or not they target specific system functionalities and issues.

However, in real-world settings, the general usability problems we discussed are not discrete: EHR design and usability issues contribute to each other. For instance, poor information displays and poor design of EHR functions may impede clinicians' workflow and, consequently, increase their cognitive burden. In terms of toolkit development, the initial toolkit should help users correctly identify broad issues that span across an EHR's functionalities and not focus solely on discrete system functions. Consequently, usability evaluation methods should also be reviewed, in part, based upon their ability to support general assessments.

Summary. EHR systems have been shown to suffer from several common usability issues that, like primary care setting attributes, need to inform toolkit development—notably the criteria for review and selection of evaluation methods. In Chapter 5, then, we describe the range of methods and tools for measuring the usability of EHR systems.

Chapter 5. Methods and Tools to Measure Usability

To determine usability evaluation methods that could be incorporated into a toolkit for primary care clinicians, we first compiled a list of conventional methods. Although there are many usability methods that should be employed in the design and development of EHRs, we focused primarily on evaluation methods, since the toolkit will be used by primary care practices to evaluate EHR systems. We begin by discussing our review of the literature to identify usability evaluation methods, and follow with a description of each method under consideration for inclusion in the toolkit.

Search Strategy and Literature Review Process

To identify usability evaluation methods that should be taken into consideration for inclusion in the toolkit, we first compiled a master list of methods. We used a user-centered design framework to begin the search for methods in our review. This framework is outlined in the NIST *Guide to the Processes Approach for Improving the Usability of Electronic Health Records*,[35] but also documented elsewhere in detail.[85-87] The framework consists of seven phases: (1) planning, (2) determining context of use, (3) determining user needs and requirements, (4) developing information architecture, (5) developing a low-fidelity prototype, (6) developing a high-fidelity prototype, and (7) formal usability testing including user feedback. Since the intended toolkit users will only be accessing the toolkit to evaluate their current EHR, we reviewed only usability testing and evaluation methods (phase 7). To help structure our search and review, we then developed inclusion criteria for methods focusing on usability testing and evaluation. Inclusion criteria included were the following:

- Evaluation methods that could identify usability problems with an HER.

- Studies published in English.

- Studies published between 1980 and 2010.

We excluded methods that are used for specific applications such as situation awareness rating techniques, usability questionnaires that measure multimedia or Web sites such as Measurement of Usability of Multimedia (MUMMS) and Web site Analysis and Measurement Inventory (WAMMI). We additionally excluded animal studies, commentaries, editorials, dissertations, and letters. To develop the master list of usability testing methods using these inclusion and exclusion criteria, we initially consulted textbooks,[88-98] usability Web sites,[99,100] and usability reports published by AHRQ,[1,2] HIMSS,[24] and NIST.[35]

We then conducted an inventory of methods by reviewing the literature using seven different databases, which included PubMed[®], CINAHL[®], Psych Info, Web of Science, Science Direct, ACM Digital Library, and ProQuest. We searched the literature for usability evaluation methods to complete the list of testing methods using the following search terms: *"usability evaluation methods," "usability inspection methods," "usability testing methods,"* or *"usability engineering methods."* We reviewed the titles and abstracts of 5,948 articles to determine their relevance in

this search for additional usability methods that could be added to the master list of methods. If the usability method was labeled as being used in a usability study, we included it in the master list. The final master list of methods compiled from this search included heuristic evaluation, cognitive walkthrough, predictive modeling (Goals, Operators, Methods, and Selection Rules [GOMS] and keystroke level models), remote/automated evaluation, laboratory testing, failure mode effects analysis, critical incident technique, Subjective Workload Assessment Technique, NASA Task Load Index, Subjective Mental Effort Questionnaire, and usability questionnaires (System Usability Scale, Questionnaire for User Interaction Satisfaction, Computer System Usability Questionnaire, Software Usability Measurement Inventory, After Scenario Questionnaire, Usefulness, Satisfaction and Ease of Use Questionnaire, Perdue Usability Testing Questionnaire, and End-User Computing Satisfaction Questionnaire).

Once we compiled the master list of methods to be reviewed for inclusion in the toolkit, we then searched the seven databases listed above by the title of each method. Inclusionary criteria for this search were similar to those developed for identifying methods initially, and included the following:

- English language only.

- Published between 1980 and 2010.

- Provided detailed information on how the method is carried out, as well as its advantages, disadvantages, and psychometric evaluation (when available).

- Described how the method has been used to evaluate usability of health IT applications, specifically EHRs.

Using the title of each usability testing method as keywords, the primary author then performed a secondary search of the literature. This search resulted in 3,402 articles focusing on usability evaluation methods. The primary author reviewed the title and abstracts (when applicable) of these articles relative to the inclusion criteria, which resulted in 256 relevant studies. Full-text versions of relevant studies were retrieved and reviewed. We abstracted from each article, the entire reference; a description of the method, benefits, disadvantages of the method; psychometric evaluation when available; constructs measured; and if applicable, the type of study, number of subjects, and study outcome. We included a total of 137 articles.

The following section provides a description of the candidate usability evaluation methods under consideration for inclusion in the usability toolkit. When considering whether the candidate method is appropriate for use in the toolkit, the following criteria were taken into consideration: (1) the intended users are primary care clinicians and nonexperts in usability evaluation, (2) the environment is the primary care setting, and (3) the method must be able to provide quantitative and some qualitative results that could summarize the usability problems with the EHR, provide educational feedback regarding these reports, and enable the provision of the results in a report format.

Candidate Usability Evaluation Methods

Usability evaluation methods can provide feedback about problems with system design before the system is implemented. All of the methods listed below can help identify usability problems with a given interface.

Heuristic Evaluation

Heuristic evaluation is one of the most commonly used discount evaluation methods due to its low cost.[101] It is an inspection method that can prospectively uncover problems with a user interface, indicate the severity of the problems, and make suggestions for fixing the problems. Heuristic evaluation can uncover both major and minor problems not necessarily found with user testing.[85] Although major problems are generally easier to discover than minor problems and are the most important to fix, minor problems can just as easily contribute to data entry errors and are easier to find via heuristic evaluation than by other evaluation methods.[85] Heuristic evaluations that are performed by two or more usability experts can identify more than 50 percent of the usability problems with an interface.[85,102] The method requires that a small set of 3 to 5 experts evaluate a user interface based on their knowledge of human cognition and interface design rules of thumb or heuristics.[21,85,88,103,104] Once the experts identify the heuristics violated within an interface, experts rate the problems in terms of severity on a scale from 1, indicating a cosmetic problem (fix can wait), to 4, indicating a catastrophic problem (immediate fix).[21,91,101] Some examples of heuristics include visibility (users should always be informed of the system state), consistency (interface design standards and conventions should be employed), match (user model matches system model), minimalist (limited use of extraneous information), memory (minimize memory load by using recognition versus recall), flexibility (shortcuts to accelerate performance), message (good error messages), error avoidance (prevent errors), closure (clear closure on all tasks), reversible actions (undo functions), control (avoid surprising actions), feedback (provide informative feedback about actions), language (utilize the users' language), and documentation (help options).[85,88]

However, there are some minor drawbacks with this method. Heuristics have not been standardized, the heuristics that are published are not well-defined and thus can be interpreted in different ways, and the results of the evaluation are highly dependent upon the expertise of the evaluators.[105,106] Heuristic evaluations are good at exposing the majority of usability problems within an interface.[107] However, heuristic evaluations cannot reveal all problems within a system.[108] Using this technique along with other methods may reveal both local (particular interface screens) and global problems (system issues).[92,105] There are numerous examples of its use in evaluating health IT applications.[102,109-113]

Cognitive Walkthrough

Cognitive walkthrough is a usability inspection method that compares the users' and designers' conceptual model and can identify numerous problems within an interface.[93,114,115] It can be used to evaluate an interface for ease of learning[115,116] and to disclose many problems that a first-time user would encounter with system functionality and ease of system use. It defines how well the interface supports "exploratory learning," or how well the first time user can

perform a task without formal training.[93] It is a technique that focuses on errors in design that would interfere with the users performing a task. It also explains mismatches between the users' and the designers' conception of a task.[93] A cognitive walkthrough is conducted by an analyst using predefined scenarios addressing four steps that identify the users' goals and how easy it is for users to meet these goals.[93] Before beginning this type of analysis, the designer must know the users, the respective tasks they will be performing, and the accurate order of actions for each task. As the evaluator steps through the actions to accomplish a task, he or she tries to answer four questions: (1) will the user try to achieve the correct effect, (2) will the user notice that the correct action is available, (3) will the user associate the correct action with the desired effect, and (4) will the user notice that progress is being made toward accomplishment of his or her goal.[93] If the evaluator answers "no" to any of the questions regarding the action leading to the goal, then that action is considered to have usability problems.[93]

Cognitive walkthroughs tend to find more severe problems,[117] but find fewer problems than a heuristic evaluation,[107] are labor intensive, and require a usability expert.[118] Cognitive walkthroughs have been successfully used to identify problems with health IT applications.[69,111,119,120]

Laboratory Testing

Controlled User Testing. Laboratory-based usability testing in a controlled laboratory environment is one way to validate interface design decisions and to test alternative interfaces.[85] It is considered the gold standard of usability engineering methods.[121] Lab-based usability testing includes both qualitative and quantitative studies since it collects both objective data such as performance metrics (e.g., time to accomplish the task, number of key strokes, errors, and severity of errors) and subjective data such as the vocalizations of users thinking aloud as they work through representative tasks or scenarios. Controlled user testing involves creating a series of commonly used task scenarios and asking the user to step through these tasks while thinking aloud[91,95] about what they are accomplishing while carrying out the tasks. The user is under the observation of a trained usability expert who records the events of the session.[122] Task performance can be measured and the evaluator only provides input if the user cannot complete the task without assistance or if the user stops talking. Usability testing also involves creating a series of commonly used task scenarios. This technique can be used to test different versions of a system with end users.

The evaluation walkthrough or a pluralistic usability walkthrough is another type of controlled user test in which a usability expert walks through the system with the intended users observing and commenting on the system.[85] The walkthrough process involves creating a problem list and assigning a severity level to each of the problems found in the walkthrough. The pluralistic usability walkthrough generally involves not only the representative users and usability experts, but additionally the developers.

In addition to performance metrics in controlled user testing, the evaluator can note the behavior of the user and any other observations that he or she finds pertinent to the design of the system. User testing in a laboratory environment including walkthroughs generally involves five or more representative users.[96,123-126] These types of usability tests uncover hidden functional and

interface design flaws.[118] There are numerous examples of studies using this approach successfully in the health care domain.[18,127-135]

Think aloud methods[91] are often used in controlled user testing and require users to talk aloud about what they are doing and thinking[95] as they use an interface or work through scenarios with a prototype. The aim of this technique is to collect procedural information about mental processing wherein the investigator can make deductions about problems a potential user may have with an interface. It has two benefits: it provides the evaluator information about design problems and it shows the evaluator the learnability of the system.[105] The use of audio-video recordings while the subjects are working with the interface also provides a rich source of data for later coding and analysis.[127] However, the reliability and validity of the think-aloud protocol has been questioned because the subjects' cognitive processes may be interrupted by the evaluator, for example, if the evaluator asks clarifying questions about a participant's comment;[136,137] thus it is important to also analyze what the subjects pay attention to within the interface, and not just their opinions.[105]

Eye-tracking may also be used with controlled user testing. Eye movement studies suggest that changes in gaze are directed to the demands of whatever task is at hand.[138] Eye-trackers record readers' eye movements as they fixate on words or groups of words or on objects within an interface. By superimposing the location of the gaze onto the text, the investigator can determine exactly where the subject is looking and the length of time of the gaze. Fixation time can provide measures of where the subject is focusing his or her attention.[139,140] Numerous studies have shown that eye gaze plays a central role in the demands of a task. These studies further suggest that gaze control and saccadic eye movements perform a significant role in mediating cognition.[141-143] Additionally, duration of eye gaze may provide a rough estimate of duration of cognitive processes, including a person's mental workload.[143] Thus, eye gaze data are quantitative data that are considered an objective approach in the study of cognitive behavior.

Remote Evaluation

Remote evaluation provides ecologically valid data from remote locations[144] and has several benefits such as eliminating the need for participants to travel to the laboratory, decreased costs associated with evaluation, and potential ability to engage a diverse group of participants.[145-148] Remote evaluation methods can be used to collect the results of surveys and questionnaires and to collect information regarding users' navigation, keystrokes, task time, usage logs, remote recordings of computer screens, qualitative data such as think-aloud protocols, and error rates through Web-based screen-recording software.[146,149,150] Remote evaluations have been used to evaluate Web-based information systems.[151-153] There are several different synchronous or asynchronous approaches to conducting remote evaluations and these have been classified as portable evaluations (usability equipment taken to user site), remote questionnaires or surveys (Web-based usability questionnaires), remotely controlled evaluations (the evaluator controls the users' computer remotely), video conferencing as an extension of the usability laboratory, instrumented remote evaluation (electronically tracking users actions remotely), and semi-instrumented remote evaluation (users send only negative task events to evaluators).[154,155] Studies examining the differences between laboratory testing vs. synchronous vs. asynchronous remote usability evaluations have shown that laboratory and synchronous remote evaluations

identify a similar number of usability problems, while remote asynchronous methods identify fewer problems. Yet, due to cost and time savings, using the asynchronous approach is still worthwhile.[149,150]

Usability Questionnaires

One of the most common ways to collect self-reported data across tasks is to use usability questionnaires. Subjective usability questionnaires measure the users' experience and perceptions with a particular system after using the system in question. These questionnaires can identify areas within a system that need improvement and can provide a measurement of the overall usability of a system. The questionnaires measure a diverse set of constructs such as satisfaction, efficiency, effectiveness, learnability, perceived usefulness, ease of use, information quality, and interface quality. There are many valid and reliable usability questionnaires that are both proprietary and nonproprietary. Some examples include System Usability Scale (SUS), Questionnaire for User Interaction Satisfaction (QUIS), Computer System Usability Questionnaire (CSUQ), Software Usability Measurement Inventory (SUMI), After Scenario Questionnaire (ASQ), Usefulness, Satisfaction and Ease Of Use Questionnaire (USE), Perdue Usability Testing Questionnaire (PUTQ), and End-User Computing Satisfaction Questionnaire (EUCS). For a complete description of these questionnaires, see Chapter 7.

Other Human Factors Methods

Predictive Modeling—Goals, Operators, Methods, and Selection Rules (GOMS) and Keystroke-Level Models (KLM)

GOMS is a cognitive modeling technique that analyzes human performance in interactive systems[156,157], such as the speed and the steps the user takes to complete a goal (i.e., ordering medications). It is a type of task analysis that analyzes low-level perceptual-motor issues, such as keystroking, and can be used to predict task performance; addresses the efficiency and complexity of a system; and can optimize parts of a system interface.[158] The goals are what the user is intending to accomplish. The operators, which can be perceptual, cognitive, or motor, are the actions the user takes to accomplish the goals. The methods are the procedures and operators of the goals and subgoals. Finally, the selection rules determine the method to use from potential alternative methods to accomplish the goal.[17]

GOMS and keystroke-level models can be used in the design phase, during development, and after implementation with existing systems to determine human performance and interaction with a system.[159] GOMS is the first step in a cognitive task analysis. It describes procedural knowledge the user needs to carry out specific tasks within a system. Although GOMS is less expensive than empirical laboratory testing, it must be tested for validity by usability experts who watch the users carry out their tasks within a system.

The Keystroke Level Model (KLM) can identify numerous problems with an interface.[160,161] It shows differences in execution times of each performed task by summing up the time taken for keystrokes, pointing, clicking, thinking, waiting, and deciding. The KLM is tedious to perform

because it requires the usability expert to determine the goals and subgoals of a task and the methods to reach the goals and subgoals, make all assumptions such as typing expertise, code each movement of the user such as mouse clicks or keystrokes, and determine the time required for the user to make decisions and time for the system to respond to their actions. Finally, the expert must sum the times for keystroke or mental operators. KLM can show problems with the predicted execution times of an application and point out particular areas where a user might be spending an inordinate amount of time with the functionality of an application. When time is a factor in performing tasks such as in an environment where productivity is critical, KLM should be considered.[160,161] See Johnson, Endoh, and Saitwal for examples of how GOMS and KLM have been used in health care.[17,162,163]

Risk Assessment

Risk assessment identifies the probability of human error occurrence within a system. These approaches include Failure Mode Effects Analysis (FMEA) and critical incident techniques.[90,97,164]

Failure Mode Effects Analysis (FMEA). FMEA is a method originally used by the military to examine the effects of human errors on systems and can examine failures of individuals or teams with tasks or functions.[90] It is generally conducted by a team of experts including end users, designers, developers, and human factors experts who brainstorm within scenarios to determine anticipated probabilities of harm associated with tasks.[97] The process uses a bottom-up process that examines the tasks of the users, speculates on potential failures within the task, and determines the consequences from the error.[90,165] FMEA has several advantages such as its ease of use, and its ability to be used in different application areas.[90] It requires very few resources and assists the evaluator with determining the severity of the errors by including the consequences of the errors. However, the accuracy of this test relies strongly on evaluator expertise.[90] Drawbacks to this analysis are that it collects qualitative data and the process can be time-consuming. FMEA has been used in the health IT domain to evaluate the impact of computerized provider order entry on reducing errors in pediatrics and demonstrated a reduction in ordering errors.[166]

Critical Incident Technique (CIT). CIT is another method that examines the cause for human-computer interaction problems that occur in task performance.[90] It is a way to collect information about features, functions, or events within an environment that result in critical incidents such as usability problems that can cause medical errors. The technique is conducted by an experienced analyst who observes the users' behavior as the critical incident unfolds, and thus the cause can be identified and resolved.[167] It can also be measured through questionnaires and structured interviews.[168] Users can also self-report critical incidents that caused significant losses that are recorded in a database and analyzed for trends or clusters of problems related to system or human issues.[90,164] These trends or clusters can then be used to identify solutions for these problems. There are several advantages to this technique: it is helpful for identifying rare events; it has high face validity; it can provide information on types of errors, when they occur, and the probability of their occurrence; and it is cost effective.[90] The disadvantages lie in the subjective and potentially fallible issue with self-reporting of events.[90] CIT has been successfully used to

assess an online evidence system and elicited examples from clinicians on how this system affected practice and the delivery of care.[169]

Workload Evaluation

Workload evaluation assesses the workload of the user while he or she works on a prototype or developed system. Mental workload has been defined as "the cost of performing a task in terms of reduction in the capacity to perform additional tasks that use the same processing resource."[170] Workload is measured to calculate the associated costs of performing tasks to determine user and system performance. Cognitive load is the amount of cognitive resources, particularly working memory that is used during thinking, learning, problem solving, and reasoning.[171] Since working memory is limited, distractions, new information, and complex information have the potential to interfere with clinical decisionmaking and can lead to errors.[172] To assess how usability of a system affects workload (outcome), a controlled test using a standard clinical context would need to be done across different EHR products. This evaluation would then provide insight into "usability" of one system compared to another. In medicine, it is important to have the ability to measure outside of the controlled laboratory how EHRs may be contributing to the cognitive load of the clinicians. Although this is generally carried out with a cognitive task analysis[18] that is expert intensive, the techniques shown below are some examples of how to measure workload within a usability toolkit. There are both objective and subjective methods.[173] The objective methods include the empirical techniques of cognitive task analysis. The questionnaires are the subjective methods.

Subjective Workload Assessment Technique (SWAT). SWAT is the most commonly used technique to measure mental workload. It is a multidimensional scale that measures time load, mental effort, and psychological stress on three levels (low, medium, and high) for a task being carried out and provides an overall workload rating.[94,174] SWAT must be normalized with a conjoint measurement and scaling procedure to ensure validity prior to use. This requires users perform a card-sorting technique.[94,175] Although it is considered a reliable and valid tool to measure workload, it is a lengthy procedure that makes this technique more complicated than others. There are two major limitations with this technique: (1) the scale must be normalized for each subject with a card-sorting procedure, and (2) it is not very sensitive for low mental workloads.[176] The simplified SWAT reduces the pretest completion time and has been proposed as an equivalent alternative method.[176] No evidence of its use was found as a method to measure workload in relation to EHRs.

NASA Task Load Index (NASA-TLX). The NASA Task Load Index is a multidimensional scale that uses six components on a 20-point scale to measure workload experience: mental demand, physical demand, temporal demand, performance, effort, and frustration.[94,177-179] Although NASA-TLX has been used in single domains, it has not been validated to be used in multitask situations.[177] This method is conducted by asking participants to rate their workload after completing a task on the six subscales. After completing all of the tasks, the participants are then asked to make 15 paired post-test comparisons, which normalize the ratings for each participant.[94,177,178] A weighting procedure is then employed to provide one summarized score. The instrument has been successfully used in different contexts, such as simulated flight tasks, air combat, and vigilance task performance.[178] Two of the six workload scales, mental workload

and frustration, were used to measure the usability of redesigned computerized clinical reminders.[58] Hertzum and Simonsen used the NASA-TLX to determine the effects of an electronic patient record on mental workload and found significant improvements with using the EPR compared to using paper records.[180]

Subjective Mental Effort Questionnaire (SMEQ). The Subjective Mental Effort Questionnaire contains one scale with nine labeled anchors (scale runs from 0-150) that, for example, ranges from "Not at all hard to do" to "Tremendously hard to do" and measures single task effort.[181,182] It is used after users complete each task in a series of tasks.[183] One advantage of SMEQ is its large number of response choices. The questionnaire has been shown to have high reliability and validity.[184] No evidence of its use in relation to EHRs was found in the literature.

Summary. There are many general usability evaluation methods that have been successfully used in health care and other domains to assess different aspects of usability of a system. Some generic methods described here have been successfully used to assess different aspects of EHR usability and some have not been used in this domain. The next chapter provides an assessment of these methods and suggestions for the methods that may be incorporated into a usability toolkit. Most require human factors expertise; some could be administered by nonexperts. For inclusion in the toolkit, it is important to consider self-administration, as primary care clinics will not have usability experts to administer these methods and any related instruments.

Chapter 6. Assessment of Usability Methods

Usability Toolkit Selection Criteria

Based on the scope of the project, guidance from AHRQ and the Technical Expert Panel, and factors identified through the consideration of EHR functions, attributes of primary care settings, and EHR usability problems, we defined high-level criteria for selecting methods for inclusion in the initial toolkit. These criteria include the following:

1. **Applicability to current systems.** The methods in the toolkit should be appropriate for tests of currently available systems.

2. **Specificity to EHRs.** Toolkit methods should be able to assess usability problems both general to EHRs as well as those associated with specific EHR functions described in Chapter 3. For example, the ability to turn off or suspend alerts provides flexibility (usability construct) to a specific EHR function, but the "consistent" placement of buttons is a general usability goals across all screens in the EHR.

3. **Usability of output.** The methods within the toolkit must provide meaningful results for primary care clinicians.

4. **Rigorousness.** The methods within the toolkit must be reliable, valid, and able to measure the various usability factors described in Chapter 1.

5. **Ease of understanding.** Toolkit methods must be easy to understand for nonusability experts such as primary care clinicians.

6. **Efficiency.** As noted in Chapter 2, usability methods must accommodate busy primary care settings. Toolkit methods must be easy to use and should not take more than 30 minutes to complete.

7. **Self-administration.** As noted at the end of Chapter 4, toolkit users should be able to apply the method and report the results themselves; the evaluation method should not require usability experts to administer and interpret.

8. **Self-contained.** The toolkit should not require installation or integration with an EHR.

Applicability of Usability Methods for EHR Toolkit

Chapter 5 reviewed a large number of usability evaluation methods. To simplify assessment of these methods and assign weights to various criteria, we grouped the criteria above into two rankings: low and high. A method was rated low if it did not meet any criteria or met some criteria but could not be self-administered. A method was rated high if it met all criteria. Criteria 5 to 7 were particularly important since it is essential that the methods in the toolkit be rigorous, yet easy to apply, rapid, and not require usability experts to interpret. We then assigned one of

these ratings to each method; methods with the highest ratings were deemed most appropriate for inclusion in the toolkit. In addition to these rankings, we also assessed the relative advantages and disadvantages of each method. Table 2 shows the type of method, a brief description of the method, its advantages and the disadvantages, and ranking of appropriateness for inclusion in the toolkit.

Recommended Methods for the Toolkit

The recommendations for the toolkit need to be based on a crosswalk between the requirements of the toolkit and the attributes of the methods. The first requirement for the toolkit is that it is practical for primary care clinicians. This requirement eliminates many methods, specifically, heuristic method, cognitive walkthrough, and laboratory testing, predictive modeling, FMEA, and the critical incident technique. The second requirement for the toolkit is that the method needs to support different types of clinicians, specialties, workflows, and a variety of EHRs. This requirement eliminates the workload evaluation methods since workloads will vary by specialty, workflows, and EHR. Remote evaluation is eliminated given the need for specialized software to record user events. Table 2 provides a summary of the crosswalk between the methods and its appropriateness (based on the toolkit requirements).

The only method not eliminated is usability questionnaires. They have the following attributes:

- Completed in an approximate 30-minute period.

- Assess a variety of EHR functionalities and target usability issues identified in Chapter 4 to varying degrees.

- Provide an assessment, which has many benefits: (1) for each question, individual users of the toolkit could be guided to different educational materials depending on their ratings and (2) can target usability for high-risk EHR functionalities.

- Include measurements at different levels of detail. For instance, these instruments can feature overall satisfaction questions as well as specific questions that inform users regarding improvements they can make to their own implementation, or inform designers about what needs to be redesigned.

- Include open-ended questions that capture important details of usability issues identified by specific questions. Although not analyzed, this information can be captured in a report for clinicians to share with vendors.

  For the toolkit to be developed in this project, we recommend usability questionnaires as the methods for the EHR usability toolkit.

Table 2. Advantages and disadvantages of usability and other human factors methods

Method	Description of Method	Advantages	Disadvantages	Appropriateness Ranking
Heuristic Evaluation[21,85,88,91,92,101,107,185]	Usability experts evaluate a system using a set of design principles/guidelines.	Low cost and addresses both local (interfaces) and global (system) usability problems.	Requires usability experts to complete the analysis, and it may overlook some usability problems that could only be found with user input.	Low
Cognitive Walkthrough[91,93]	Experts imitate users stepping through the interface to carry out typical tasks. Finds mismatches between users' and designers' conceptualization of a task.	Focuses on ease of learning for first time users.	Does not determine all problems with an interface. Requires expertise.	Low
Controlled user testing[85,91,95,122,127]	Users test the system performing representative tasks using verbal protocols. Testing gathers information about the users' performance, includes post-tests of usability and observations made by the evaluator.	Performance measurements can be obtained in addition to verbal protocol information. Quantitative results are easy to compare.	Detailed planning is required prior to running these tests. Requires experts to run the tests in controlled laboratory.	Low

Table 2. Advantages and disadvantages of usability and other human factors methods (continued)

Method	Description of Method	Advantages	Disadvantages	Appropriateness Ranking
Remote Evaluation[144-150,154,155]	System records events as user works through the tasks and collects results of questionnaires. Includes asynchronous and synchronous approaches.	Accurate performance measures can be obtained. Data can be ready for analysis from questionnaires.	Software can be costly.	Low
Usability Questionnaires	Questionnaires that measure efficiency, satisfaction, learnability, system usefulness, information quality and many other measures.	Questionnaires are easy to administer online and provide written feedback and scores. Many are reliable and validated.	May not be specific to EHR systems; may only focus on assessing overall usability.	High
Predictive Modeling— GOMS/KLM[156,157,160,161]	Determines user goals to complete a task, operators to perform the goal, methods to accomplish the goal, and selection rules to reach the goal. Is part of the cognitive task analysis.	Calculates the time to reach the goal. Includes Key-Stroke level models.	Very time intensive and requires usability expertise.	Low

Table 2. Advantages and disadvantages of usability and other human factors methods (continued)

Method	Description of Method	Advantages	Disadvantages	Appropriateness Ranking
Failure Modes and Effects Analysis (FMEA)[90]	Analyzes human reliability, identifies potential failure modes, can be used to study human errors based on tasks and functions.	Cost effective test and can determine errors by severity, permits descriptive information on different types of errors.	Depends on expertise of analyst, can be time consuming to analyze.	Low
Critical Incident Technique[90,164,167]	Identifies and determines design flaw via self-report.	Cost effective. A method of gathering data that can be analyzed for trends. Helpful for rare events, has high face validity, provides information on types of errors.	Dependent on users' verbal reports.	Low
Subjective Workload Assessment Technique (SWAT)[94,174,176]	Evaluates workload measuring time load, mental effort, and psychological stress.	Most frequently cited in workload literature; theoretically grounded.	Scale must be normalized for each subject by means of a card sorting technique, large amount of subject preparation and training. Subjective rating techniques that uses three levels: low, medium, high. Low sensitivity for low mental workloads.	Low

Table 2. Advantages and disadvantages of usability and other human factors methods (continued)

Method	Description of Method	Advantages	Disadvantages	Appropriateness Ranking
Simplified Subjective Workload Assessment Technique (SWAT)[176]	Evaluates cognitive/workload measuring time load, mental effort, and psychological stress.	Theoretically grounded. Correlated well with original SWAT.	Needs validation in a medical environment.	Low
NASA Task Load Index[94,177,178]	Evaluates workload measuring mental demand, physical demand, temporal demand, performance, effort, and frustration. Measures each component subscale with 20 levels.	Uses an adjustment to normalize ratings.	Will not determine many usability issues. Scale must be normalized for each subject but less time intensive than SWAT.	Low
Subjective Mental Effort Questionnaire (SMEQ)[181,182,184]	Subjective measure of mental effort.	Contains one scale with nine labels that measures subjective mental effort after each task completed. Time limited, easy to use.	Requires analysis by usability experts to interpret results.	Low

Summary. This chapter reviews a wide range of usability evaluation methods, and assesses their advantages, disadvantages, and appropriateness for inclusion in the toolkit. Commonly used usability evaluation techniques, such as heuristic evaluations, and more rigorous methods, such as laboratory testing, are impractical for primary care settings given the usability expertise and resources required for these methods. Usability questionnaires would provide the basis of a toolkit for primary care clinicians to enable educational insights into usability issues with their current systems. Chapter 7 provides a review of candidate usability questionnaires that could be used in the toolkit and offers further recommendations for which instruments might be most appropriate.

Chapter 7. Review of Usability Questionnaires

Based on the analysis in Chapter 5, we have identified nine usability questionnaires that could consistently identify general usability issues, are easy to understand, could be completed quickly, and could be self-administered. We describe these nine questionnaires, outlining their advantages and disadvantages, with examples of their use in evaluating health information systems, including EHRs. We compare these questionnaires by their reliability, validity, and attributes of usability measured (Table 3). To ensure a questionnaire provides high-quality results, it should be both reliable and valid. Reliability is a measure of consistency or repeatability of a measurement. In other words, the estimate for reliability should be the same each time a measurement is used with the same subjects and under the same conditions.[186] In psychometrics, reliability should be at least 0.70.[186] Validity assesses what the questionnaire claims to measure or it is the accuracy of the measurement.[186] Although reliability is a prerequisite for validity, it does not ensure validity. While a measure could consistently measure a construct(s), it does not necessarily mean that it is measuring the intended construct(s). The chapter concludes with options for including instruments in the toolkit.

System Usability Scale (SUS)

The System Usability Scale (SUS) developed by Digital Equipment Corporation is a 10-item scale that is scored on a 5-point Likert scale on strength of agreement.[187] SUS assesses the general usability of a system.[188] The scale consists of five positive and five negative statements and the scale provides a single summated rating for the system being evaluated. The scale is given to respondents after testing or working with a system. The score ranges from 0-100 with higher scores indicating better usability.[187] SUS has several advantages: (1) can be used to test different types of interfaces; (2) it provides one score that is easy to interpret; (3) it provides an easy and quick measure to determine the users' perception of the usability of the system; and (4) it is nonproprietary and available for use providing that the source is acknowledged.[188] One disadvantage is that, due to variable findings regarding the instrument's reliability, the SUS should be used in combination with another method to determine the usability of a system. The coefficient alpha, when originally tested, was found to be 0.85 and was based on 77 cases.[189] Later testing by Bangor et al. found the alpha coefficient to be 0.91 based on 2,324 cases.[188] Additionally, Bangor et al. showed some evidence of concurrent validity. The SUS items were based on three usability criteria: effectiveness, efficiency, and satisfaction. However, Lewis and Sauro[190] recently proposed a two-factor structure (usability and learnability) as opposed to a unidimensional factor model (measuring just usability).[187-189] Borci et al. proposed that usability and learnability are independent components of the SUS rating and showed that the factor reliability by the alpha coefficient for usability was 0.81 and 0.76 for learnability factors, thus they showed a positive correlation of usability and learnability.[191] However, their results should be considered preliminary since they only tested one system and their user population consisted of college students. Both limitations may have affected the association of the two factors. In health care, SUS has been used to evaluate the usability of an emergency department information system and investigate usability gaps;[192] a computerized decision aid for patients on adjuvant radioactive iodine treatment;[193] an oncology data retrieval, visualization, exploration, and

analysis system;[194] a method for designing intelligent alarms on monitoring equipment;[195] and a handheld computer self-management tool for patients with diabetes.[196]

Questionnaire for User Interaction Satisfaction (QUIS)

The Questionnaire for User Interaction Satisfaction was developed at the University of Maryland and was designed to evaluate the users' satisfaction with different aspects of an interface.[197] QUIS was developed in 1988. In its current version, QUIS 7.0. uses a hierarchical approach in which each factor has a main question followed by sub-questions. QUIS consists of demographic, overall system satisfaction and interface questions on terminology, screen factors, and system feedback, system capabilities, and learning factors, as well as questions on system components such as technical manuals, online tutorials, multimedia, voice recognition, virtual environments, Internet access, and software installation.[198] Each question measures the users' perceptions on a 9-point categorical scale. Additional space is provided to allow the users to make comments on the interface factors. QUIS is available in two different formats: a long form with 71 questions and a short form with 27 questions. The questionnaire is also available in an online format. The overall reliability of QUIS has a Cronbach's alpha of 0.94[197] and 0.95.[198] QUIS is a proprietary questionnaire with licensing fees. The questionnaire has been used in many usability evaluation studies in health care. QUIS has been used to study overall user satisfaction with a newly implemented EMR system;[199] in a formative pre- and post-usability evaluation to determine physician acceptance of a redesigned EMR to determine overall user satisfaction with the system and user satisfaction with screen layout, terminology used, system learnability, and system capabilities;[200] to determine user satisfaction with an electronic discharge summary in a pediatric intensive care unit;[201] to determine nurses' satisfaction with an electronic medication administration record prototype;[202] to determine nurses' evaluation of a decision support computer program for pain management;[203] in the evaluation of a nurse practitioner outcomes database used to track client variables for evaluating clinical outcomes;[204] to evaluate user satisfaction with physician order entry systems;[205] and to determine primary care physician satisfaction with EMRs.[206]

Computer System Usability Questionnaire (CSUQ)

The Computer System Usability Questionnaire (CSUQ) was also developed by IBM and is a slightly modified version of the Post-Study System Usability Questionnaire (PSSUQ).[207,208] The wording of the CSUQ refers to work situations and not usability testing situations, as PSSUQ does. It is a nonproprietary computer-based interface measurement tool.[188] CSUQ measures the same three factors as PSSUQ: system usefulness, informational quality, and interface quality. Overall, the CSUQ's coefficient alpha is 0.95, with coefficient alphas of 0.93 for system usefulness, 0.91 for informational quality, and 0.89 for interface quality.[209]The CSUQ has validity and reliability similar to that of the PSSUQ.[209] The questionnaire has been shown to work well outside of the usability laboratory. It has been used to determine physician satisfaction with effectiveness, efficiency, responsiveness, and learnability of a redesigned EMR system.[200] It has also been used to determine ease of use of a guideline-based decision support system;[210] and it has been used to compare users' overall satisfaction with and perception of system usefulness, information quality, and interface between an original and redesigned family history tracking and pedigree drawing program.[17]

Software Usability Measurement Inventory (SUMI)

The Software Usability Measurement Inventory (SUMI) was developed by the Human Factors Research Group.[211] SUMI is intended to be completed by users who have experience with the system being evaluated. SUMI consists of 50 items on an "agree, don't know, or disagree" three-point scale and provides an overarching usability rating across five subscales.[212] The coefficient alpha for SUMI is 0.89 and ranges from 0.71 to 0.92 for each of the questionnaire's five subscales: efficiency, affect, helpfulness, control, and learnability.[189] Efficiency is a measurement of the users' perception on how the software assists them in their work. Affect measures how well the users like the software. Helpfulness measures the adequacy of the system's help function and documentation as well as the degree to which the software is self-explanatory. Control is a measurement of the degree to which the users feel like they are in control of the software, and learnability measures the speed with which users feel like they can learn the software. SUMI is a proprietary questionnaire.[188] SUMI has not been widely used in the evaluation of health IT systems, but Narasimhadevara et al. used it to evaluate an interactive transplant nursing system in a hospital setting.[213]

After Scenario Questionnaire (ASQ)

The After Scenario Questionnaire (ASQ) was developed by IBM and measures user satisfaction with a three-item scale.[207,214,215] This brief nonproprietary questionnaire is designed to be completed directly following scenario usability studies.[188,216] The items are measured on a seven-point Likert scale that is anchored at the end points with "strongly agree" and "strongly disagree." The items measure ease of task completion, time required to complete the tasks, and satisfaction with support information. It should be noted that the findings from the psychometric evaluation of ASQ have limited generalizability due to the less than optimal sample size for the factor analysis .[215] The coefficient alpha for ASQ is 0.93.[207] There was no evidence of use of ASQ to measure the usability of health information systems.

Usefulness, Satisfaction, and Ease of Use Questionnaire (USE)

The Usefulness, Satisfaction, and Ease of Use (USE) questionnaire was developed by Lund and can be used to evaluate any interface.[188] It is a nonproprietary 30-item questionnaire that measures usefulness, satisfaction, and ease of use of an interface using a seven-point Likert agreement scale.[217] Although Lund reports high Cronbach's alphas for the questionnaire's items, he does not provide the actual measurements. However, according to Lund a partial correlation suggests that two of the factors measured by this questionnaire, usefulness and ease of use, influence each other, meaning that improvement in one will influence the other. Based on results from the literature review, it does not appear this questionnaire has been used to evaluate health information systems.

Perdue Usability Testing Questionnaire (PUTQ)

The Perdue Usability Testing Questionnaire (PUTQ) is a 100-item questionnaire that is based on human information processing theory and includes eight factors.[218] These factors are compatibility, consistency, flexibility, learnability, minimal action, minimal memory load, perceptual limitation, and user guidance. The reported Cronbach's alpha's ranged from 0.59 to 0.81 with an average of 0.70.[218] The PUTQ can be used to evaluate any interface, but the items in the questionnaire are primarily focused on graphical user interfaces.[218] There is no evidence of its use to evaluate health information systems.

End-User Computing Satisfaction Questionnaire (EUCS)

The End-User Computing Satisfaction Questionnaire (EUCS) was developed by Doll and Torzadeh in 1988. EUCS is a multidimensional instrument and measures user satisfaction with all types of applications.[219] The questionnaire uses a five-point Likert scale. It is a reliable and valid instrument that can be used as a general measure of user satisfaction with an information system.[220,221] EUCS has 12 items that measure five components of end-user satisfaction: content, accuracy, format, ease of use, and timeliness. Doll and Torzadeh report the reliability of EUCS to be 0.92. The reliability of each factor was 0.89 for content, 0.91 for accuracy, 0.78 for format, 0.85 for ease of use, and 0.82 for timeliness.[219] This instrument has been applied to various information technologies such as email applications[222] and decision support.[223] In a cross-sectional survey, the instrument was used to compare the use of an EMR system to complete clinical tasks and showed moderate satisfaction with the EMR.[224]

Table 3. Characteristics of usability questionnaires reviewed

	SUS[188]	SUMI[189]	ASQ[215]	CSUQ[207]	QUIS[197]		USE[217]	PUTQ[218]	EUCS[136]
Survey Length (number of items)	10	50	3	19	27/71		30	100	12
Reliability	0.85-0.91	0.89	0.93	0.95	0.94		NR*	0.59-0.81	0.92
Validity	Convergent	Construct	Concurrent	Content & Construct	External		NR	Content & Criterion	Convergent & Discriminant
Overall Usability	X			X					
Learnability	X	X			X				
Efficiency		X	X						X
Helpfulness		X							
Control		X							
Effectiveness			X		X				X
Satisfaction		Affect	X		X		X		
Usefulness				X			X		
Ease of Use							X		X
Information quality				X					
Interface quality				X	X				
Demographics					X				
Interface factors**					X				
System components***					X				
Compatibility								X	
Consistency								X	
Flexibility								X	
Minimal action								X	
Minimal memory load								X	
Perceptual limitation								X	
User guidance								X	
Content									X
Format									X

* NR = not reported.

** Interface factors include terminology and system feedback, screen, learning, and system capabilities.

***Components include technical manuals, online tutorials, multimedia, voice recognition, virtual environments, Internet access, and software installations.

Questionnaire Recommendations

The questionnaires reviewed measure a range of factors as outlined in Table 3, and all use Likert scales to determine strength of agreement. Some questionnaires measure few factors such as the System Usability Scale (SUS), which simply provides a measure of overall system usability, while others, such as the Questionnaire for User Interaction Satisfaction (QUIS) measure many different factors. All of the questionnaires included in this review had published psychometrics with the exception of the Usefulness, Satisfaction, and Ease of Use (USE) questionnaire. All of the reliabilities of the remainder of the questionnaires exceeded 0.80 with the exception of the Perdue Usability Testing Questionnaire (PUTQ), which had an average reliability of 0.70. With the exception of USE, all questionnaires have some measurement of validity. Although measurement of usability goes beyond subjective ratings of a system's usability and should include measurable outcomes such as task completion time, error rate, ease of learning, and so forth, obtaining these other measurable criteria requires usability expertise and is not realistic to include in the toolkit.

All of the usability questionnaires reviewed in this chapter were developed to measure overall system usability or measure general usability issues—none were developed specifically for EHRs. We summarize some of the advantages and disadvantages with each questionnaire we reviewed as follows:

- SUS is a short nonproprietary questionnaire, could be easily self-administered, and has good reliability and validity. However, it only measures overall usability and is not comprehensive enough for determining usability issues with an EHR.

- SUMI is more comprehensive than SUS, has identified reliability and validity, but it is proprietary and, again, not specific to an EHR.

- ASQ is a short three-item usability questionnaire that assesses user satisfaction after completing scenarios. Although it has good reliability and validity, it is not comprehensive enough for use in the usability toolkit.

- USE does not have any published psychometric evaluation data.

- PUTQ is a 100-item questionnaire with variable factor reliability and could be used to evaluate any interface. There was no evidence of its use in evaluating EHRs.

- EUCS is a short nonproprietary questionnaire that can be easily self-administered and has good reliability and validity. However, it is not comprehensive enough for determining specific usability issues with the EHR.

- CSUQ is a short nonproprietary questionnaire, could be easily self-administered, and has the best reliability metrics of all of the questionnaires reviewed. However, it only measures three factors of usability: system usefulness, informational quality, and interface quality.

- QUIS is a proprietary questionnaire that measures not only quantitative usability issues, but also includes a qualitative component. Its reliability is very near the reliability of CSUQ, but measures more factors than CSUQ.

Based on this outline of the questionnaires, none in their current state could adequately assess the usability problems specific to an EHR as outlined in Chapter 4—an important criterion in assessing usability methods. Therefore, we are recommending the development of a set of EHR usability questions that target known EHR usability issues and primary care setting attributes.

Summary. This chapter reviews nine usability questionnaires (Appendix C contains links to and some samples of usability questionnaires reviewed in our analysis), and assesses their advantages, disadvantages, and appropriateness for inclusion in the toolkit.

Chapter 8: Conclusions

The purpose of this report is to inform the development of a toolkit for primary care providers to assess EHR usability. Each chapter provides either a component of the framework for the toolkit or directs the approach.

Chapter 2, Defining EHRs and Usability, provides two core aspects of the conceptual framework for the toolkit. The EHR functionalities (i.e., health information display, order entry, results management, CDS, electronic communications/coordination, patient support, reporting and population health management, and administrative process) serve as the clinical framework for the toolkit. The TURF framework (useful, usable, and satisfying) serves as the usability framework.

The findings of Chapter 3, Primary Care Attributes, and Chapter 4, Usability Issues in EHRs, will help target high-risk areas for the toolkit to address. Moreover, the toolkit will be developed to help users flag usability issues known to impact patient safety and physician use.

The methods and tools to measure usability provide an opportunity to assess the different approaches that the toolkit might employ, each having its limitations and benefits (Chapter 5). However, given the requirements for the toolkit (Chapter 6), it is clear that the only method that can be incorporated in a toolkit for clinicians is the questionnaire. Unfortunately, none of the currently available questionnaires have been developed for evaluating the usability of an EHR, nor have any been validated to use in their evaluation (Chapter 7).

This toolkit we propose would consist of three parts: analysis, education, and tailoring. The analysis component would consist of questions sets addressing a usability issue in the appropriate clinical context. The education component would provide information on that usability issue and why it is important to a clinician to identify and understand. Finally, the tailoring component will provide a summary that clinicians can share with vendors and EHR certifying bodies to complement their user-centered design programs.

References

1. Armijo D, McDonnell C, Werner K. Electronic health record usability: Interface design considerations. Rockville (MD): Agency for Healthcare Research and Quality; 2009. AHRQ Publication No. 09(10)-0091-2-EF.

2. Armijo D, McDonnell C, Werner K. Electronic health record usability: Evaluation and use case framework. Rockville (MD): Agency for Healthcare Research and Quality; 2009. AHRQ Publication No. 09(10)-0091-1-EF.

3. Committee on Quality of Health Care in America, Institute of Medicine. Crossing the quality chasm: A new health system for the 21st century. Washington, DC: National Academies Press; 2001.

4. Ford EW, Menachemi N, Phillips MT. Predicting the adoption of electronic health records by physicians: When will health care be paperless? J Am Med Inform Assoc 2006;13(1):106-12.

5. Overhage JM, Middleton B, Miller RA, et al. Does national regulatory mandate of provider order entry portend greater benefit than risk for health care delivery? The 2001 ACMI debate. J Am Med Inform Assoc 2002;9(3):199-208.

6. Coble JM, Karat J, Orland MJ, et al. Iterative usability testing: Ensuring a usable clinical workstation. In: Masys DR, editor. The emergence of 'Internetable' health care. Systems that really work. Proceedings of the AMIA Annual Fall Symposium; 1997 Oct 25-29; Nashville, TN. Philadelphia (PA): Hanley & Belfus, Inc.; 1997. p. 744-8.

7. Lun KC. New user interfaces. Int J Biomed Comput 1995;39(1):147-50.

8. Tang PC, Patel VL. Major issues in user interface design for health professional workstations: Summary and recommendations. Int J Biomed Comput 1994;34(1-4): 139-48.

9. Terazzi A, Giordano A, Minuco G. How can usability measurement affect the re-engineering process of clinical software procedures? Int J Med Inform 1998;52(1-3):229-34.

10. Bates DW, Kuperman GJ, Rittenberg E, et al. A randomized trial of a computer-based intervention to reduce utilization of redundant laboratory tests. Am J Med 1999;106(2):144-50.

11. Tierney WM, McDonald CJ, Martin DK, et al. Computerized display of past test results effect on outpatient testing. Ann Intern Med 1987;107(4):569-74.

12. Linder J, Schnipper JL, Tsurikova R, et al. Barriers to electronic health record use during patient visits. In: Bates DW, editor. Biomedical and Health Informatics: From Foundations to Applications to Policy. Proceedings of the AMIA Annual Fall Symposium; 2006 Nov 11-15; Washington, DC. Red Hook (NY): Curran Associates, Inc.; 2006. p. 499-503.

13. Zhang J. Human-centered computing in health information systems. Part 1: analysis and design. J Biomed Inform 2005;38(1):1-3.

14. Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: The nature of patient care information system-related errors. J Am Med Inform Assoc 2004;11(2):104-12.

15. Campbell EM, Sittig DF, Ash JS, et al. Types of unintended consequences related to computerized provider order entry. J Am Med Inform Assoc 2006;13(5):547-56.

16. Koppel R, Metlay JP, Cohen A, et al. Role of computerized physician order entry systems in facilitating medication errors. JAMA 2005;293:1197-203.

17. Johnson CM, Johnson TR, Zhang J. A user-centered framework for redesigning health care interfaces. J Biomed Inform 2005;38(1):75-87.

18. Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. J Biomed Inform 2004;37(1):56-76.

19. Patel VL, Kushniruk AW, Yang S, et al. Impact of a computer-based patient record system on data collection, knowledge, organization, and reasoning. J Am Med Inform Assoc 2000;7(6):569-85.

20. Rose AF, Schnipper JL, Park ER, et al. Using qualitative studies to improve the usability of an EMR. J Biomed Inform 2005;38(1):51-60.

21. Zhang J, Johnson TR, Patel VL, et al. Using usability heuristics to evaluate patient safety of medical devices. J Biomed Inform 2003;36(1):23-30.

22. Andre TS, Hartson HR, Belz SM, et al. The user action framework: A reliable foundation for usability engineering support tools. Int J Hum-Comput St 2001;54:107-36.

23. National Institute of Standards and Technology, Industry Usability Reporting Project Working Group. Common industry specifications for usability - Requirements. Gaithersburg (MD): National Institute of Standards and Technology; 2007. NISTIR 7432.

24. Belden J, Grayson R, Barnes J. Defining and testing EMR usability: Principles and proposed methods of EMR usability evaluation and rating. Chicago (IL): Healthcare Information and Management Systems; 2009.

25. Garets D, Davis M. Electronic medical records vs. electronic health records: Yes, there is a difference. Chicago (IL): Healthcare Information and Management Systems Society; 2006. Updated January 26, 2006.

26. The National Alliance for Health Information Technology. Report to the Office of the National Coordinator for health information technology on defining key health information technology terms. Washington, DC: Department of Health & Human Services; 2008.

27. International Organization for Standardization. Health informatics - Electronic health record - Definition, scope and context. Geneva (CH): International Organization for Standardization; 2005. ISO 20514.

28. Available at: Healthcare Information and Management Systems. EHR topics and tools. http://www.himss.org/ASP/topics_ehr.asp. Accessed January 21, 2011.

29. Hayrinen K, Saranto K, Nykanen P. Definition, structure, content, use and impacts of electronic health records: A review of the research literature. Int J Med Inform 2008;77(5):291-304.

30. Institute of Medicine of the National Academies: Committee on Data Standards for Patient Safety. Key capabilities of an electronic health record system: Letter report. Washington, DC: National Academies Press; 2003.

31. Osheroff JA, Teich JM, Middleton B, et al. A roadmap for national action on clinical decision support. J Am Med Inform Assoc 2007;14(2):141-5.

32. Garg AX, Adhikari NKJ, McDonald H, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: A systematic review. JAMA 2005;293(10):1223-38.

33. Carayon P, Karsh B. Incorporating health information technology into workflow redesign: Summary report (Prepared by The Center for Quality and Productivity Improvement (CQPI), University of Wisconsin-Madison under Contract No. HHSA 290-2008-10036C). Rockville, MD: Agency for Healthcare Research and Quality, October 2010. AHRQ Publication No. 10-0098-EF.

34. Zhang J, Walji MF. TURF: Toward a unified framework of EHR usability. J Biomed Inform (In press); 2011.

35. Lowry S, Schumacher RM. NIST guide to the processes approach for improving the usability of electronic health records. Gaithersburg (MD): National Institute of Standards and Technology; 2010. NISTIR 7741.

36. Pelayo S, Leroy N, Guerlinger S, et al. Cognitive analysis of physicians' medication ordering activity. Stud Health Technol Inform 2005;116:929-34.

37. Wipfli R, Lovis C. Alerts in clinical information systems: Building frameworks and prototypes. Stud Health Technol Inform 2010;155:163-9.

38. Bodenheimer T. Primary care - Will it survive? N Engl J Med 2006;355(9):861-4.

39. Smith M, Dang D, Lee J. E-prescribing: clinical implications for patients with diabetes. J Diabetes Sci Technol 2009;3(5):1215-8.

40. Goh JM, Gao G, Agarwal R. Evolving work routines: Adaptive routinization of information technology in healthcare. Inform Syst Res 2011;Special issue: The role of information systems in healthcare organizations: Synergies from an interdisciplinary perspective(forthcoming).

41. Khajouei R, Jaspers MWM. CPOE system design aspects and their qualitative effect on usability. Stud Health Technol Inform 2008;136:309-14.

42. Krall MA, Sittig DF. Clinician's assessments of outpatient electronic medical record alert and reminder usability and usefulness requirements. In: Kohane IS, editor. Bio*medical Informatics: One Discipline. Proceedings of the AMIA Annual Fall Symposium; 2002 Nov 9-13; San Antonio, TX. Philadelphia (PA): Hanley & Belfus, Inc.; 2002. p. 400-4.

43. Kawamoto K, Houlihan CA, Balas EA, et al. Improving clinical practice using clinical decision support systems: A systematic review of trials to identify features critical to success. Br Med J 2005;330(7494):765-72.

44. Tamblyn R, Huang A, Kawasumi Y, et al. The development and evaluation of an integrated electronic prescribing and drug management system for primary care. J Am Med Inform Assoc 2006;13(2):148-59.

45. Wang SJ, Blumenfeld BH, Roche SE, et al. End of visit: Design considerations for an ambulatory order entry module. In: Kohane IS, editor. Bio*medical Informatics: One Discipline. Proceedings of the AMIA Annual Fall Symposium; 2002 Nov 9-13; San Antonio, TX. Philadelphia (PA): Hanley & Belfus, Inc.; 2002. p. 864-8.

46. Kushniruk AW, Triola MM, Borycki EM, et al. Technology induced error and usability: The relationship between usability problems and prescription errors when using a handheld application. Int J Med Inform 2005;74 (7-8):519-26.

47. Bailey R, Barnum C, Bosley J, et al. Research-based web design and usability guidelines. Washington, DC: U.S. Dept. of Health and Human Services; 2006.

48. Poon EG, Gandhi TK, Sequist TD, et al. "I wish I had seen this test result earlier!": Dissatisfaction with test result management systems in primary care. Arch Intern Med 2004;164(20):2223-8.

49. Shojania KG, Jennings A, Mayhew A, et al. Effect of point-of-care computer reminders on physician behaviour: A systematic review. Can Med Assoc J 2010;182(5): E216-25.

50. Bates DW, Cohen M, Leape LL, et al. Reducing the frequency of errors in medicine using information technology. J Am Med Inform Assoc 2001;8(4):299-308.

51. Phansalkar S, Edworthy J, Hellier E, et al. A review of human factors principles for the design and implementation of medication safety alerts in clinical information systems. J Am Med Inform Assoc 2010;17(5):493-501.

52. Cuvo JB, Dinh AK, Fahrenholz CG, et al. Quality data and documentation for EHRs in physician practice. J AHIMA 2008;79(8):43-8.

53. Ventres W, Kooienga S, Vuckovic N, et al. Physicians, patients, and the electronic health record: An ethnographic analysis. Ann Fam Med 2006;4(2):124-31.

54. Staggers N, Jennings BM, Lasome CEM. A usability assessment of AHLTA in ambulatory clinics at a military medical center. Mil Med 2010;175(7):518-24.

55. Van Vleck TT, Wilcox A, Stetson PD, et al. Content and structure of clinical problem lists: A corpus analysis. In: Unknown, editor. Biomedical and health informatics: From foundations to applications to policy. Proceedings of the AMIA Annual Fall Symposium; 2008 Nov 8-12; Washington, DC. Published online; 2008. p. 753-7.

56. Elder NC, McEwen TR, Flach J, et al. The management of test results in primary care: Does an electronic medical record make a difference? Fam Med 2010;45(5):327-33.

57. Tamblyn R, Huang A, Taylor L, et al. A randomized trial of the effectiveness of on-demand versus computer-triggered drug decision support in primary care. J Am Med Inform Assoc 2008;15(4):430-8.

58. Saleem JJ, Patterson ES, Militello L, et al. Impact of clinical reminder redesign on learnability, efficiency, usability, and workload for ambulatory clinic nurses. J Am Med Inform Assoc 2007;14(5):632-40.

59. Schumacher R, Webb J, Johnson K. How to select an electronic health record system that healthcare professionals can use. Oakbrook Terrace (IL): User Centric, Inc.; 2009. Version 1.1.

60. Aarts J, Ash J, Berg M. Extending the understanding of computerized physician order entry: Implications for professional collaboration, workflow and quality of care. Int J Med Inform 2007;76(suppl 1):S4-13.

61. Horsky J, McColgan K, Pang JE, et al. Complementary methods of system usability evaluation: Surveys and observations during software design and development cycles. J Biomed Inform 2010;43(5):782-90.

62. Trafton J, Martins S, Michel M, et al. Evaluation of the acceptability and usability of a decision support system to encourage safe and effective use of opioid therapy for chronic, noncancer pain by primary care providers. Pain Med 2010;11(4):575-85.

63. Zheng K, Padman R, Johnson MP, et al. An interface-driven analysis of user interactions with an electronic health records system. J Am Med Inform Assoc 2009;16(2):228-37.

64. Singh H, Mani S, Espadas D, et al. Prescription errors and outcomes related to inconsistent information transmitted through computerized order entry: A prospective study. Arch Intern Med 2009;169(10):982-9.

65. Hyun S, Johnson SB, Stetson PD, et al. Development and evaluation of nursing user interface screens using multiple methods. J Biomed Inform 2009;42(6):1004-12.

66. Grossman JM, Gerland A, Reed MC, et al. Physicians' experiences using commercial e-prescribing systems. Health Aff 2007;26(3):w393-404.

67. Orrico KB. Sources and types of discrepancies between electronic medical records and actual outpatient medication use. J Manag Care Pharm 2008;14(7):626-31.

68. Goldman RE, Dubé C, Lapane KL. Beyond the basics: Refills by electronic prescribing. Int J Med Inform 2010;79(7):507-14.

69. Peute LWP, Jaspers MWM. The significance of a usability evaluation of an emerging laboratory order entry system. Int J Med Inform 2007;76(2-3):157-68.

70. Yackel TR, Embi PJ. Unintended errors with EHR-based result management: A case series. J Am Med Inform Assoc 2010;17(1):104-7.

71. Chaffee BW, Zimmerman CR. Developing and implementing clinical decision support for use in a computerized prescriber-order-entry system. Am J Health Syst Pharm 2010;67(5):391-400.

72. Paterno MD, Maviglia SM, Gorman PN, et al. Tiering drug-drug interaction alerts by severity increases compliance rates. J Am Med Inform Assoc 2009;16(1): 40-6.

73. Shah NR, Seger AC, Seger DL, et al. Improving acceptance of computerized prescribing alerts in ambulatory care. J Am Med Inform Assoc 2006;13(1):5-11.

74. Steele AW, Eisert S, Witter J, et al. The effect of automated alerts on provider ordering behavior in an outpatient setting. PLoS Med 2005;2(9):e255.

75. Weingart SN, Toth M, Sands DZ, et al. Physicians' decisions to override computerized drug alerts in primary care. Arch Intern Med 2003;163(21):2625-31.

76. Kuperman GJ, Reichley RM, Bailey TC. Using commercial knowledge bases for clinical decision support. J Am Med Inform Assoc 2006;13(4):369-71.

77. Lamy JB, Venot A, Bar-Hen A, et al. Design of a graphical and interactive interface for facilitating access to drug contraindications, cautions for use, interactions and adverse effects. BMC Med Inform Decis Mak 2008;8:21.

78. Rosenbloom ST, Geissbuhler AJ, Dupont WD, et al. Effect of CPOE user interface design on user-initiated access to educational and patient information during clinical care. J Am Med Inform Assoc 2005;12(4):458-73.

79. Teich JM, Osheroff JA, Pifer EA, et al. Clinical decision support in electronic prescribing: Recommendations and an action plan: Report of the joint clinical decision support workgroup. J Am Med Inform Assoc 2005;12(4):365-76.

80. Osheroff JA. Improving medication use and outcomes with clinical decision support: A step-by-step guide. Chicago: Healthcare Information and Management Systems Society; 2009.

81. Hong Y, Ehlers K, Gillis R, et al. A usability study of patient-friendly terminology in an EMR system. Stud Health Technol Inform 2010;160(Pt 1):136-40.

82. Jaja C, Pares-Avila J, Wolpin S, et al. Usability evaluation of the interactive personal patient profile-prostate decision support system with african american men. J Natl Med Assoc 2010;102(4):290-7.

83. Marchionini G, Rimer BK, Wildemuth B. Evidence base for personal health record usability: Final report to the National Cancer Institute (Prepared by School of Information and Library Science, University of North Carolina at Chapel Hill under Supplement to Grant No.

R01 CA105786-03S1). Chapel Hill, NC: National Cancer Institute, February 2007.

84. Schnipper JL, Linder JA, Palchuk MB, et al. "Smart forms" in an electronic medical record: Documentation-based clinical decision support to improve disease management. J Am Med Inform Assoc 2008;15(4):513-23.

85. Nielsen J. Usability engineering. Boston: Academic Press; 1993.

86. Bevan N. UsabilityNet methods for user centred design. In: Jacko JA, Stephanidis C, editors. Human-Computer Interaction: Theory and Practice (part 1), Volume 1 (Human Factors and Ergonomics). Proceedings of HCI International 2003; 2003 Jun 22-27; Crete, GR. Mahwah (NJ): Lawrence Erlbaum Associates, Inc.; 2003. p. 434-8.

87. Mayhew DJ. The usability engineering lifecycle: A practitioner's handbook for user interface design. San Diego (CA): Academic Press; 1999.

88. Shneiderman B, Plaisant C. Designing the user interface: Strategies for effective human-computer interaction. 5th ed. Boston: Addison-Wesley; 2010.

89. Hackos JT, Redish JC. User and task analysis for interface design. New York: John Wiley & Sons, Inc.; 1998.

90. Kirwan B, Ainsworth LK. A guide to task analysis. Philadelphia: Taylor & Francis, Inc.; 1992.

91. Nielsen J, Mack RL. Usability inspection methods. New York: John Wiley & Sons, Inc.; 1994.

92. Dumas JS, Redish JC. A practical guide to usability testing. Portland (OR): Intellect Books; 1999.

93. Wharton C, Rieman J, Lewis C, et al. The cognitive walkthrough method: A practitioner's guide. In: Nielsen J and Mack RL, editors. Usability inspection methods. New York: John Wiley & Sons, Inc; 1994. p. 105-40.

94. Charlton SG, O'Brien TG. Handbook of human factors testing and evaluation. Mahwah (NJ): Lawrence Erlbaum Associates, Inc.; 2002.

95. Ericsson KA, Simon HA. Protocol analysis: Verbal reports as data revised edition. Cambridge (MA): The MIT Press; 1984.

96. Tullis TS, Albert B. Measuring the user experience: Collecting, analyzing and presenting usability metrics. Burlington (MA): Morgan Kaufman Publishers; 2008.

97. Association for the Advancement of Medical Instrumentation. Human factors engineering - Design of medical devices. Arlington (VA): Association for the Advancement of Medical Instrumentation; 2010. ANSI/AAMI HE75:2009.

98. Vicente KJ. Cognitive work analysis: Toward safe, productive, and healthy computer-based work. Mahwah (NJ): Lawrence Erlbaum Associates; 1999.

99. Nielsen J. Available at: Useit.com. Jakob Nielsen's website. http://www.useit.com/. Accessed February 25, 2011.

100. Available at: Usability.gov. Usability methods. http://www.usability.gov/methods/index.html. Accessed February 25, 2011.

101. Nielsen J. Guerilla HCI: Discount usability engineering to penetrate intimidation barrier. In: Bias RG and Mayhew DJ, editors. Cost-justifying usability. San Diego (CA): Academic Press; 1994. p. 245-72.

102. Thyvalikakath TP, Monaco V, Thambuganipalle H, et al. Comparative study of heuristic evaluation and usability testing methods. Stud Health Technol Inform 2009;143:322-7.

103. Molich R, Nielsen J. Improving a human-computer dialogue. Commun ACM 1990;33(3):338-48.

104. Tognazzini B. Available at: AskTog.com. First principles of information design. http://www.asktog.com/basics/firstPrinciples.html. Accessed February 11, 2011.

105. Jaspers MWM. A comparison of usability methods for testing interactive health technologies: Methodological aspects and empirical evidence. Int J Med Inform 2009;78(5):340-53.

106. Lindgaard G. Notions of thoroughness, efficiency, and validity: Are they valid in HCI practice? Int J Ind Ergonom 2006;36(12):1069-74.

107. Jeffries R, Miller J, Wharton C, et al. User interface evaluations in the real world: A comparison of four techniques. In: Robertson SP, Olson GM, Olson JS, editors.

Reaching through technology. CHI '91 proceedings of the SIGCHI conference on human factors in computing systems; 1991 Apr 27-May 2; New Orleans, LA. New York: ACM Press; 1991. p. 119-24.

108. Hartson HR. Human-computer interaction: Interdisciplinary roots and trends. J Syst Software 1998;43(2):103-18.

109. Beuscart-Zephir M, Pelayo S, Anceaux F, et al. Cognitive analysis of physicians and nurses cooperation in the medication ordering and administration process. Int J Med Inform 2007;76(1):65-77.

110. Choi J, Bakken S. Web-based education for low-literate parents in Neonatal Intensive Care Unit: Development of a website and heuristic evaluation and usability testing. Int J Med Inform 2010;79(8):565-75.

111. Karahoca A, Bayraktar E, Tatoglu E, et al. Information system design for a hospital emergency department: A usability analysis of software prototypes. J Biomed Inform 2010;43(2):224-32.

112. Sox CM, Gribbons WM, Loring BA, et al. Patient-centered design of an information management module for a personally controlled health record. J Med Internet Res 2010;12(3):e36.

113. Edwards PJ, Moloney KP, Jacko JA, et al. Evaluating usability of a commercial electronic health record: A case study. Int J Hum-Comput St 2008;66(10):718-28.

114. Rieman J, Franzke M and Redmiles D. Usability evaluation with the cognitive walkthrough. In: Katz I, Mack R, Marks L, editors. Mosaic of creativity. Proceedings of the CHI '95 Conference on human factors in computing systems; 1995 May 7-11; Denver, CO. New York: ACM Press; 1995. p. 387-8.

115. Polson P, Lewis C, Rieman J, et al. Cognitive walkthrough: A method for theory-based evaluation of user interfaces. Int J Man Mach Stud 1992;36(5):741-73.

116. Polson P, Lewis CH. Theory-based design for easily learned interfaces. Hum-Comput Interact 1990;5(2):191-220.

117. Sears A. Heuristic walkthroughs: Finding the problems without the noise. Int J Hum Comput Interact 1997;9(3):213-34.

118. Beuscart-Zéphir M, Elkin P, Pelayo S, et al. The human factors engineering approach to biomedical informatics projects: State of the art, results, benefits, challenges. Yearb Med Inform 2007;:109-27.

119. Cohen T, Kaufman D, White T, et al. Cognitive evaluation of an innovative psychiatric clinical knowledge enhancement system. Stud Health Technol Inform 2004;107(Pt 2):1295-9.

120. Khajouei R, de Jongh D, Jaspers M. Usability evaluation of a computerized physician order entry for medication ordering. Stud Health Technol Inform 2009;150:532-6.

121. Newman WM. On simulation, measurement, and piecewise usability evaluation. Hum Comput Interact 1998;13(3):316-23.

122. Maguire M. Methods to support human-centered design. Int J Hum-Comput St 2001;55:587-634.

123. Faulkner L. Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. Behav Res Meth Ins C 2003;35(3):379-83.

124. Lewis JR. Sample sizes for usability studies: Additional considerations. Hum Factors 1994;36(2):368-78.

125. Nielsen J and Landauer TK. A mathematical model of the finding of usability problems. In: Ashlund S, Mullet K, Henderson A, et al., editors. Bridges between worlds. CHI '93 Proceedings of the INTERACT '93 and CHI '93 conference on human factors in computing systems; 1993 Apr 24-29; Amsterdam, NL. New York: ACM Press; 1993. p. 206-13.

126. Virzi RA. Refining the test phase of usability evaluation: How many subjects is enough? Hum Factors 1992;34(4):457-68.

127. Kushniruk AW, Patel VL. Cognitive computer-based video analysis: its application in assessing the usability of medical systems. Medinfo 1995;8(Pt 2):1566-9.

128. Borycki EM, Lemieux-Charles L, Nagle L, et al. Novice nurse information needs in paper and hybrid electronic-paper environments: A qualitative analysis. Stud Health Technol Inform 2009;150:913-7.

129. Currie LM, Graham M, Allen M, et al. Clinical information needs in context: An observational study of clinicians while using a clinical information system. In: Musen M, editor. Biomedical and health informatics: From foundations to applications to policy. Proceedings of the 2003 AMIA Annual Fall Symposium; 2003 Nov 8-12; Washington, DC. Bethesda (MD): American Medical Informatics Association; 2003. p. 190-4.

130. Jaspers MWM, Steen T, Van Den Bos C, et al. The use of cognitive methods in analyzing clinicians' task behavior. St Heal T 2002;93:25-31.

131. Jaspers MW, Steen T, Bos Cvd, et al. The think aloud method: A guide to user interface design. Int J Med Inform 2004;73(11-12):781-95.

132. Lilholt LH, Pedersen SS, Madsen I, et al. Development of methods for usability evaluations of EHR systems. Stud Health Technol Inform 2006;124:341-6.

133. Wu RC, Orr MS, Chignell M, et al. Usability of a mobile electronic medical record prototype: A verbal protocol analysis. Inform Health Soc Care 2008;33(2): 139-49.

134. Carroll C, Marsden P, Soden P, et al. Involving users in the design and usability evaluation of a clinical decision support system. Comput Meth Prog Bio 2002;69(2):123-35.

135. Malhotra S, Laxmisan A, Keselman A, et al. Designing the design phase of critical care devices: A cognitive approach. J Biomed Inform 2005;38(1):34-50.

136. Boren MT, Ramey J. Thinking aloud: Reconciling theory and practice. IEEE T Prof Commun 2000;43(3):261-78.

137. Nisbett RE, Wilson TD. Telling more than we know: Verbal reports on mental processes. Psych Rev 1977;84(3):231-59.

138. Yarbus AL. Eye movements during perception of complex objects. In: Riggs LA, editor. Eye movements and vision. New York: Plenum Press; 1967. p. 171-96.

139. Kowler E, Anton S. Reading twisted text: Implications for the role of saccades. Vision Res 1987;27(1):45-60.

140. Bower GH, Clapper JP. Experimental methods in cognitive science. In: Posner MI, editor. Foundations of cognitive neuroscience. Cambridge (MA): The MIT Press; 1989. p. 245-300.

141. Ballard DH, Hayhoe M, Pelz J. Memory representations in natural tasks. J Cognitive Neurosci 1995;7(1):66-80.

142. Chase WG, Simon HA. Perception in chess. Cognitive Psychol 1973;4(1):55-81.

143. Just MA, Carpenter PC. Eye fixations and cognitive processes. Cognitive Psychol 1976;8:441-80.

144. West R, Lehman KR. Automated summative usability studies: An empirical evaluation. In: Grinter R, Rodden T, Aoki P, et al., editors. CHI '06 Conference on human factors in computing systems. CHI '06 proceedings of the SIGCHI conference on human factors in computing systems; 2006 Apr 24-27; Montreal, CA. New York: ACM Press; 2006. p. 631-9.

145. Castillo JC, Hartson HR, Hix D. Remote usability evaluation: Can users report their own critical incidents? In: Lund A, Karat C, editors. Making the impossible possible. CHI '98 conference proceedings on human factors in computing systems; 1998 Apr 18-23; Los Angeles, CA. New York: ACM Press; 1998. p. 253-4.

146. Dixon BE. Enhancing the informatics evaluation toolkit with remote usability testing. In: Unknown, editor. Biomedical and health informatics: From foundations to applications to policy. Proceedings of the AMIA Annual Fall Symposium; 2009 Nov 14-18; San Francisco, CA. Bethesda (MD): American Medical Informatics Association; 2009. p. 147-51.

147. Petrie H, Hamilton F, King N, et al. Remote usability evaluations with disabled people. In: Grinter R, Rodden T, Aoki P, et al., editors. CHI '06 Conference on human factors in computing systems. CHI '06 proceedings of the SIGCHI conference on human factors in computing systems; 2006 Apr 22-27; Montreal, CA. New York: ACM Press; 2006. p. 1133-41.

148. Winckler MAA, Freitas CMDS, Valdeni de Lima J. Usability remote evaluation for WWW. In: Turner T, Szwillus G, Czerwinski M, et al., editors. CHI '00 extended abstracts on human factors in computing systems. CHI '00 Proceedings of the SIGCHI conference on human factors in computing systems; 2000 Apr 1-6; The Hague, NL. New York: ACM Press; 2000. p. 131-2.

149. Andreasen MS, Nielsen HV, Schroder SO, et al. What happened to remote usability testing? An empirical study of three methods. In: Rosson MB, editor. CHI '07 Conference on human factors in computing systems. CHI '07 Proceedings of the SIGCHI conference on human factors in computing systems; 2007 Apr 30-May 3; San Jose, CA. New York: ACM Press; 2007. p. 1405-14.

150. Bruun A, Gult P, Hofmeister L, et al. Let your users do the testing: A comparison of three remote asynchronous usability testing methods. In: Olsen DR, Arthur RB, editors. CHI '09 Conference on human factors in computing systems. CHI '09 Proceedings of the 27th international conference on human factors in computing systems; 2009 Apr 4-9; Boston, MA. New York: ACM Press; 2009. p. 1619-28.

151. Kushniruk AW, Borycki EM, Kuwata S, et al. Emerging approaches to evaluating the usability of health information systems. In: Kushniruk AW and Borycki EM, editors. Human, social, and organizational aspects of health information systems. Hershey (PA): IGI Global; 2008. p. 1-22.

152. Kushniruk AW, Ho F. The virtual usability laboratory: Evaluating web-based health systems. In proceeding of e-Health 2004, Victoria, BC.

153. Kushniruk AW, Patel C, Patel VL, et al. Televaluation of clinical information systems: An integrative approach to assessing web-based systems. Int J Med Inform 2001;61(1):45-70.

154. Hartson HR, Castillo JC, Kelso J. Remote evaluation: The network as an extension of the usability laboratory. In: Tauber MJ, editor. Common ground. CHI '96 Proceedings of the SIGCHI conference on human factors in computing systems; 1996 Apr 13-18; Vancouver, CA. New York: ACM Press; 1996. p. 228-35.

155. Hartson HR, Castillo JC. Remote evaluations for post-deployment usability improvement. In: Catarci T, Costabile MF, Santucci G, et al., editors. 4th International working conference on advanced visual interfaces. AVI '98 Proceedings of the working conference on advanced visual interfaces; 1998 May 24-27; L'Aquila, IT. New York: ACM Press; 1998. p. 22-9.

156. John BE, Kieras DE. The GOMS family of user interface analysis techniques: Comparison and contrast. ACM T Comput-Hum Int 1996;3(4):320-51.

157. John BE, Vera AH, Newell A. Toward real-time GOMS: A model of expert behavior in a highly interactive task. Behav Inform Technol 1994;13(4):255-267.

158. John B, Kieras D. Using GOMS for user interface design and evaluation: Which technique? ACM T Comput-Hum Int 1996;3(4):287-319.

159. Kieras D. GOMS models for task analysis. In: Diaper D and Stanton N, editors. The handbook of task analysis for human-computer interaction. Mahwah (NJ): Lawrence Erlbaum Associates; 2004. p. 83-116.

160. Card SK, Moran TP, Newell A. The psychology of human-computer interaction. Hillsdale (NJ): Lawrence Erlbaum Associates; 1983.

161. Kieras D. Available at: Ann Arbor: Department of Psychology, University of Michigan. Using the keystroke-level model to estimate execution times. Unpublished report. ftp://141.212.113.110/people/kieras/GOMS/KLM.pdf. Accessed February 11, 2011.

162. Endoh A, Minato K, Komori M, et al. Quantitative comparison of human computer interaction for direct prescription entry systems. Medinfo 1995;8(Pt 2):1101-5.

163. Saitwal H, Feng X, Walji M, et al. Assessing performance of an electronic health record (EHR) using cognitive task analysis. Int J Med Inform 2010;79(7):501-6.

164. Flanagan JC. The critical incident technique. Psychol Bull 1954;51(4):327-58.

165. Joint Commission on Accreditation of Healthcare. Team identifies ways to mitigate risks in new accreditation process. Jt Comm Perspect 2004;24(1):15-6.

166. Kim GR, Chen AR, Arceci RJ, et al. Error reduction in pediatric chemotherapy: Computerized order entry and failure modes and effects analysis. Arch Pediatr Adolesc Med 2006;160(5):495-8.

167. Carroll JM, Koenemann J, Rosson MB, et al. Critical incidents and critical threads in empirical evaluation. In: Alty J, Diaper D, Guest SP, editors. People and Computers VIII. Proceedings of the HCI '93 Conference; 1993 Aug 8-13; Orlando, FL. Cambridge (MA): Cambridge University Press; 1993. p. 279-92.

168. Maddox ME. Available at: Federal Aviation Administration. Chapter 5: Facility design. https://hfskyway.faa.gov/(A(uPo9A-guywEkAAAAZGUzZmRlM2MtMDY4YS00YzJkLTk4MmQtZDZjNjUwZjk4NWU20t5zCUX8V7snznPEaSLy0V1-60o1))/HFTest/Bibliography%20of%20Publications%5CHuman%20Factor%20Maintenance%5CHuman%20Factors%20Guide%20for%20Aviation%20Maintenance%20-%20Chapter%205%20%20Facility%20Design.pdf. Accessed February 25, 2011.

169. Westbrook JI, Coeira EW, Gosling AS, et al. Critical incidents and journey mapping as techniques to evaluate the impact of online evidence retrieval systems on health care delivery and patient outcomes. Int J Med Inform 2007;76(2-3):234-45.

170. Kramer AF, Sirevaag EJ, Braune R. A psychophysiological assessment of operator workload during simulated flight missions. Hum Factors 1987;29(2):145-60.

171. Chandler P, Sweller J. Cognitive load theory and the format of instruction. Cognition Instruct 1991;8(4):293-332.

172. Sweller J. Cognitive load during problem-solving: effects on learning. Cognitive Sci 1988;12(2):257-85.

173. Bevan N, Macleod M. Usability measurement in context. Behav Inform Technol 1994;13(1):132-45.

174. Embrey D, Balckett C, Marsden P, et al. Development of a human cognitive workload assessment tool. Dalton (UK): Human Reliability Associates; 2006. MCA Final Report.

175. Reid GB, Nygren TE. The subjective workload assessment technique: A scaling procedure for measuring mental workload. In: Hancock PA and Meshkati N, editors. Human mental workload. Amsterdam (NL): North-Holland Press; 1988. p. 185-218.

176. Luximon A, Goonetilleke RS. Simplified subjective workload assessment technique. Ergonomics 2001;44(3):229-43.

177. DiDomenico A, Nussbaum MA. Interactive effects of physical and mental workload on subjective workload assessment. Int J Ind Ergonom 2008;38(11-12):977-83.

178. Rubio S, Diaz E, Martin J, et al. Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. Appl Psychol-Int Rev 2004;53(1):61-86.

179. Hart SG, Staveland LE. Development of the NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Hancock PA and Meshkati N, editors. Human mental workload. Amsterdam (NL): North-Holland Press; 1988. p. 139-84.

180. Hertzum M, Simonsen J. Positive effects of electronic patient records on three clinical activities. Int J Med Inform 2008;77(12):809-17.

181. Bevan N. Measuring usability as quality of use. Software Qual J 1995;4(2):115-30.

182. Sauro J, Dumas JS. Comparison of three one-question, post-task usability questionnaires. In: Olsen DR, Arthur RB, editors. CHI '09 Conference on human factors in computing systems. CHI '09 Proceedings of the 27th international conference on human factors in computing systems; 2009 Apr 4-9; Boston, MA. New York: ACM Press; 2009. p. 1599-608.

183. Hassenzahl M, Sandweg N. From mental effort to perceived usability: Transforming experiences into summary assessments. In: Dykstra-Erickson E, Tscheligi M, editors. Connect. CHI '04 Conference on human factors in computing systems; 2004 Apr 24-29; Vienna, AT. New York: ACM Press; 2004. p. 1283-6.

184. Kirakowski J, Cierlik B. Measuring the usability of web sites. In: Unknown, editor. Human-system interaction: The sky's no limit. Human factors and ergonomics society 42nd annual meeting; 1998 Oct 5-9; Santa Monica, CA. Chicago: Human factors and ergonomics society; 1998. p. 424-8.

185. Cockton G, Lavery D, Woolrych A. Inspection-based evaluations. In: Jacko JA and Sears A, editors. Human-computer interaction: Development process. Mahwah (NJ): Lawrence Erlbaum Associates, Inc.; 2003. p. 1118-38.

186. Nunnally JC, Bernstein I. Psychometric theory (3rd ed.). New York: McGraw-Hill; 1994.

187. Brooke J. SUS - A quick and dirty usability scale. In: Jordon PW, Thomas B, Weerdmeester BA, et al., editors. Usability evaluation in industry. London: Taylor & Francis; 1996. p. 189-94.

188. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput Interact 2008;24(6):574-94.

189. Kirakowski J. Available at: SUMI Questionnaire Homepage. The use of questionnaire methods for usability assessment (unpublished manuscript). http://sumi.ucc.ie/sumiapp.html. Accessed February 10, 2011.

190. Lewis JR and Sauro J. The factor structure of the system usability scale. In: Kurosu M, editor. Human Computer Interaction International Conference (HCII 2009). Proceedings of the 1st International Conference on Human Centered Design; 2009 Jul 19-24; San Diego, CA. Heidelberg (DE): Springer-Verlag Berlin; 2009. p. 94-103.

191. Borsci S, Federici S, Lauriola M. On the dimensionality of the System Usability Scale: A test of alternative measurement models. Cogn Process 2009;10(3):193-7.

192. Delasobera BE, Dubin J, Childs C, et al. Effect on emergency department patient satisfaction of a simple in-room information computer display. Ann Emerg Med 2010;56(3):S121.

193. Sawka AM, Straus S, Gafni A, et al. A usability study of a computerized decision aid to help patients with, early stage papillary thyroid carcinoma in, decisionmaking on adjuvant radioactive iodine treatment. Patient Educ Couns 2010;Aug 21:Epub ahead of print.

194. Klimov D, Shahar Y, Taieb-Maimon M. Intelligent visualization and exploration of time-oriented data of multiple patients. Artif Intell Med 2010;49(1):11-31.

195. Otero A, Felix P, Barro S, et al. Addressing the flaws of current critical alarms: A fuzzy constraint satisfaction approach. Artif Intell Med 2009;47(3):219-38.

196. Lutes KD, Chang K and Baggili IM. Diabetic e-management system (DEMS). In: Latifi S, editor. New Generations. ITNG '06 Proceedings of the Third International Conference on Information Technology; 2006 Apr 10-12; Las Vegas, NV. Washington, DC: IEEE Computer Society; 2006. p. 619-24.

197. Chin JP, Diehl VA and Norman KL. Development of an instrument measuring user satisfaction of the human-computer interface. In: Soloway E, Frye D, Sheppard SB, editors. CHI '88 Conference on human factors in computing systems. CHI '88 Proceedings of the SIGCHI conference on Human factors in computing systems; 1988 May 15-19; Washington, DC. New York: ACM Press; 1988. p. 213-8.

198. Harper B, Slaughter L, Norman K. Questionnaire administration via the WWW: A validation & reliability study for a user satisfaction questionnaire. In: Lobodzinski S, Tomek I, editors. World Conference on the WWW, Internet & Intranet. Proceedings of WebNet 97; 1997 Nov 1-5; Toronto, CA. Charlottesville (VA): AACE; 1997. p. 808-10.

199. Al-Azmi SF, Al-Enezi N, Chowdhury RI. Users' attitudes to an electronic medical record system and its correlates: A multivariate analysis. HIM J 2009;38(2):34-40.

200. Jaspers MWM, Peute LWP, Lauteslager A, et al. Pre-post evauation of physicians' satisfaction with a redesigned electronic medical record system. Stud Health Technol Inform 2008;136:303-8.

201. Burkle T, Engel PA. When usage and user satisfaction differ: The case of an electronic discharge summary. Stud Health Technol Inform 2007;12(pt2):1093-7.

202. Staggers N, Kobus D, Brown C. Nurses' evaluation of a novel design for an electronic medication administration record. Comput Inform Nurs 2007;25(2):67-75.

203. Im EO, Chee W. Evaluation of the decision support computer program for cancer pain management. Oncol Nurs Forum 2006;33(5):977-82.

204. Hortman PA, Thompson CB. Evaluation of user interface satisfaction of a clinical outcomes database. Comput Inform Nurs 2005;23(6):301-7.

205. Murff HJ, Kannry J. Physician satisfaction with two order entry systems. J Am Med Inform Assoc 2001;8(5):499-509.

206. Sittig DF, Kuperman GJ and Fiskio J. Evaluating physician satisfaction regarding user interactions with an electronic medical record system. In: Lorenzi NM, editor. Transforming health care through informatics: Cornerstones for a new information management paradigm. Proceedings of the AMIA Annual Fall Symposium; 1999 Nov 6-10; Washington, DC. Philadelphia (PA): Hanley & Belfus, Inc.; 1999. p. 400-4.

207. Lewis JR. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. Int J Hum Comput Interact 1995;7(1):57-78.

208. Lewis JR. Psychometric evaluation of the computer system usability questionnaire: The CSUQ. Indianapolis (IN): IBM; 1992. Technical Report 54.723.

209. Lewis JR. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. Indianapolis (IN): IBM; 1993. Technical Report 54.786.

210. Goud R, Jaspers MW, Hasman A, et al. Subjective usability of the CARDSS guideline-based decision support system. Stud Health Technol Inform 2008;136:193-8.

211. Kirakowski J, Corbett M. SUMI: The software measurement inventory. Brit J Educ Technol 1993;(24):210-2.

212. Van Veenendall E. Questionnaire based usability testing. In: Unknown, editor. EURO & Y2K: The industrial impact. Conference Proceedings of the European Software Quality Week; 1998 Nov 9-13; Brussels, BE. San Francisco (CA): Software Research, Inc.; 1998. p. 1-9.

213. Narasimhadevara A, Radhakrishnan T, Leung B, et al. On designing a usable interactive system to support transplant nursing. J Biomed Inform 2008;14(1):137-51.

214. Lewis JR. An after-scenario questionnaire for usability studies: Psychometric evaluation over three trials. Indianapolis (IN): IBM; 1991. Technical Report 54.584.

215. Lewis JR. A psychometric evaluation of an after-scenario questionnaire for computer usability studies: The ASQ. ACM SIGCHI Bull 1991;23(1):78-81.

216. Lewis JR. Usability testing. In: Salvendy G, editor. Handbook of human factors and ergonomics. Third Edition. Hoboken (NJ): John Wiley & Sons, Inc.; 2006. p. 1275-316.

217. Lund A. Available at: STC Usability SIG Newsletter. Measuring usability with the USE questionnaire. http://www.stcsig.org/usability/newsletter/0110_measuring_with_use.html. Accessed February 11, 2011.

218. Lin HX, Choong YY, Salvendy G. A proposed index of usability: A method for comparing the relative usability of different software systems. Behav Inform Technol 1997;16(4):267-78.

219. Doll WJ, Torkzadeh G. The measurement of end-user computing satisfaction. MIS Quar 1988;12(2):259-74.

220. Harrison AW, Rainer RK. A general measure of user computing satisfaction. Comput Human Behav 1996;12(1):79-92.

221. Doll WJ, Xia W, Torkzadeh G. A confirmatory factor analysis of the end-user computing satisfaction instrument. MIS Quart 1994;18(4):453-61.

222. Adams DA, Nelson RR, Todd PA. Perceived usefulness, ease of use, and usage of information technology: A replication. MIS Quart 1992;16(2):227-47.

223. McHaney R, Cronan TP and Hagmann C. Examination of validity in DSS success surrogates. In: Bartkus K, editor. Proceedings of DSI'98 Decision Sciences Institute Annual Meeting; 1998 Nov 21-24; Las Vegas, NV. Atlanta (GA): Decision Sciences Institute; 1998. p. 792-3.

224. Laerum H, Ellingsen G, Faxvaag A. Doctors' use of electronic medical records systems in hospitals: Cross sectional survey. BMJ 2001;323(7325):1344-8.

Appendix A: Technical Expert Panel

Jeffery L. Belden, M.D.
Associate Professor of Clinical Family and
Community Medicine
Affiliated Faculty, Information Experience
Laboratory, University of Missouri
School of Medicine

Edna Boone, M.A., CPHIMS
Senior Director, Healthcare Information
Systems, Healthcare Information and
Management Systems Society (HIMSS)

Keith A. Butler, Ph.D.
Principal Research Scientist, University of
Washington, Seattle

Joseph Cafazzo, Ph.D., P.Eng.
Senior Director, Healthcare Human Factors
Assistant Professor, IBBME and HPME,
Faculty of Medicine, University of
Toronto

Ben-Tzion Karsh, Ph.D.
Associate Professor, Department of
Industrial Engineering, University of
Wisconsin-Madison

Robert M. Kolodner, M.D.
Open Health Tools

Andre Kushniruk, Ph.D.
Professor, School of Health Information
Science, University of Victoria

Matthew Quinn, M.B.A.
Special Expert Health IT, Information
Technology Laboratory, Information
Access Division, National Institute of
Standards and Technology

Robert J. Schumacher, Ph.D.
Managing Director, User Centric

Nancy Staggers, Ph.D., R.N., FAAN
Director, Informatics Program, College of
Nursing, University of Utah
Adjunct Professor, Biomedical Informatics,
University of Utah
Professor, College of Nursing, University of
Utah

Jiajie Zhang, Ph.D.
Professor, University of Texas at Houston
Associate Dean for Research, University of
Texas at Houston
Principal Investigator and Co-director,
National Center for Cognitive
Informatics and Decision Making in
Healthcare
Co-director, Center for Cognitive
Informatics and Decisionmaking, School
of Biomedical Informatics, University of
Texas at Houston

Appendix B: Annotated References Describing Methods Cited

Adams DA, Nelson RR, Todd PA. Perceived usefulness, ease of use, and usage of information technology: A replication. MIS Quart 1992;16(2):227-47.

Adams et al. report on two studies which focus on the psychometric properties of ease-of-use and usefulness scales related to attitudes toward voice and electronic mail and the ability to show discriminate validity. These studies supported the validity and reliability of the ease-of-use and usefulness scales and their ability to rate these properties among different software packages.

Al-Azmi SF, Al-Enezi N, Chowdhury RI. Users' attitudes to an electronic medical record system and its correlates: A multivariate analysis. HIM J 2009;38(2):34-40.

The questionnaire for user interaction satisfaction tool was used to evaluate medical receptionists' responses to a newly implemented electronic medical record system in primary care centers. The majority of the respondents believed that the new system was easy, flexible, and satisfying, while more than one-third of the respondents found the system to be inadequate.

Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput Interact 2008;24(6):574-94.

This article presents nearly 10 years of system usability scale (SUS) data collected while conducting 206 studies and 2,324 individual surveys. Results showed the system usability scale to be very robust. This presents a statement-by-statement analysis, factor analysis, and reliability analysis, and describes the various uses of SUS, the specific details of the SUS score, and explains a modification to SUS.

Boren MT, Ramey J. Thinking aloud: Reconciling theory and practice. IEEE T Prof Commun 2000;43(3):261-78.

The authors review how use of think-aloud protocols differs from the seminal work of Ericsson and Simon. They present how their new approach deals with problems that require interventions which are not outlined by Ericsson and Simon.

Borsci S, Federici S, Lauriola M. On the dimensionality of the System Usability Scale: A test of alternative measurement models. Cogn Process 2009;10(3):193-7.

The authors tested the unidimensional model of SUS and the two-factor model of SUS proposed by Lewis and Sauro (usability and learnability), and did not find a fit to the data. They hypothesized that usability and learnability are independent factors of SUS.

Brooke J. SUS - A quick and dirty usability scale. In: Jordon PW, Thomas B, Weerdmeester BA, et al., editors. Usability evaluation in industry. London: Taylor & Francis; 1996. p. 189-94.

Brooke, the founder of the System Usability Scale (SUS), discusses this robust and reliable scale and the technique of selecting questions to elicit extreme expressions of attitude when constructing the Likert based scale. The scale, which is administered after participants have used the system, but prior to discussion about the system, provides a usability score from 0 to 100.

Burkle T, Engel PA. When usage and user satisfaction differ: The case of an electronic discharge summary. Stud Health Technol Inform 2007;12(pt2):1093-7.

Burkle and Engel employed a longitudinal study over an 11-month period utilizing system logs and questionnaires including a modified questionnaire of user satisfaction (QUIS) to assess system use and user satisfaction before and after introduction of an electronic discharge letter application in a German university hospital. As compared to a previous study in an Austrian hospital using comparable methods, the researchers noted that user satisfaction was lower as use of the new discharge letter increased.

Chin JP, Diehl VA, Norman KL. Development of an instrument measuring user satisfaction of the human-computer interface. In: Soloway E, Frye D, Sheppard SB, editors. CHI '88 Conference on human factors in computing systems. CHI '88 Proceedings of the SIGCHI conference on Human factors in computing systems; 1988 May 15-19; Washington, DC. New York: ACM Press; 1988. p. 213-8.

Chin et al. discuss their part of the research effort to develop the Questionnaire for User Interface Satisfaction (QUIS) using two pairs of software categories: liked/disliked and the standard command line system (CLS) and a menu-driven application (MDA). For the QUIS, which measures subjective user ratings of the human

computer interface, the study established external validity, but did not establish construct or predictive validity. The reliability of QUIS was 0.94.

Doll WJ, Torkzadeh G. The measurement of end-user computing satisfaction. MIS Quar 1988;12(2): 259-74.

Doll and Torkzadeh report on the development and testing of an instrument to measure end-user computer satisfaction, which showed construct and discriminant instrument validity, adequate reliability, as well as criterion-related validity. End-user design involvement was positively correlated with end-user satisfaction and the authors noted that further validation was necessary.

Doll WJ, Xia W, Torkzadeh G. A confirmatory factor analysis of the end-user computing satisfaction instrument. MIS Quart 1994;18(4):453-61.

Doll and Torkzadeh validate the end-user computing satisfaction instrument. The cross-validation study shows evidence of a multifaceted construct that consists of five subscales. These subscales are content, accuracy, format, ease of use, and timeliness. The items show adequate reliability and validity.

Goud R, Jaspers MW, Hasman A, et al. Subjective usability of the CARDSS guideline-based decision support system. Stud Health Technol Inform 2008;136:193 8.

A modified version of the computer system usability questionnaire was used to assess user acceptance of a clinical decision support system for outpatient cardiac rehabilitation. Satisfaction with the system was positively correlated with those clinicians who integrated the system into their usual routine.

Harrison AW, Rainer RK. A general measure of user computing satisfaction. Comput Human Behav 1996;12(1):79-92.

The results of this study support the work of Doll and Torkzadeh (1988) with the End-User Computer Satisfaction Instrument for measuring user satisfaction with computer applications. The authors additionally note that their research indicates this usability instrument may be used to measure general user satisfaction with all personnel across all applications rather than with just one specific application.

Harper B, Slaughter L, Norman K. Questionnaire administration via the WWW: A validation & reliability study for a user satisfaction questionnaire. In: Lobodzinski S, Tomek I, editors. World Conference on the WWW, Internet & Intranet. Proceedings of WebNet 97; 1997 Nov 1-5; Toronto, CA. Charlottesville (VA): AACE; 1997. p. 808-10.

Harper et al. discuss the advantages of utilizing the WWW to administer the QUIS, namely lower costs, less time for data collection, access to a highly specialized user population, and minimal delay between exposure to an interface and measures of satisfaction. Other advantages noted were the ability to include a greater number of subjects, thereby increasing the reliability of statistical results.

Hortman PA, Thompson CB. Evaluation of user interface satisfaction of a clinical outcomes database. Comput Inform Nurs 2005;23(6):301-7.

The study by Hortman and Thompson used the questionnaire for user interaction satisfaction (QUIS) for an evaluation of an outcomes database used by nurse practitioners to evaluate clinical patient outcomes and process improvement projects. They conclude that interface designers can become too close to the design to recognize problems, and information from the QUIS ratings can be used to modify screen design prior to database implementation to increase usability.

Im EO, Chee W. Evaluation of the decision support computer program for cancer pain management. Oncol Nurs Forum 2006;33(5):977-82.

Im et al. used the questionnaire for user interaction satisfaction to measure acceptance of a decision support computer program for cancer pain management. Findings showed the decision support computer program to have accuracy and acceptability and to offer valuable evidence-based practice guidelines for managing pain in cancer patients.

Jaspers MWM, Peute LWP, Lauteslager A, et al. Pre-post evaluation of physicians' satisfaction with a redesigned electronic medical record system. Stud Health Technol Inform 2008;136:303-8.

The authors used the questionnaire for user interaction satisfaction and the computer usability satisfaction questionnaire to determine what aspects of a redesigned electronic medical record system met user satisfaction. Both questionnaires were distributed to 150 clinicians who had used both the old version of the electronic medical record and the redesigned electronic medical record. They found that overall user satisfaction was high for both

systems, but also determined problems with screen layout and interaction with the newly designed system.

Johnson CM, Johnson TR, Zhang J. A user-centered framework for redesigning health care interfaces. J Biomed Inform 2005;38(1):75-87.

A case study used the computer system usability questionnaire to compare user satisfaction with the original family history tracking and pedigree drawing program with a redesigned system. The redesigned system showed improvement in system usefulness, information quality, and interface quality.

Kirakowski J. Available at: SUMI Questionnaire Homepage. The use of questionnaire methods for usability assessment (unpublished manuscript). http://sumi.ucc.ie/sumipapp.html. Accessed February 10, 2011.

The history and development of Software Usability Measurement Inventory (SUMI) is reviewed and requirements outlined. This software evaluation scale uses a 50-item questionnaire to measure efficiency, affect, helpfulness, control, and learnability as tested by three kinds of validity studies.

Klimov D, Shahar Y, Taieb-Maimon M. Intelligent visualization and exploration of time-oriented data of multiple patients. Artif Intell Med 2010;49(1):11-31.

The authors designed and developed a system to allow aggregate multipatient data review known as Visualization of Time Oriented Records (VISTORS). A system usability scale (SUS) questionnaire was used to determine the usability of VISTORS. Their use of the SUS revealed that the system was rated less usable by clinicians than by medical informaticians.

Laerum H, Ellingsen G, Faxvaag A. Doctors' use of electronic medical records systems in hospitals: Cross sectional survey. BMJ 2001;323(7325): 1344 8.

The researchers adopted the end-user computing satisfaction questionnaire to compare three electronic medical record systems by doctors in 19 Norwegian hospitals assessing general clinical tasks. Doctors were surveyed utilizing cross sectional survey questionnaires along with semistructured telephone interviews with IT professionals regarding system implementation. Results showed that user satisfaction was positive, but there were some differences among the systems.

Lewis JR. An after-scenario questionnaire for usability studies: Psychometric evaluation over three trials. Indianapolis (IN): IBM; 1991. Technical Report 54.584.

In this report, Lewis compares the psychometric properties of the printer scenario questionnaire with the After-Scenario Questionnaire (ASQ). The printer scenario questionnaire showed less reliability than the After-Scenario Questionnaire. For scenario-based studies, Lewis strongly suggests using the After-Scenario Questionnaire.

Lewis JR. A psychometric evaluation of an after-scenario questionnaire for computer usability studies: The ASQ. ACM SIGCHI Bull 1991;23(1):78-81.

Lewis evaluated the psychometric properties of the three-item After-Scenario Questionnaire. Although the results should be considered preliminary, Lewis showed that this questionnaire has adequate sensitivity, reliability, and concurrent validity.

Lewis JR. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. Int J Hum Comput Interact 1995;7(1):57-78.

Lewis reports on the item construction, item selection, and psychometric evaluation of the After-Scenario Questionnaire (ASQ), the Printer Scenario Questionnaire (PSQ), the Post-Study System Usability Questionnaire (PSSUQ), and the Computer System Usability Questionnaire (CSUQ). A sample of each scale is provided within this report.

Lewis JR. Usability testing. In: Salvendy G, editor. Handbook of human factors and ergonomics. Third Edition. Hoboken (NJ): John Wiley & Sons, Inc.; 2006. p. 1275-316.

Lewis offers both the basics of usability testing and advanced topics such as sample size estimations, confidence interval computations, and important concepts concerning the use of standardized usability questionnaires. This report provides information on the SUS, QUIS, CUSI, SUMI, PSSUQ, and the CSUQ.

Lewis JR, Sauro J. The factor structure of the system usability scale. In: Kurosu M, editor. Human Computer Interaction International Conference (HCII 2009). Proceedings of the 1st International Conference on Human Centered Design; 2009 Jul 19-24; San Diego, CA. Heidelberg (DE): Springer-Verlag Berlin; 2009. p. 94-103.

Lewis and Sauro present the results of their factor analysis on the 10-item system usability scale. They showed that the

scale has two factors; usability (eight items) and learnability (two items). These scales correlate highly with the overall SUS and have coefficient alphas of .91 (usability) and .70 (learnability). This study provided new evidence that the system usability scale could provide information on overall usability or on both usability and learnability.

Lin HX, Choong YY, Salvendy G. A proposed index of usability: A method for comparing the relative usability of different software systems. Behav Inform Technol 1997;16(4):267-78.

The authors present the theoretical framework for the development of the Perdue Usability Testing Questionnaire. The questionnaire is based on human information processing theory. The questionnaire showed good construct and content validity and good criterion-based validity and reliability.

Lund A. Available at: STC Usability SIG Newsletter. Measuring usability with the USE questionnaire. http://www.stcsig.org/usability/newsletter/0110_measuring_with_use.html. Accessed February 11, 2011.

Lund offers a review of the background and development of the USE questionnaire which measures usability across many different domains. While it is noted that the USE has been utilized successfully by many different entities, Lund states that the development of the questionnaire is not completed and that it should be subjected to a complete psychometric instrumentation process.

Lutes KD, Chang K and Baggili IM. Diabetic e-management system (DEMS). In: Latifi S, editor. New Generations. ITNG '06 Proceedings of the Third International Conference on Information Technology; 2006 Apr 10-12; Las Vegas, NV. Washington, DC: IEEE Computer Society; 2006. p. 619-24.

Lutes et al. utilized the system usability scale (SUS) to evaluate the usability of a handheld computer system and custom software which allowed patients to enter and upload health data for use by clinicians. Results of the SUS scores revealed that the patients thought the system to be usable.

Murff HJ, Kannry J. Physician satisfaction with two order entry systems. J Am Med Inform Assoc 2001;8(5): 499-509.

Physician satisfaction was assessed with the questionnaire for user interaction satisfaction to determine user satisfaction with two physician order entry systems; one was a commercial system and the other was the Department of Veterans Affairs Computerized Patient Record System (CPRS). Results showed that the physicians were more satisfied with the CPRS and this was correlated with their ability to carry out their tasks in a "straightforward" way.

Narasimhadevara A, Radhakrishnan T, Leung B, et al. On designing a usable interactive system to support transplant nursing. J Biomed Inform 2008;14(1):137-51.

The authors utilized a combination of agile programming and user centered design to develop an interactive system for transplant nurses. The system was evaluated with the system usability measurement inventory and found to be a well accepted, very usable, and high-quality product.

Nunnally JC, Bernstein I. Psychometric theory (3rd ed.). New York: McGraw-Hill; 1994.

This is a classic comprehensive textbook in psychometric theory. It takes into consideration broad measurement problems that occur in research.

Otero A, Felix P, Barro S, et al. Addressing the flaws of current critical alarms: A fuzzy constraint satisfaction approach. Artif Intell Med 2009;47(3):219-38.

Otero et al. utilize the system usability scale (SUS) in their design of an intelligent alarm system which addresses the current flaws and limitations of threshold alarms. SUS scores revealed that the physicians believed the system to be significantly more usable (76.8) than did the nurses (57.5).

Sawka AM, Straus S, Gafni A, et al. A usability study of a computerized decision aid to help patients with early-stage papillary thyroid carcinoma in decisionmaking on adjuvant radioactive iodine treatment. Patient Educ Couns 2010;Aug 21:Epub ahead of print.

Sawka et al. used the system usability scale (SUS) to determine the usability of a patient-directed decision aid for patients with early papillary thyroid cancer for the purpose of evaluating treatment choices. Results showed the decision aid to be acceptable and to significantly increase medical knowledge in patients with and without thyroid cancer.

Sittig DF, Kuperman GJ and Fiskio J. Evaluating physician satisfaction regarding user interactions with an electronic medical record system. In Proceedings of the American Medical Association Annual Symposium, 1999; 400-404.

The authors used the questionnaire for user interaction satisfaction to assess the usability of an electronic medical

record. The questionnaire was distributed to 75 primary care clinicians. The authors found that user satisfaction was most correlated with screen design and layout and not the response time of the system.

Staggers N, Kobus D, Brown C. Nurses' evaluation of a novel design for an electronic medication administration record. Comput Inform Nurs 2007;25(2):67-75.

This study used the questionnaire for user interaction satisfaction to evaluate an electronic medication administration record prototype. The authors found the usability evaluation to be positive based on the quantitative scores and comments from the instrument.

Van Veenendall E. Questionnaire based usability testing. In: Unknown, editor. EURO & Y2K: The industrial impact. Conference Proceedings of the European Software Quality Week; 1998 Nov 9-13; Brussels, BE. San Francisco (CA): Software Research, Inc.; 1998. p. 1-9.

The authors describe the Software Usability Measurement Inventory and discuss practical applications, cost benefits, and the advantages and disadvantages of the questionnaire.

Appendix C: Web Sites for Nonproprietary Usability Questionnaires

This appendix lists Web addresses where the nonproprietary (publicly available) usability questionnaires we reviewed for this report may be accessed.

Web sites for nonproprietary usability questionnaires

    a. After-Scenario Questionnaire (ASQ) – http://hcibib.org/perlman/question.cgi?form=ASQ

    b. Computer System Usability Questionnaire (CSUQ) – http://hcibib.org/perlman/question.cgi

    c. Post-Study System Usability Questionnaire (PSSUQ) – http://drjim.0catch.com/usabqtr.pdf

    d. Purdue Usability Testing Questionnaire (PUTQ) – http://hcibib.org/perlman/question.cgi?form=PUTQ

    e. Usefulness, Satisfaction, and Ease of Use Questionnaire (USE) – http://www.stcsig.org/usability/newsletter/0110_measuring_with_use.html

    f. End-User Computing Satisfaction Instrument (EUCS) – http://www.jstor.org/stable/pdfplus/248851.pdf

Appendix D. Sections From the Questionnaire for User Interface Satisfaction (QUIS)

This appendix provides a sample of parts from the Questionnaire for User Interaction Satisfaction—a proprietary questionnaire.

QUIS Sample Parts (Copyright © 1984, 1993, 1998. University of Maryland. All rights reserved.).

Identification number: _____

System code: _____

Age: _____

Gender: ____ male

____ female

PART 1: System Experience

1.1   How long have you worked on this system?

__ less than 1 hour	__ 6 months to less than 1 year
__ 1 hour to less than 1 day	__ 1 year to less than 2 years
__ 1 day to less than 1 week	__ 2 years to less than 3 years
__ 1 week to less than 1 month	__ 3 years or more
__ 1 month to less than 6 months	

PART 3: Overall User Reactions

Please circle the numbers which most appropriately reflect your impressions about using this computer system. Not Applicable = NA.

3.1 Overall reactions to the system:	terrible wonderful 1 2 3 4 5 6 7 8 9		NA
3.2	frustrating satisfying 1 2 3 4 5 6 7 8 9		NA
3.3	dull stimulating 1 2 3 4 5 6 7 8 9		NA
3.4	difficult easy 1 2 3 4 5 6 7 8 9		NA
3.5	inadequate power adequate power 1 2 3 4 5 6 7 8 9		NA
3.6	rigid flexible 1 2 3 4 5 6 7 8 9		NA

PART 4: Screen

4.1 Characters on the computer screen	hard to read easy to read 1 2 3 4 5 6 7 8 9		NA
4.2 Highlighting on the screen	unhelpful helpful 1 2 3 4 5 6 7 8 9		NA
4.3 Screen layouts were helpful	never always 1 2 3 4 5 6 7 8 9		NA
4.4 Sequence of screens	confusing clear 1 2 3 4 5 6 7 8 9		NA

Please write your comments about the screens here:

_____

_____

_____

_____

PART 5: Terminology and System Information

5.1	Use of terminology throughout system	inconsistent consistent 1 2 3 4 5 6 7 8 9	NA
5.2	Terminology relates well to the work you are doing	never always 1 2 3 4 5 6 7 8 9	NA
5.3	Messages which appear on screen	inconsistent consistent 1 2 3 4 5 6 7 8 9	NA
5.4	Messages which appear on screen	confusing clear 1 2 3 4 5 6 7 8 9	NA
5.5	Computer keeps you informed about what it is doing	never always 1 2 3 4 5 6 7 8 9	NA
5.6	Error messages	unhelpful helpful 1 2 3 4 5 6 7 8 9	NA

Please write your comments about terminology and system information here:

_____

_____

_____

PART 7: System Capabilities

7.1	System speed	too slow fast enough 1 2 3 4 5 6 7 8 9	NA
7.2	The system is reliable	never always 1 2 3 4 5 6 7 8 9	NA
7.3	System tends to be	noisy quiet 1 2 3 4 5 6 7 8 9	NA
7.4	Correcting your mistakes	difficult easy 1 2 3 4 5 6 7 8 9	NA
7.5	Ease of operation depends on your level of experience	never always 1 2 3 4 5 6 7 8 9	NA

Please write your comments about system capabilities here:

_____

_____

_____

PART 8: Technical Manuals and On-line help

8.1 Technical manuals are	confusing	clear	
	1 2 3 4 5 6 7 8 9		NA
8.2 Information from the manual is easily	never	always	
understood	1 2 3 4 5 6 7 8 9		NA
8.3 Amount of help given	inadequate	adequate	
	1 2 3 4 5 6 7 8 9		NA

Please write your comments about technical manuals and on-line help here:

_____

_____

_____

_____