Assessment Protocol: Applying Usability Assessment Methods to Evaluate Clinical Decision Support (CDS) Interventions

Department of Veterans Affairs (VA)
Office of Informatics and Information Governance (OIIG)
Health Informatics (HI)
Office of Knowledge Based Systems (KBS)

Assessment Protocol: Applying Usability Assessment Methods to Evaluate Clinical Decision Support (CDS) Interventions:

by , Department of Veterans Affairs (VA), Office of Informatics and Information Governance (OIIG), Health Informatics (HI), and Office of Knowledge Based Systems (KBS)

Publication date September 2017

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

This Assessment Protocol is based upon the material in the *General Model to Evaluate CDS Interventions* (DVA, 2017b) and the *Practical Guide for Clinical Decision Support (CDS) Evaluation* (DVA, 2017c). Although this document is based on its companion documents, the Assessment Protocol is intended to function as a stand-alone resource to guide the assessment of a CDS intervention using select usability methods: heuristic evaluation, user observation for identifying usability problems and workflow, and summative usability testing. **The assessment protocol provides an overview of the methods and applies the steps in the CDS Impact Evaluation Framework to illustrate how the method can be applied to evaluate a CDS intervention.**

This protocol can be used in a number of ways. For example, the Assessment Protocol could be used to design an evaluation to compare a new CDS intervention with an existing manual workflow or to compare one CDS intervention with another. This document will begin with a list of assumptions and constraints. This will be followed by a step-by-step plan to conduct an assessment of a CDS intervention. For the sake of clarity, aspects of the General Model are referenced and may be accessed as needed. For a detailed discussion of the General Model, please refer to the *General Model to Evaluate CDS Interventions* report.

This work is important for two reasons. Clinical quality improvement efforts generally, the informatics field and the U.S. Department of Veterans Affairs (VA) specifically do not yet use a model-based approach to evaluate CDS interventions. Likewise, a clear step-by-step approach to CDS evaluations is not available. This protocol offers both a model-based and step-by-step approach. Evaluators will be able to use the assessment methods described in this document as an improved approach to CDS evaluations, potentially leading to richer results in the short term, as well as potentially improved CDS interventions and possibly better care outcomes in the long term.

Chapter 2. Assumptions

For the Assessment Protocol to be useful from a practical standpoint, certain assumptions must be made. This document can be used by readers with advanced informatics training, although the intended audience consists of clinicians at a local level, perhaps on or working with a quality improvement (QI) team, who would like to perform an evaluation or assessment of an implemented CDS intervention. The Assessment Protocol provides several salient methods to evaluate CDS interventions, with a concentration on less typical QI methods. The intent is not to outline the entire suite of available methods but to provide evaluators with a select set of commonly employed usability methods. In addition, this Assessment Protocol assumes that no sophisticated instrumentation will be available to the assessment team (e.g., eye-tracking hardware, a usability lab, etc.). The overall objective of the Assessment Protocol is to evaluate an *existing* CDS intervention in use by clinicians. This protocol can inform the *development* of a CDS intervention and may provide useful context for requirements, but the primary focus is intended for a post-implementation phase where the evaluation team has little familiarity with detailed, formalized CDS evaluation or usability studies. Finally, one method is listed as an expert technique. Those trained in the technique using the material in this document can be considered experts.

A *CDS Evaluation Model* adapted from Osheroff and colleagues was discussed as part of the General Model in section 1.5.2 (DVA, 2017b). The five steps of that model included: 1) create CDS interventions, 2) verify and validate, 3) monitor and measure, 4) evaluate impact effectiveness, and 5) modify and maintain. This Assessment Protocol outlines processes for CDS evaluation primarily at the evaluate impact effectiveness step. However, heuristic evaluation and user observation techniques in particular may also be used in earlier steps such as: verify and validate as well as monitor and modify.

2.1. Introductory Material About Evaluating CDS Interventions

Readers are invited to view introductory material in the General Model about CDS interventions as needed (DVA, 2017b):

- Prevalence and Importance of CDS Interventions (section 1.2);
- Definition of CDS Interventions (section 1.3.1), CDS impact (section 1.3.2), and assessments versus evaluation (section 1.3.3);
- Reasons to measure CDS impact (section 1.4);
- Introduction to Continuous Quality Improvement (CQI) and CDS Assessments and Evaluations (section 1.5); and
- CQI and CDS Organizational Drivers (section 1.5.1.1).

Chapter 3. Approach

The Assessment Protocol follows the step-by-step processes in the CDS Impact Evaluation Framework outlined in the Practical Guide: (1) Define Assessment Goals, (2) Develop Assessment Plan, (3) Prepare for Assessment, (4) Conduct the Assessment, (5) Analyze Data, and (6) Synthesize and Communicate Findings with one exception. Evaluators will want to choose a method earlier in the process—as they develop assessment goals versus waiting until the planning process. This is important because the method will drive goal refinement and other steps in the framework. That is, any usability method is tightly linked to elements of the assessment goals and plan. Readers are referred to details about the CDS impact evaluation process in the Practical Guide (DVA, 2017c).

Based upon the assumptions stated earlier, three main methods to conduct an assessment will be presented as options for the evaluators. Within the larger set of available techniques, the three main approaches to conducting a CDS intervention evaluation in this Assessment Protocol are heuristic evaluation, user observations, and summative usability testing. Each will be discussed in detail. These may be used independently or in combination, depending on the desired outcomes and other constraints (e.g., time).

Chapter 4. Define Assessment Goals

Readers are referred to material on defining assessment goals in the Practical Guide, chapter 3 (DVA, 2017c). The Assessment Protocol builds on that material.

4.1. Define Primary Assessment Goals for Usability Assessments

Table 4.1, “Sample Primary Goals for Usability Assessments”shows examples of primary assessment goals related to the three select usability methods. Evaluators may want to conduct assessments using more than one usability method as multimethod evaluations can identify different and complementary usability problems (Georgsson & Staggers, 2016a).

Table 4.1. Sample Primary Goals for Usability Assessments

General Model Category	Sample Primary Goals
Effectiveness	• Determine why the CDS intervention changed only 30% of clinicians’ behavior • Determine why only 20% of clinicians are using the suicide risk CDS intervention • Determine whether the costs for lab tests on magnesium levels were lower after CDS implementation
Safety	• Determine why a medication alert is being overridden frequently • Determine the workflow among physicians, nurse practitioners, and staff nurses for a CDS alert pre- and post-implementation
Efficiency	• Determine whether the CDS intervention meets its predetermined benchmark for data entry times • Determine the workflow among physicians, nurse practitioners, and staff nurses for a CDS alert pre- and post-implementation
Satisfaction	• Determine if the CDS intervention meets the threshold for acceptable usability (e.g., 60 on a scale of 100)

Once the primary assessment goals and target CDS intervention are known, evaluators will want to create more specific objectives for the CDS usability assessment. For example, see Table 4.2, “Sample of Specific Usability Assessment Goals and Objectives ”.

Table 4.2. Sample of Specific Usability Assessment Goals and Objectives

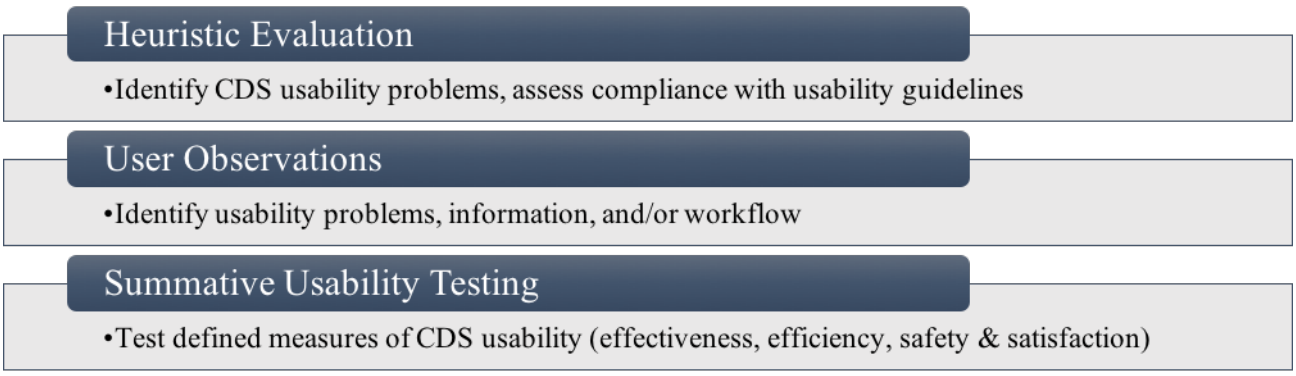
General Model Category	Primary Assessment Goals	Specific Assessment Objectives
Effectiveness	Determine why a medication alert changed only 30% of clinicians’ behavior	• Measure the number of times an alert is overridden and the reason for the override • Determine the number of alerts by role (physician, nurse practitioner, staff nurse) and that the CDS configuration considers whether it should avoid firing in designated cases (by role, setting, etc.)
Safety	Determine why a medication alert is	• Observe clinicians as they interact with this CDS intervention to detect potential safety issues

General Model Category	Primary Assessment Goals	Specific Assessment Objectives
	being overridden frequently	• Interview clinicians about their reasons for alert overrides
Efficiency	• Determine whether provider workflow is interrupted with the new alert • Determine the average time providers take to address this alert	• Observe the number and kinds of interruptive alerts for the CDS by role in a naturalistic setting • Time 15 users as they interact with the new CDS intervention during a representative task
Satisfaction	No goal selected for this category in this particular evaluation plan	N/A

Chapter 5. Develop Assessment Plan

The next step is to develop an assessment plan for the CDS intervention for each of the three main types of usability methods: (1) heuristic evaluation, (2) user observations, and (3) summative usability tests (Figure 5.1, “Introduction to the Three Usability Assessment Methods”).

Figure 5.1. Introduction to the Three Usability Assessment Methods



Each method varies in the type of data created, use in particular development stages of the product, the time and resources needed, complexity, and level of granularity or detail provided. Thus, evaluators will want to tailor the assessment plan to the chosen method. When planning the assessment, an important consideration is that the environment, population of users, user interface, and other device characteristics (such as placement, size, accessibility, etc.) play a role in overall usability. These characteristics can be considered in addition to usability problems to due functionality, fit to workflow and interoperability.

Table 5.1, “Overview of CDS Intervention Assessment Methods” provides a high-level overview of each of the three methods and the benefits and disadvantages of each. In addition to these attributes, evaluators will want to consider the complexity of the CDS intervention being evaluated (e.g., a paper prototype may be more feasible to evaluate than waiting for a fully programmed intervention. Moreover, evaluations of an early prototype can identify crucial problems before they are programmed and difficult to fix.) The benefits and disadvantages in Table 5.1, “Overview of CDS Intervention Assessment Methods” may be used as part of the method selection process.

Table 5.1. Overview of CDS Intervention Assessment Methods

Method	Description	When to Use	Resources	Benefits	Disadvantages
Heuristic Evaluation	An expert method (e.g., someone trained in using the technique) to inspect an intervention, find problems, categorize them, and assign severity scores	Anytime in the systems life cycle (pre- or post-implementation)	• 3–5 single-domain experts (e.g., trained in the method or experts in usability) • 2–3 dual-domain experts (e.g., clinical and usability) • Any product from an early prototype to a fielded CDS intervention	• Cost effective • Can be completed virtually • Time effective • Typically finds many existing usability problems	• Excludes end users • Can generate a list of primarily minor usability problems if the evaluators are not also domain experts • May not provide guidance on how to remedy identified issues • Concentrates on usability problems versus helpful features
User Observations	Direct observations of actual users as they interact with CDS interventions in order to identify usability problems and/or workflow processes	Anytime in the systems life cycle (pre- or post-implementation)	• One for discrete goals (e.g., determining task sequence or observing an adverse event) • Minimum of 3–5 representative users for each user group for main goal(s) • Observer evaluator(s) • Any product from an early prototype to a fielded CDS intervention	• Viewing actual users who perform the work in context • Easy to learn • Can be relatively fast with a small number of users and less complex tasks • Robust amount of issues with a small number of users	• Can be time-consuming and costly depending upon the number of users and/or the complexity of the tasks • Observations can be obtrusive to users • Results can be difficult to replicate due to differences in users and observer evaluators • Evaluators can find it easy to be distracted with the complexity of activities in naturalistic settings
Summative Usability Testing	A more formal assessment of CDS intervention in a more controlled environment involving multiple representative participants, discrete measures, defined tasks, and at least one moderator	Summative usability testing is conducted at the end of the development cycle or after fielding	• Study moderator, researchers/ study support • Near-final prototype or finalized version of a CDS intervention • Formal protocol and controlled setting • 15–20 participants	• Includes end users • Objective data • Allows for comparisons • Controlled study that reduces variance across elements • Formal usability lab not needed	• More complex to conduct than heuristic evaluation and user observation • Higher resource demands (people, costs, time) • Takes longer to plan and execute

5.1. Heuristic Evaluation

Heuristic evaluation is one of the most well-established and commonly employed usability methods where heuristics (i.e., rules of thumb) are used to categorize identified usability problems. The technique is considered an expert method where those trained in the technique (or usability experts) use tasks to interact with and evaluate a CDS intervention to:

- Find usability problems;
- Assign the problems to a specific category of heuristic; and
- Rate each problem by severity (Nielsen & Molich, 1990).

The technique can be used during any phase of the life cycle. Heuristic evaluation is particularly useful because it is a discount usability technique—meaning it is relatively quick to complete, cost effective, and resource efficient (Jeffries, Miller, Wharton, & Uyeda, 1991; Nielsen & Molich, 1990). Improvements to the original method that can enhance its effectiveness include the use of dual-domain experts (who possess training in this usability method and clinical expertise); use of validated tasks; and employing an expanded method for severity ratings (Georgsson & Staggers, 2016a). These improvements can help evaluators then translate usability problems into design requirements. Primary disadvantages of the technique are that it does not involve actual end users, and it concentrates on identifying rather than solving usability problems.

5.2. User Observations

Evaluators will find direct user observations helpful and insightful for evaluating CDS interventions. Evaluators may perform naturalistic or controlled observations to find usability problems or construct workflow processes. Naturalistic observations are conducted in users' actual work environments; they involve an observer evaluator and participant who is a representative CDS user. The participant interacts with the target CDS intervention while talking aloud, called a "thinking aloud" protocol (Nielsen Norman Group, 2012; Nielsen, 1993). Evaluators capture issues through note-taking or analyses of recordings. Controlled observations are similar but conducted outside the users' work environment, typically in a quiet room or lab. An observer captures information about the interaction (e.g., usability problems or steps to complete a task).

Workflow analysis is a product of observations (or interviews) to document activities completed toward a specific goal, such as ordering a medication. The activities depict task and/or information flow, often across stakeholders (Osheroff, et al., 2012, p. 88). Workflow analysis can be either basic or complex, as the analyses can be completed at various levels of complexity. These analyses can have multiple purposes: (a) to document a process for use in initial design or process redesign, (b) to be used for other usability tools such as task identification or scenario building, or (c) to identify usability problems with a deployed CDS intervention. To be fair, whole academic courses are available about this workflow and task analysis. In contrast, this document offers tools to help evaluators construct basic and useful workflow process maps based upon succinct user observations.

User observations are important to understand for several reasons: (a) CDS interventions are tightly intertwined with care processes and workflow; (b) outlining workflow can identify gaps or bottlenecks in processes; (c) the optimal timing of a CDS intervention can help assure it is more effective (Osheroff, et al., 2012); and (d) it can be used to evaluate the unintended consequences of CDS interventions, such as interruptions.

User observations are currently underutilized as a CQI method. There are advantages to the observation of actual users in actual settings. Evaluators can see how a CDS intervention is being used (or not) to complete work, where gaps or redundancies exist, and where usability problems exist in the process of completing work. The technique is relatively easy to learn, and it can reveal a robust set of usability problems even with a small group of users. Its disadvantages are that it can be time-consuming and more costly, depending upon the number of users, and the results can be difficult to replicate due to variability in evaluators and users.

5.3. Summative Usability Testing

A usability test, fundamentally, measures how—or whether—a product supports end users' work for both cognitive and behavioral aspects. Usability tests can be very formal and rigorous using advanced research methods, or they can be more basic and less formal. Even basic testing reveals important information about the product. Summative usability testing concentrates on assessing a CDS intervention after a product is near final in development, or already implemented, and emphasizes more objective measures. This kind of usability test is called *summative testing*. Excellent resources for this type of testing are National Institute of Standards and Technology (NIST) Internal Report 7804 (Lowry, Quinn, & Ramaiah, 2012; Rubin & Chisnell, 2008; Tullis and Albert, 2013).

Each method is now described in more detail using the steps in the CDS Impact Evaluation process.

Chapter 6. Heuristic Evaluation

6.1. Refine Assessment Goals and Create Specific Objectives

Evaluators define major goals and select heuristic evaluation as the method of choice (or one of a suite of methods) in previous steps of the CDS Impact Evaluation Framework. The next step is to refine specific objectives for the heuristic evaluation assessment. Heuristic evaluation objectives can be at a more general (versus discrete) level. For example:

- What are the usability problems with this CDS intervention?
- What are their severity levels?

6.2. Prepare for the Assessment

Readers are referred to the Practical Guide for general information on preparing for the assessment, chapter 5 (DVA, 2017c). As with all evaluations, one of the most important steps in a usability evaluation of any kind is creating a plan for the assessment. In fact, the planning step may consume more time than the data collection. Evaluators will want to keep the goal and main method in mind as they step through each of these considerations to create scenarios, tasks, the test environment, participant and moderator instructions, training materials, and required instruments.

6.2.1. Obtain Approvals

Institutional Review Board (IRB) approvals are usually not required for heuristic evaluations. However, evaluators may be in a site that does require a managerial approval (e.g., to obtain funding or to allocate time to work on the assessment).

6.2.2. Select Team Members

Considerations for selecting the team of evaluators for heuristic evaluation include:

- Clinical expertise;
- Expertise related to the CDS intervention;
- Informatics knowledge;
- Usability expertise;
- Research expertise;
- Experience in using heuristic evaluation in the past; or
- A combination of the above.

A lead evaluator should be selected to guide the process. This evaluator should have formal knowledge, training, and expertise in this method. Optimally, the lead evaluator would have performed heuristic evaluations in the past, made findings available internally in a local organization and/or published findings in the literature. The other evaluators can have a variety of backgrounds with expertise in at least one of the areas listed here. Variability in backgrounds may lead to a more comprehensive evaluation across the team members.

Heuristic evaluation requires minimal single-domain or dual-domain evaluators. A single domain expert is one who is an expert in one particular field such as a clinical field or usability. A dual-domain evaluator is one who is an expert in both. The distinction is important to note. Three to five evaluators find between 74% and

87% usability problems on average (Nielsen, 1992). The number of usability problems found by dual-domain evaluators is even higher at 81% to 90% and requires only two to three evaluators (Nielsen, 1992).

6.2.3. Select Participants

Because heuristic evaluation is an expert method, no participants are needed for the main assessment. However, evaluators may wish to consult with either participants or clinical experts to verify findings and/or discuss implications of the findings.

6.2.4. Determine the Setting and Create the Software CDS Testing Environment

Heuristic evaluations of CDS interventions are completed in a quiet room, such as a conference room, a computer lab, a virtual environment, or other space with network access to the CDS intervention. The lead evaluator will need to create or make available a computer environment for use by the team. While authorized evaluators might view the production environment to verify an issue, the production environment is not suitable for evaluation because of patient privacy concerns and because any required data entry would inadvertently change patient records. The evaluators may need to work with the IT department to create a separate computer environment, similar to a training environment (or the training environment itself), to conduct the evaluation and work through any details about how and when to access the CDS intervention site. Evaluators will want to verify that the training environment, if it is used, is consistent with the production environment. It may not be the same version. Also, heuristic evaluations can occur with fictitious data that mirrors actual patients' and clinicians' records, so the lead evaluator should be prepared to populate records with synthetic data if needed. Another consideration is that the environment, like a training environment, must allow refresh (reset) of the application back to default data as new evaluators begin their assessments.

6.2.5. Select Tools and Instrumentation

6.2.5.1. Create Tasks

Heuristic evaluations are completed using defined tasks. In fact, heuristic evaluations are most useful when performed in context with specific tasks, (e.g., a context of medication administration with the set of related tasks). Most often these are independently created without the use of a scenario (a narrative about the setting and users) as is common with other methods such as usability testing. However, the evaluators may decide to create and employ a scenario. This is especially important if evaluators are not experts in the domain being assessed. Please see guidance in Section 8.2.5.1, "Write Scenarios" to create scenarios.

Tasks for any evaluation should represent the common activities users employ to interact with the application at hand and be in a sequence that makes sense to the end user. They are tailored to each application being evaluated and are typically specific and discrete (e.g., not "Navigate around the interface" but "Navigate to the vital sign section of the SMARTForm"). Optimally, tasks are derived from end users, from usability testing, or they may be validated by actual end users (Georgsson, Staggers, & Weir, 2016b). For example, common tasks for the evaluation of an electronic medication administration record (eMAR) would be:

- Order and modify medication(s).
- Verify medication orders.
- Prepare medication.
- Access medication information/help.
- Administer and chart medication(s).
- Generate reports and review eMAR (Staggers, Iribarren, Guo, & Weir, 2015).

Sample tasks to evaluate applications, such as SMARTForms, could include:

1. Navigate to and access the correct form.
2. Enter assessment data (e.g., vital signs, physical assessment, mental status, ethyl alcohol (ETOH) history, etc.), as specified by the evaluators with consideration to the purpose of the evaluation and capabilities of the users. Evaluators will want to specify exactly which data users should enter.
3. Find fields that are empty and need to be completed by users.
4. Determine the fall risk score based upon entered data.
5. Evaluate suggested interventions from the CDS.
6. Find the suicide risk score.
7. Locate all information that is abnormal and/or needs follow-up.
8. Create the “story of the patient” to communicate during a handoff. In a SMARTForm, this would be the ability to see a summary of all the abnormal values or findings in the form, for example.

6.2.5.2. Prepare Briefing Instructions

Using standard instructions, called a script, the lead evaluator can guide the other evaluators though the processes for the planned heuristic evaluation. This can be completed verbally or electronically. Elements in this script can include:

- A thank you to other evaluators;
- A reminder about the purpose of the assessment;
- Where to access the CDS intervention;
- How to receive training on the CDS intervention;
- The tasks to interact with the CDS intervention;
- The list of heuristics being used;
- A data collection form;
- A timeframe for completing the evaluation; and
- How to upload or communicate findings to the lead evaluator.

Sample Instructions

> Thank you for participating in this assessment. As a reminder, we are conducting a heuristic evaluation of [name the CDS intervention]. You can access the CDS intervention at [provide the location or URL]. Training on this CDS intervention is available at [provide the location or online training location]. After training, you will complete your assessments individually using the list of heuristics provided to you. Please use the list of tasks [provide the list] to interact with the CDS intervention and outline any usability problems on the form provided [provide the form]. The assessment period is over the next [name the dates]. You may upload your findings at [provide the URL] or communicate directly with [give the contact person]. Questions are welcome at any time. After our individual assessments, we will compile the findings into a master list and complete further analyses.

6.2.5.3. Develop Training Materials

For heuristic evaluations, training may be needed on: (a) the CDS intervention and (b) the heuristic evaluation method itself. All team members should have standardized training on the CDS intervention, its purpose(s), its

uses to date, and its navigation as correlated to typical workflow. The goal is for the set of evaluators to have the same baseline knowledge and skills about the CDS intervention. A standard training method is recommended to decrease variability and increase the likelihood that major usability problems will be found across the evaluators.

The lead evaluator may need to train team members on the heuristic evaluation technique. This Assessment Protocol may be used as a resource. Other resources include Harrington, et al. (2011); Nielsen & Mack, (1994); and Zhang, et al. (2003) .

6.2.5.4. Develop Test Materials Packet

For any type of assessment, evaluators will need to gather and/or develop needed testing materials (e.g., demographic forms, specific heuristic categories, and data collection forms). A lead evaluator compiles the materials needed for the assessment and organizes the process. Activities include developing and executing the heuristic evaluation plan; crafting a timeline; and communicating the details about how to obtain access to the application, how to record findings, and where to upload or send the findings.

6.2.5.4.1. Develop a Demographic Questionnaire

The lead evaluator for the heuristic evaluation will want to collect demographic data about the other evaluators who complete the heuristic evaluation to ensure a representative cross-section of salient characteristics such as experience with the method and CDS intervention. Any specific information should be linked to the purpose of the evaluation. For instance, race is not related to application interactions so would not be collected, but specialty could affect the interaction. Typical types of demographic data might include:

- Profession;
- Identification as single- or dual-domain evaluator;
- Educational level;
- Specialty;
- Years in their field;
- Usability experience or education;
- Training on the method; and
- Previous publications or projects using heuristic evaluation.

6.2.5.4.2. Create Heuristic Categories

To identify usability problems, evaluators will need to create a reference document on heuristic categories and sample evaluation items. The reference document can be based upon 14 categories such as those by Zhang et al. (2003) available on the University of Texas, School of Biomedical Informatics, National Center for Cognitive Informatics & Decision Making in Healthcare website [https://sbmi.uth.edu/nccd/ehrusability/design/guidelines/] (2017). A sample reference document for heuristic categories is in Table 6.1, “Heuristic Categories & Sample Items to Evaluate, Adapted from: Zhang et al. (2003) and Guo et al. (2011)”.

Table 6.1. Heuristic Categories & Sample Items to Evaluate, Adapted from: Zhang et al. (2003) and Guo et al. (2011)

Heuristic Category	Definition	Sample Items to Evaluate
Consistency and Standards	Product is consistent across all aspects: methods of navigation; messages and actions; and meaning of buttons, terms, and icons. For informatics-trained	• Screens have consistent layout (e.g., patient name is in the same location on each screen; navigation for back, exit, help, or home is in the same location) • Icons are intuitive, make sense to clinicians

Heuristic Category	Definition	Sample Items to Evaluate
	evaluators, other methods might be used. For example, that the product is congruent with known screen design principles for color and screen layout. Advanced evaluators: Product is consistent with International Organization for Standardization (ISO) usability guidelines.	• Colors for high, low values are consistent across screens • Colors have consistent meaning (red = critical) across screens, modules • Navigation methods are consistent (e.g., how to select an item, drag an item) • Actions from elements (e.g., buttons) are consistent
Visibility of System State	Users understand what the system is doing and what they can do with the product from the system messages, information, and displays.	• Items should change appearance when selected and at each step in a list or process (helpful when a user is interrupted) • Corrections, deletions, modifications, copy/pastes should be visible and change appearance (e.g., color)
Match Between System and World	The technology matches the way users think and work, uses appropriate information flow, has typical options that users need, and includes expected actions by the system.	• Information flow matches users' thought processes • Key information is prominent for linked or "chunked" information such as a bed number on a patient roster • Critical information is available when needed (e.g., not buried in an electronic "sticky note" or awaiting a system refresh) • Complete information is available for the task (e.g., auto-population works) • Associated information is linked to appropriate fields (e.g., current medications are auto-populated into the discharge form) • Depth of information is appropriate to the clinical specialty (e.g., in-depth neuro assessment for Neuro ICU, high level for GYN) • Ability to synthesize, collate information (e.g., see the "big picture") • Intuitive format (e.g., blood pressure of 150/90 in one field, not in separate fields or with 90 listed first) • Avoids use of hybrid systems (e.g., a combination of paper and electronic or use of other methods to calculate titrations) • Data are charted in one clear place (or auto-populated) so users do not have to look in several places to find critical information (e.g., medication reactions)
Minimalist	Product does not display superfluous information. System and screen design are targeted to primary information users' needs. Use of progressive disclosure to display details of a category of information only	• Information is tailored to the user's specialty (e.g., surgeons do not see full preventive medication alerts) • Users can click on a category (e.g., history) to see full history • Use of graphs and images versus text to show patterns with discrete data available as needed

Heuristic Category	Definition	Sample Items to Evaluate
	when needed. The exception can be designs for expert users where screen density is preferred.	
Minimize Memory Load	Product minimizes the amount of information and tasks users have to memorize to adequately use the technology. Product makes use of sample formats for data input such as a calendar for date format.	• Minimal clicks, especially for common tasks (e.g., entering orders and vital signs, finding summary data, entering data into a CDS intervention) • Collated information so users do not have to search for information (e.g., shift change or across many screens in a SMARTForm) • Obvious indicators for missed fields (e.g., missed assessment data, missed meds)
Informative Feedback	The technology provides prompt and useful feedback about users' interactions and actions (e.g., feedback that orders were placed).	Clear messages after actions (e.g., "assessment saved," "order entered," or "order still pending")
Flexibility and Efficiency	The product can be tailored and customized to suit individuals' needs. It includes novice and expert capabilities (e.g., string searches).	• Availability to tailor to specialty and expertise (e.g., primary care vs. anesthesiology notes) • Ability to edit comments, notes (with indicators) according to role (e.g., aids cannot edit RN or MD notes)
Good Error Messages	Error messages tell users what error occurred and how users can recover from the error. Messages are not abstract or general such as "Forbidden!" Messages need to be precise and polite and not blame the user.	Clear error messages (e.g., "need reason for alert override" or "mandatory field left blank," with a pointer to which one)
Avoid Catastrophic Errors	Product must prevent catastrophic errors as much as possible (e.g., pediatric medication order dosing mixed between kilograms and pounds).	• Alerts for out-of-range values for a field (e.g., temp of 400º) • Consistent units of measure and data values in a graph and throughout the whole system (e.g., kilograms for patient weights, centigrade for temperatures) • Data is patient-centered and flows with the patient (e.g., Bar code medication administration [BCMA] is not available in the operating room [OR], emergency department [ED], or interventional radiology)
Clear Closure	Users should know when a task is completed and all information is accepted. Displays should include progress toward 100% completion using a series of bars or other graphical depictions.	Clear indicator when process is completed (e.g., with a message, change in appearance, or the like)
Reversible Actions	Whenever possible, actions and interactions should be able to be undone within legal limits in electronic health records. If actions cannot be reversed,	• Availability of an undo button wherever appropriate • Clear indicator of edited material including user ID in notes, orders

Heuristic Category	Definition	Sample Items to Evaluate
	there is a consistent procedure to document the correction of any misinformation in the system.	
Use the Users' Language	The technology uses language and terms the targeted users can comprehend and expect. Health terms are used appropriately.	• Terms and abbreviations make sense to users (e.g., not HSM, an obscure abbreviation used in the VA's eMAR), and assure that medical terms like CPR are only used to mean what clinical staff think they mean (e.g., cardiopulmonary resuscitation) • Check all terms for clinical meaning versus developer input, and tailor the terms to the users (e.g., terms specific to pharmacy may not be understood by nurses and vice versa)
Users in Control	Users initiate actions versus having the perception that the technology is in control. Avoid surprising actions, ending up in unexpected places, and loud sounds to notify users.	• Application can be tailored to suppress sounds when patients are with the user • No unexpected system actions (e.g., maintenance cycle during shift change or prime medication times)
Help and Documentation	Product provides help for users within the context the actions occur. It embeds help functions throughout the application.	• Help is context-sensitive (e.g., tailored to the location in the application versus having to search the entire help document) • Help is easily accessible from any screen • Help is granular enough to solve the user's issue

6.3. Conduct Assessment

The evaluators examine the application individually, having no collaboration with others (Nielsen & Mack, 1994). This is advantageous over other methods where scheduling issues may delay assessment. Evaluators may choose to capture screenshots for catastrophic usability problems or to augment the list of problems.

6.3.1. Collect Data

Evaluators individually step through the CDS intervention using the validated tasks described in Section 6.2.5.1, "Create Tasks" to identify usability problems (see Table 6.1, "Heuristic Categories & Sample Items to Evaluate, Adapted from: Zhang et al. (2003) and Guo et al. (2011)" and Table 6.2, "Sample List of Usability Problems and Their Descriptions"). Usability problems are an element of a product that makes it difficult, impossible, inefficient, and/or unpleasant for a user to complete a task or goal in a typical way (Lavery, Cockton & Atkinson, 1997). Usability problems include a broad array of issues that may encompass functionality (or lack of it), systems integration (or its absence), missing information, incomplete processes, mismatches with workflow or the way users think, lack of information synthesis to support "at a glance" information as well problems with user interface design, confusing icons, and excessive required clicks.

Sample usability problems might be:

- Mismatches in the way a CDS intervention is organized compared to the way a clinician thinks;
- A mismatch with expected workflow and the timing of a CDS alert;
- Fields out of view, so they are missed;
- Required data is not imported into a CDS intervention; or

- The use of unconventional abbreviations that clinicians have to memorize (or ignore) such as HSM, a term used in the VA's electronic medication administration record (eMAR), or internal entry number (IEN) abbreviations used in the VA's eMAR. When asked about the two abbreviations nurses had no idea what either of them meant (Staggers, Iribarren, Guo, & Weir, 2015).

Usability problems and their descriptions are recorded on a data collection form, such as the one in Table 6.2, "Sample List of Usability Problems and Their Descriptions". Evaluators can use the list of heuristics and sample items in Table 6.1, "Heuristic Categories & Sample Items to Evaluate, Adapted from: Zhang et al. (2003) and Guo et al. (2011)" as they complete tasks.

Data collection forms can be as simple as entering information into a Word table or an Access database and can be organized by task, (e.g., see Table 6.5, "Summary Table for an Electronic Medication Administration Record (eMAR) Example (Guo, Iribarren, Kapsandoy, Perri, & Staggers, 2011)" from an eMAR evaluation). A similar and expanded tool is available from the University of Texas, School of Biomedical Informatics, National Center for Cognitive Informatics & Decision Making in Healthcare website [https://sbmi.uth.edu/nccd/ehrusability/design/guidelines/] (2017).

Table 6.2. Sample List of Usability Problems and Their Descriptions

Task	Description of Usability Problem	Location/Screenshots
Administer Medications	• BCMA screen does not have an indicator for automatic refresh; nurses must continually click refresh (a multiple-step process) or exit out and reenter to avoid missing new or discontinued orders. • No structured option is available to document medication patch placement location; nurses must use a free-text field.	A list outlining where the problem exists and/or screenshots. For example, in this case, the usability issue is located on the BCMA screen, one of four possible screens in the eMAR.
Prepare Medications	• No easy way to organize/prepare for medication administration, especially across patients; work-around is to run missed medication reports often. • Nurses must document IVs using "IV bag chronology," which is not intuitive.	
Other tasks	Describe identified usability problems.	

6.4. Analyze Data

The lead evaluator will consolidate usability problems into a master list (in Word or in a database), totaling how frequently evaluators found the same problems. Evaluators may need to discuss the identified usability problems and clarify descriptions before more analyses occur. Evaluators then assign the consolidated usability problems into a specific heuristic category as listed in Table 6.1, "Heuristic Categories & Sample Items to Evaluate, Adapted from: Zhang et al. (2003) and Guo et al. (2011)". More than one heuristic may be assigned to each usability problem. This process may be completed individually and then discussed as a group to verify category assignments. Alternately, the process may be completed using group discussion. Once a usability problem is categorized using a heuristic, it is known as a *heuristic violation*.

The evaluators assign severity ratings to each usability problem using the rating scale in Table 6.3, "Severity Rating Scale, Adapted from: (Georgsson, Staggers, & Weir, 2016; Nielsen & Mack, 1994)". Again, this process may be completed individually or as a group, but usually the process is completed individually so averages can be computed. For usability-trained evaluators, the traditional Nielsen or Zhang severity ratings can be expanded to include in-depth severity ratings across these dimensions (Georgsson, Staggers, & Weir, 2016b): frequency, impact, and persistence (Table 6.4, "Severity Factors: Advanced (Usability Experts)"). Each dimension is individually scored, summed, and averaged across the evaluators, using descriptive statistics.

Table 6.3. Severity Rating Scale, Adapted from: (Georgsson, Staggers, & Weir, 2016; Nielsen & Mack, 1994)

Severity Rating	Description	Priority
0	Not a usability problem.	None
1	Cosmetic—No need to fix unless extra time is available.	Low
2	Minor—Annoying issue with minor impact.	Medium
3	Major—Important to fix. Issues with major impact in use or during training or both. Consider the numbers and kinds of users affected.	High
4	Catastrophic—Severe issue that must be corrected, especially issues related to patient safety. Imperative to fix this before product can be released (if in development).	Critical

Table 6.4. Severity Factors: Advanced (Usability Experts)

Factor	Questions
Problem frequency	Is it common or rare?
Problem impact (if it occurs)	Will it be easy or difficult for the users to overcome?
Problem persistence	Is it a one-time problem that users can overcome once they know about it, or will users repeatedly be bothered by the problem?

The data collection form is expanded to include all elements: task, usability problem description, heuristic violations, and severity ratings. A beginning summary table for the previous example on eMAR might look like Table 6.5, “Summary Table for an Electronic Medication Administration Record (eMAR) Example (Guo, Iribarren, Kapsandoy, Perri, & Staggers, 2011)”.

Table 6.5. Summary Table for an Electronic Medication Administration Record (eMAR) Example (Guo, Iribarren, Kapsandoy, Perri, & Staggers, 2011)

Task	Description of Usability Problem	Heuristic Violation	Severity
Administer Medications	BCMA screen does not have indicator for automatic refresh. Nurses must continually click refresh (multiple-step process) or exit out and reenter to avoid missing new orders.	• Match with the real word • Minimize memory load	4 (catastrophic)
Prepare Medications	No easy way to organize/prepare for medication administration across patients. Work-around is to run missed medication reports frequently.	• Match with the real world • Flexibility of the system	4 (catastrophic)

6.5. Synthesize and Communicate Findings

Guidance on this topic is available in the Practical Guide, chapter 8 (DVA, 2017c).

6.5.1. Generate Report

Evaluators may find guidance in NIST 7742 (Schumacher & Lowry, 2010) helpful for generating reports after a heuristic evaluation is completed. The 7742 report format is condensed and adapted for heuristic evaluation in Appendix C, *Heuristic Evaluation Report Sample*.

Chapter 7. User Observations

Readers may find a brief online overview helpful on controlled and naturalistic user observations on the Interaction Design Foundation's website [https://www.interaction-design.org/literature/article/how-to-conduct-user-observations](2017). Likewise, Goodman, et al. provides a discussion of field visits, another term for observations in naturalistic work settings (Goodman, Kuniavsky, & Moed, 2012, pp. 211-242). Detailed information about observation studies is also available from the Trent Focus Group (Fox, 1998). Steps to conduct user observations are outlined in this chapter.

7.1. Refine Assessment Goals and Create Specific Objectives

Evaluators define the primary assessment goal(s) and select user observations as a method. Now, evaluators will refine specific objectives for the particular assessment.

Sample objectives for naturalistic and controlled user observations might be:

- What kinds of usability problems do users experience with this particular CDS intervention?
- How well does the CDS intervention support decisions and actions?
- How many and what kinds of interruptions are caused by or related to the CDS intervention?
- How many steps does it take to complete the CDS intervention process? How does this compare to pre-CDS intervention?
- How many and what kind (if any) additional information sources need to be accessed to complete the CDS actions?
- Is the CDS intervention process intact? Are stakeholders completing their appropriate portion of the CDS intervention? If not, why not? Is the CDS intervention going to the correct stakeholders in the correct order?
- What is the difference in the way the CDS intervention was intended to be used versus the way it is actually used (particularly in a naturalistic environment)?

Sample objectives for workflow analysis might be:

- What is the information flow to complete a specified task or goal?
- How does the current workflow compare to the previous one (baseline or prior to this CDS intervention)?
- Who are the stakeholders involved in this task or goal?
- What are the steps being taken to complete the goal? (This can be in contrast to how the process was originally envisioned.)

Evaluators can use a guide called the AEIOU framework from the Doblin Group eLab to think about what to observe in work settings and then later to assist in coding observations (Goodman, Kuniavsky, & Moed, 2012, pp. 231-232). The AEIOU framework is helpful for detailed (micro) observations versus higher level (macro) activities:

- Activities (A) are actions participants want to accomplish;
- Environments (E) include the entire set of locations where the activities occur;
- Interactions (I) occur between the participants and someone or something else (e.g., CDS interventions or people or other tools);
- Objects (O) are items in the environment (i.e., elements participants use in their work); and

- Users (U) (i.e., participants in a CDS evaluation) are the people involved in the activities.

7.2. Prepare for the Assessment

Creating a plan for user observations is imperative. It is especially important in naturalistic user observations because those settings can be complex and distracting. Also, the number and types of observations need to be limited in scope so they are not overwhelming. This planning step is worth the effort to assure the assessment goes smoothly. Evaluators will want to keep the project goal and main method in mind as they step through each of the steps in the CDS Impact Evaluation Framework.

7.2.1. Obtain Approvals

Evaluators are referred to the Practical Guide section 5.1 on obtaining approvals (DVA, 2017c).

7.2.1.1. Develop Consent Form (If Needed)

If the evaluators are conducting a more formal study, they will need to provide informed consent forms for participants to sign. See Veterans Health Administration (VHA) Handbook 1004.01 [https://www.va.gov/vhapublications/viewpublication.asp?pub_ID=2055](VHA, 2009).

7.2.2. Select Team Members

Criteria for selecting team members should be developed and used. Sample considerations are:

- At least initial training on user observations and how to detect usability problems; and
- The number of participants to be observed.

If a relatively small number of participants are to be observed, then one team member can be sufficient. However, additional team members may be necessary if a larger number of participants are being observed and a larger set of data are being analyzed.

7.2.3. Select Participants

Considerations for participants include the number and types of participants. While a small number of participants can be observed (1 for discrete goals, 3–5 for a main goal), the key is to select participants who are representative of typical CDS intervention users. Evaluators will want to think through the CDS process to identify representative end users (e.g., the type of CDS and target users). Even one participant can be highly insightful for discrete goals such as determining a task sequence, providing information for scenario development, or identifying adverse events or “show-stoppers” in a workflow. A larger number of participants than 3-5 could be observed for a main goal, although the amount of collected data would then be larger as well. Criteria can include similar ones to the heuristic evaluation:

- Clinical expertise
- Familiarity with the CDS intervention

However, unlike heuristic evaluations, evaluators may want to select participants with varying levels of expertise and familiarity with the CDS intervention. For example, a clinician new to a CDS intervention may be confused about some tasks and messages while practiced, experienced clinicians may create work-arounds and/or ignore the same messages.

7.2.4. Determine the Setting

Naturalistic observations are completed “in the wild” in a clinical setting whereas controlled observations are conducted in a quiet room, such as a conference room, a computer lab, a virtual environment, or other space

with access to the CDS intervention. Workflow observations may be in either setting. When clinical settings are used, evaluators will need to consider factors such as how obtrusive the observations will be to participants and patients, and specifically, whether patients will be included in observations. If patients are included, protections for patient privacy and formal consents for the observations are required (see Section 7.2.1, “Obtain Approvals” and Section 7.2.1.1, “Develop Consent Form (If Needed)”). Evaluators will likely observe real patient data during most observations, so privacy consideration must be made about any recordings and data must be de-identified. Evaluators may need IRB approval for the project if they involve patients and protected data.

For controlled observations, evaluators may need to create or make available a computer environment for use by the team and participants. The evaluators may need to work with the IT department to create a separate computer environment, similar to a training environment (or the training environment itself), to conduct the assessment. Participants often need to interact with fictitious data that mirror actual patients’ and clinicians’ records, so evaluators should be prepared to populate records with synthetic data if needed. Another consideration is that the environment, like a training environment, must allow refresh (reset) of the application to default data as new participants start their assessments. Logically, if the training environment is used as the testing environment, evaluators will need to work through logistics such as conflicts in scheduling training and testing sessions, refreshing schedules, etc. Evaluators will want to assure that the selected environment mirrors the production environment in cases where the training environment might be a version or two earlier than production.

7.2.5. Select Tools and Instrumentation

Evaluators will compile materials for data collection. This includes instructions about the tasks to be observed and the procedure for user observations. The evaluator will want to develop forms for data collection such as one using the AEIOU framework (Section 7.1, “Refine Assessment Goals and Create Specific Objectives”) or recording the list of usability problems as in Table 6.2, “Sample List of Usability Problems and Their Descriptions”.

7.2.5.1. Write Scenarios

Scenarios are not used for naturalistic observations. In controlled observations, scenarios are optional. However, defining observation goals and/or tasks is necessary for all user observations, so evaluators can ask participants to complete activities of interest. If a scenario is desired, please use guidance in Section 8.2.5.1, “Write Scenarios”.

7.2.5.2. Create Tasks

Tasks for any assessment should represent the common activities users employ to interact with the application at hand. For user observations, these might be more general activities and tasks (e.g., “Show me how to find and organization information about the meds you will be giving at 0900”; or “Show me how you renew a prescription for this patient who is allergic to iodine.”) In naturalistic observations, the goal may be to observe an activity, such as completing a Separation from Active Duty SMARTForm or conducting a handoff without providing any specifics about tasks.

For controlled observations, evaluators will want to think through discrete tasks to observe. Readers are referred to Section 6.2.5.1, “Create Tasks” on creating tasks.

7.2.5.3. Develop a Data Collection Plan

Evaluators will want to think through a data collection plan. Data collection methods can be as simple as entering observations into a Word table or slide deck, recording video of the session for later analysis or using a tool such as the Problem Step Recorder (a program embedded in Windows for Microsoft products to capture screen shots correlated to mouse clicks or keystrokes). Be sure to consider privacy issues if video recordings are used.

A data collection form should be prepared. Data collection forms can range from notetaking using the AEIOU framework (Section 7.1, “Refine Assessment Goals and Create Specific Objectives”) to capturing screenshots to

lists of issues. Evaluators can observe overt participant behaviors, but many healthcare activities are cognitive. Therefore, evaluators will want to ask participants to think aloud (Nielsen Norman Group, 2012; Nielsen, Usability Engineering, 1993)—that is, ask participants to talk aloud as they work—about the steps they are taking, any confusing items, or even why they are using particular tools. Participants can be encouraged to express their opinions. If participants fall quiet, observers can encourage them to begin talking anew. This technique is very helpful as participants to step through their normal processes in completing a goal during controlled observations.

The Problem Step Recorder (PSR or just Step Recorder) is available on Windows 7 and later Microsoft products. It allows evaluators to capture and store screen shots correlated to mouse clicks and keystrokes. It is an excellent resource to analyze and illustrate usability problems, especially if evaluators would like to view them in more detail later. Fisher provides tips on using the PSR in his Lifewire article (Fisher, 2017). The PSR was originally developed for users to record and send computer issues to support personnel for help; however, it works well for recording usability problems. Evaluators need to be aware that Windows 7 is limited to 100 screen captures and Windows 10 is limited to 1,000. Thus, evaluators will need another method for more lengthy processes. To access the PSR, go to the START button and type in PSR. Evaluators need to start the PSR before the participant begins the tasks.

7.2.5.3.1. Develop Assessment Sequence

The evaluators may find it helpful to write out the sequence and procedures for collecting data. A sequence for user observations testing might be:

- Welcome participants, and thank them for completing the study.
- Verify participant identities.
- If using a controlled setting, orient participants to the setting, seating, available restrooms if applicable, etc.
- Obtain written informed consent if needed or verbal consent to participate.
- Obtain demographic data.
- Read introductory script.
- Ask users to complete tasks one by one, if in a controlled setting, or provide them with other instructions on the tasks.
- Thank the participants.

7.2.5.4. Prepare Briefing Instructions

A script can be to read to participants before data collection occurs. This assures consistency in the way instructions are delivered and helps avoid gaps in important information. The script is easier to write once all the elements of the assessment are outlined. A script typically contains these elements:

- The purpose of the study;
- The role of the participant;
- What the participant can expect to experience (e.g., tasks, length of time);
- Where to ask for help or to whom to address questions about the assessment;
- A statement that the assessment will have no correlation to their job performance;
- For naturalistic observations, the observer may be asked to stop an observation or not follow the participant into a patient room at any time in the process; and
- A statement that they may end the observation at any time.

Sample Script Adapted from NIST 7742

Thank you for participating in this assessment. Our session today will last [XX minutes]. During that time, we are asking you to work with the [list the type] CDS intervention.

I will ask you to complete tasks using this CDS intervention and talk aloud about your experiences as you step through the CDS interface. We are interested in the actions you take with this CDS intervention, how easy or how difficult this CDS intervention is to use, what actions are useful to you, and how we could improve it. You will be asked to complete these tasks and we will take notes as we observe and listen to you. We will be recording (your voice or the screens). Please be honest with your feedback.

The product you will be using today is [describe the state of the application (i.e., production version, early prototype, etc.)].

We will [describe what the evaluators will be doing, such as observing participant actions quietly or recording the audio and screenshots of the session]. All of the information that you provide will be kept confidential and your name will not be associated with your comments at any time.

Do you have any questions or concerns?

7.2.5.5. Develop Test Materials Packet

Evaluators will want to develop a packet of testing materials whether electronic or paper. The packet can include, for example: the consent form (usually paper), the demographics form, instructions, the scenario (if used), observation goals, and required tasks. Each packet should be coded with a participant identification number versus any personally identifiable information such as names.

7.2.5.5.1. Develop a Demographic Questionnaire

Evaluators will want to have participants fill out demographic information about themselves. Any specific information should be linked to the purpose of the assessment. For instance, race is not related to application interactions so would not be collected, but specialty could affect the interaction. Sample items for participant demographics might include:

- Profession;
- Specialty;
- Educational level;
- Length of time in the profession;
- Length of time in the institution;
- Whether the participant has used the application in their work (If so, for how long?);
- Use of other technology pertinent to the study (e.g., CDS intervention in another institution, mobile apps to assist in actions and decisions, etc.);
- Age;
- Gender; and
- Unit or ward.

7.2.5.6. Develop a Communication Plan

Readers are referred to Practical Guide section 5.5.7 for general information on developing a communication plan (DVA, 2017c).

7.3. Conduct Observations

Evaluators and participants are highly interactive during user observations. However, evaluators will want to be as unobtrusive as is reasonable in naturalistic settings. Participants should feel free to ask evaluators to wait outside a patient room and not observe some activities, for example.

7.3.1. Collect Data

The evaluators will enact the data collection plan as per Section 7.2.5.3, “Develop a Data Collection Plan”. Evaluators should be aware that user observations involve higher volumes of detailed data collection and can be time-consuming.

7.3.2. Analyze Workflows

In the analysis process, evaluators outline the process steps in the workflow including decision points (e.g., steps following a yes or no decision). The Office of the National Coordinator for Health Information Technology (ONC) provides guidance on workflow process mapping. For example, evaluators may want to reengineer workflow and then measure the impact pre- and post-CDS implementation. Or evaluators may use the techniques to map workflow at one point in time (ONC, 2017c). A HealthIT-enabled CQI (eCQI) worksheet is available for inpatient (ONC, 2015a) and outpatient assessments (ONC, 2015c). A detailed instructor’s manual is also available from the ONC (2012).

Figure B.1 in Appendix B of the Practical Guide (DVA, 2017c) shows a workflow diagram of a simple clinic visit (HITRC, 2011). Readers can see how the process evolves and the notation used to depict aspects of the visit (e.g., a decision point). A very brief description of workflow processes is in Osheroff et al. (2012, p. 106), which provides a blank sample workflow process diagram.

7.4. Analyze Data

Two common results from user observations are discussed here: usability problems and workflow processes. When identifying usability problems, evaluators can create products similar to Table 6.2, “Sample List of Usability Problems and Their Descriptions” 5 or a listing of issues, such as the number of interruptions. For data analyses during controlled observations, evaluators may want to complete a detailed analysis of data as participants step through interactions. In that case, evaluators could consider:

- The most critical items for redesign;
- Patient safety issues;
- Deviations from projected task sequences; and/or
- Failed tasks.

During observation analyses, the evaluator may uncover, for instance, that participants used multiple ways of completing the same task or, in extreme cases, they may find that some participants did not achieve the intended goal for a task (e.g., they overrode a salient medication alert because it interrupted them at the wrong time in the sequence of ordering). The evaluators may decide to further analyze the data by classifying them into heuristic evaluation categories or by grouping them into similar kinds of usability problems.

If the technique is used for design or redesign, participants could meet with the evaluator (and/or designer). This can be an interactive session, and evaluators can record any additional comments that were not captured during the observations. Issues and recommendations can then be prioritized. For example, failed tasks and patient

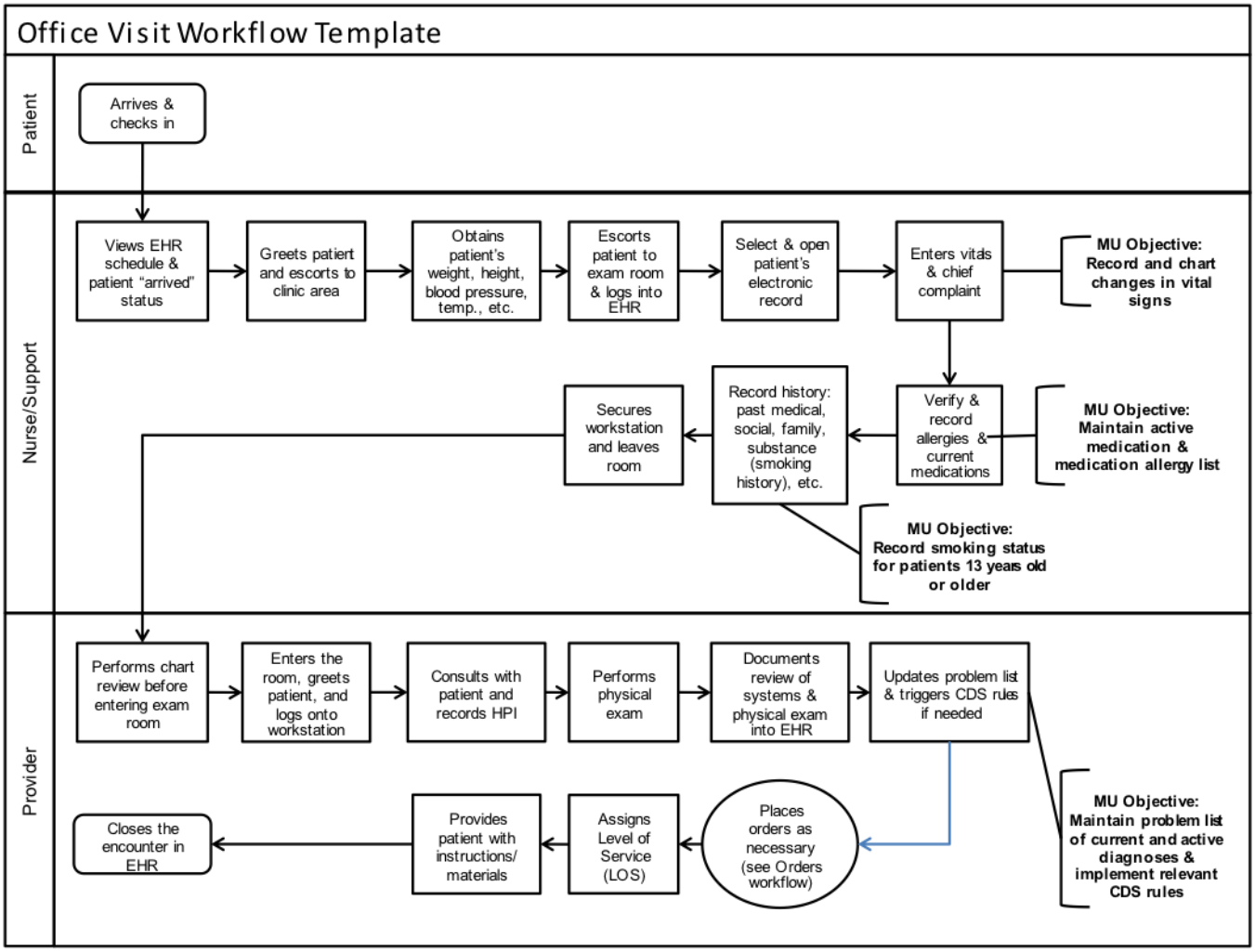
safety issues should be taken into consideration when prioritizing issues and recommendations. Likewise, issues that prevented participants from completing a task or from achieving the goal should be placed as a higher priority than issues that were only found to be inconvenient.

After the prioritization, the evaluator can analyze tasks and sequences in more detail across participants to detect patterns. Analyses using demographic data may help detect these patterns. Perhaps all of the participants had trouble completing a task except one, and that participant may have had previous experience with that type of clinical CDS intervention. Evaluators might examine demographic data related to the failed tasks.

7.4.1. Diagram Workflows

Another typical product from user observations is a diagram of workflow processes. These diagrams outline stakeholders (roles), activities, and information flow as in Figure 7.1, “Sample Workflow Diagram (HITRC, 2011)”. Previously mentioned guidance from the ONC on eCQI worksheets (Section 7.3.2, “Analyze Workflows”) can be translated into diagrams similar to Figures B.1 and B.2 in Appendix B of the Practical Guide (DVA, 2017c).

Figure 7.1. Sample Workflow Diagram (HITRC, 2011)



7.5. Synthesize and Communicate Findings

If readers need to review this material, they are referred to chapter 8 in the Practical Guide (DVA, 2017c).

7.5.1. Generate Report

Evaluators may find guidance in NIST 7742 helpful for generating reports for user observations. The 7742 report format is condensed and adapted for user observation in Appendix D.

Chapter 8. Summative Usability Testing

8.1. Refine Assessment Goals and Create Specific Objectives

Evaluators may use the categories in the General Model to frame usability test questions and/or set benchmarks for summative testing. Often summative testing includes multiple assessment objectives. Depending upon the purpose of the assessment, evaluators may want to set benchmarks for the interactions. Sample benchmarks might be:

- Navigating to an admission assessment takes no more than two clicks;
- Pulling up an admission assessment form takes less than 1 second;
- Entering admission data for a surgical patient takes no longer than 20 minutes; or
- Entering data and viewing a fall risk score takes no more than 10 minutes.

Evaluators may wish to add several subjective (qualitative) questions for participants to answer. A mixed-methods design of both quantitative and qualitative data can be very informative. General questions might ask participants to:

- Describe the most helpful features of the CDS; or
- Outline areas in the CDS that need improvement.

Sample summative usability questions or objectives mirror the categories in the General Model and are listed in Table 8.1, “Sample Summative Usability Questions/Objectives by General Model Category”.

Table 8.1. Sample Summative Usability Questions/Objectives by General Model Category

Category	Questions
Effectiveness	• Can users successfully complete required tasks? • Is the major goal accomplished (the correct diagnosis or intervention)? • What is the perceived mental workload for this CDS intervention? • What improvements do participants suggest?
Safety	• How many errors do participants make as they complete a task? • What kind of errors do users make as they complete tasks (navigation, typographical, potential patient safety errors)? • What is the severity of those errors (critical and patient safety errors, risks to patient safety)? • Do participants arrive at the correct conclusion/diagnosis for a sequence of actions?
Efficiency	• How long or how many clicks does it take participants to navigate to the needed form? • How long does it take participants to complete each task or goal? • How many clicks are required to complete each task or goal?

Category	Questions
	• How many interruptions occur during a task (if observed in a natural setting)? • How many different screens, modules, or artifacts do participants need to access to obtain the information they need to complete a goal?
Satisfaction	• How satisfied are participants with the product overall? • What aspects of the interaction are most positive? • What needs to be improved?

8.2. Prepare for the Assessment

Creating a plan is an imperative when performing summative usability testing on a CDS intervention. Because it is more formal, evaluators will want to outline each step of the assessment in detail before beginning data collection. This planning step will be time-consuming, but it is worth the effort to assure the assessment goes smoothly. Evaluators will want to keep the primary assessment goal and main method(s) in mind as they step through each of these considerations when creating: scenarios, tasks, the test environment, participant and moderator instructions, training materials, and required instruments.

8.2.1. Obtain Approvals

Readers are referred to the Practical Guide section 5.1 on obtaining approvals (DVA, 2017c). IRB approvals are often needed for summative usability testing methods.

8.2.1.1. Develop Consent Forms (If Needed)

If the evaluators are conducting a more formal study, they will need to provide informed consent forms for participants to sign. See VHA Handbook 1004.01 [https://www.va.gov/vhapublications/viewpublication.asp?pub_ID=2055](VHA, 2009).

8.2.2. Select Team Members

Criteria for selecting team members should be developed and used. Sample considerations are:

- Experience with the application;
- Experience with usability testing and/or research methods;
- Interest in improving usability; and
- The number of participants to be observed.

If a relatively small number of participants are to be observed, then even one team member can be sufficient. However, summative usability evaluators may find that two team members ease the logistics and data collection efforts. Per NIST 7804, one team member's role could be to interact with the participant while the second team member collects the needed data.

8.2.3. Select Participants

Considerations for participants in summative usability testing include the number and characteristics of participants. While design evaluations require only small numbers of users, authors recommend 15–20 participants for summative usability testing to find 80% and 95% of the usability problems, respectively (Faulkner, 2003). However, even 15 participants may be a difficult sample to obtain, especially if the participants are clinicians. In that case, evaluators will want to think through the most important participant characteristics and select participants based upon the most critical characteristics. The fewer participants in summative testing, the greater variability in user performance, the fewer identified usability problems, and

the less confidence in results. In the end, evaluators will need to balance available resource constraints against time, the likely validity of the findings and the criticality of the particular CDS evaluation or assessment being conducted.

As part of the plan, evaluators will want to develop a set of inclusion and exclusion criteria for participants. Inclusion criteria might be:

- Representative users of the application. (The interdisciplinary nature of many applications should be taken into consideration; think across physicians, nurses, pharmacists, aids or technicians, unit clerks, and administrators.);
- A range of novice versus expert users of technology in general and for this application specifically;
- A range of users new to the organization as well as long-term employees; and
- A representative age range (broad versus narrow is typically better) or other pertinent demographics.

Exclusion criteria might be:

- Participants who are less than 3 months out of their basic educational programs;
- Participants who have visual acuity issues; or
- Participants who have little or no experience with electronic health records or computerized clinical care.

8.2.3.1. Develop the Plan to Recruit Participants

Part of the assessment plan includes outlining how participants will be recruited within VHA regulations. Various methods might be used, for example:

- Fliers in clinical or public areas;
- Social media or email;
- Personal recruitment (e.g., patients) in waiting rooms or clinical areas; or
- Staff meetings.

8.2.4. Determine the Setting

Readers are referred to the second paragraph in section 7.2.4 on determining settings for controlled observations.

8.2.5. Select Tools and Instrumentation

- Summative usabiity testing methods usually include a robust set of tools and instruments. These are outlined next.

8.2.5.1. Write Scenarios

Scenarios are a story or narrative that sets the stage for a user’s interaction with an application. Scenarios should be as representative of the actual work as possible; they communicate a likely setting, characters (e.g., patients and/or clinical staff), and activities linked to the purpose of the evaluation. For instance, CDS intervention scenarios could describe a specific clinical setting, patient characteristics, and representative activities (e.g., a medical ward in an acute care or a primary care ambulatory care setting; documenting an admission assessment, suicide risk, or physical assessment).

Scenarios may be used at any stage of the systems life cycle and are particularly useful during design and early prototyping as well as with summative usability testing. Details about scenarios are available elsewhere; for example, see the textbook by Goodman, Kuniavsky, and Moed (2012).

The remainder of this section details the basic steps to construct a scenario.

8.2.5.1.1. Decide on the Basic Purpose of the Scenario

First, think about these elements:

- Consider only the most relevant settings, patients, and activities linked to the study purpose.
- Consider the most frequent activities users might complete.
- Think about other activities that, while not frequent, may be necessary to complete an activity.
- Consider a cluster or sequence of activities that are needed to complete an action.

These all can point to the basis of a scenario (e.g., determining the falls risk score for a new patient).

8.2.5.1.2. Outline the Major Elements of a Scenario

- Make the overall purpose explicit.
- Decide on a setting.
- Outline major characters or actors.
- Tell the users what major activity they are to complete.
- Outline any events that might occur.
- Create specific tasks users need to complete based upon the primary assessment goals.

Sample Scenario

You just received a new patient from the ED on your medical–surgical ward. You need to document her admission assessment and determine her fall risk and any needed interventions or support she will require while she is an inpatient and to prepare for her needs at discharge. The patient is a 79-year-old Veteran who lives alone. She has a history of diabetes Type 2 x 20 years, depression, and high blood pressure. The patient and her neighbor enjoy having evening cocktails together each day. The neighbor noticed that the patient has been moving more slowly around her apartment during the day and worries she might fall. The patient refused to go in for a routine examination. This afternoon, the neighbor found the patient in the laundry room looking dazed. The neighbor brought her into the VA and dropped her off in the ED.

In the next step, evaluators can create tasks for CDS assessment that can be combined with the scenario.

8.2.5.2. Create Tasks

Readers are referred to Section 6.2.5.1, “Create Tasks” on creating tasks.

8.2.5.3. Develop a Data Collection Plan

To plan for data collection during summative usability testing, evaluators will want to consider measures related to the usability questions and specific tasks they outlined. Table 8.2, “Sample Measures for Summative Usability Testing” outlines sample measures for summative testing.

Table 8.2. Sample Measures for Summative Usability Testing

Focus	Sample Measure
Effectiveness	• Number and percent of tasks completed (successful, partial, unsuccessful) • Cognitive workload (e.g., the National Aeronautics and Space Administration (NASA) Task Load Index (TLX) instrument) • Accuracy (errors) in navigation

Focus	Sample Measure
	• Accuracy of output or goal
Safety	• Number of errors per task • Error types (e.g., patient safety, minor) • Error severity (e.g., critical vs noncritical)
Efficiency	• Time per task • Total time to complete a sequence of tasks toward a goal • Number of clicks per task and sequence of tasks toward a goal • Number of interruptions during a task (if in a field setting)
Satisfaction	• Satisfaction score (e.g., Post-System Satisfaction Usability Questionnaire or PSSUQ instrument) • Number and types of recommendations for improvement
Subjective or Qualitative Questions	• Participants' favorite features of the application • Features needing improvement

Part of the assessment plan includes details about how to collect the data. For instance, will a stopwatch be used to time tasks or will an automated program capture the time on task and number of keystrokes (clicks)? Will the session be video recorded and analyzed? Evaluators will also want to create standard definitions where needed. Table 8.3, "Sample Measure Definitions, Adapted from NIST 7742" shows an example of definitions of various measures that point toward data analyses. Evaluators may find additional information in Tullis and Albert (2013).

Table 8.3. Sample Measure Definitions, Adapted from NIST 7742

Measures	Actions	Definition
Effectiveness: Task Success	Task times are recorded.	A task is counted as a success if the participant was able to achieve the correct outcome without assistance. If a benchmark is used, then the participant must complete the task within the time allotted on a per task basis.
Effectiveness: Partial Task Success	Task times are recorded but analyzed separately from successful tasks. Errors and error types should be collected (qualitatively).	The evaluator may wish to include this category. It can be defined as requiring moderator assistance if the participant gets lost or comes to a standstill in interactions.
Effectiveness: Task Failures	Task times may not be taken. Enumeration of errors and error types should be collected (qualitatively).	If the participant abandoned the task, did not reach the correct answer, performed the task incorrectly, or reached the end of the allotted time before successful completion, the task was counted as a failure.

8.2.5.3.1. Develop Assessment Sequence

The evaluators will find it helpful to write out the sequence and procedures for collecting data. A sequence for usability testing might be:

- Welcome participants, and thank them for completing the study.
- Orient participants to the setting, seating, available restrooms, etc.

- Verify participant identities.
- Obtain written informed consent if needed or verbal consent to participate.
- Obtain demographic data.
- Read introductory script.
- Begin data collection by providing or reading the scenario and tasks.
- Administer post-tests if needed (e.g., Post-Study System Usability Questionnaire [PSSUQ], NASA Task Load Index or TLX, subjective questions).
- Thank the participants.

Evaluators will want to develop a packet of testing materials whether electronic or paper. The packet can include, for example: the consent form (usually paper), the demographics form, instructions, the scenario (if used) with required tasks, and copies of other instruments such as the NASA TLX or PSSUQ and/or subjective questions. Each packet should be coded with a participant identification number versus any PII, such as participants' names.

8.2.5.4. Prepare Briefing Instructions

A script is created to read to participants before data collection occurs. This assures consistency in the way instructions are delivered and helps avoid gaps in important information. The script is easier to write once all the elements of the assessment are outlined. A script typically contains these elements:

- The purpose of the study;
- The role of the participant;
- What the participant can expect to experience (e.g., tasks, length of time);
- Where to ask for help or to whom to address questions about the assessment;
- A statement that the assessment will have no correlation to their job performance; and
- A statement that they may end the assessment at any time.

Sample Script Adapted from NIST 7742

> Thank you for participating in this assessment. Our session today will last [XX minutes]. During that time, you will take a look at a CDS intervention.
>
> I will ask you to complete tasks using this CDS intervention and answer some questions. We are interested in how easy or how difficult this CDS intervention is to use, what actions are useful to you, and how we could improve it. You will be asked to complete these tasks on your own, trying to do them as quickly as possible with the fewest possible errors. Please save your detailed comments until the end of a task or the end of the session as a whole when we can discuss freely. Please be honest with your feedback.
>
> The product you will be using today is [describe the state of the application (i.e., production version, early prototype, etc.)].
>
> We will [describe what the evaluators will be doing, such as observing participant actions quietly or recording the audio and screenshots of the session]. All of the information that you provide will be kept confidential and your name will not be associated with your comments at any time.
>
> Do you have any questions or concerns?

Next, evaluators can develop a script for the testing procedure. A sample script follows for moderators to read after participants understand the scenario.

Testing Procedure Sample Script

> You will be completing several tasks during this assessment and we will ask you to fill out two questionnaires afterward. Please log into the CDS intervention now. The first task is to find the admission assessment for patient [XX]. Complete the task quickly and independently with as few errors as possible. Please let me know when you are ready to begin and when you are finished with the task.

8.2.5.5. Develop Training Materials

For usability testing, the evaluators will need to decide who needs to be trained on the CDS intervention and the depth of training required. The goal is for the evaluator team and participants to have sufficient baseline knowledge about the content and navigation. Because evaluators are assessing a fielded CDS intervention, participants may not need training, but often either the team members or the participants (or both) need standardized training. A standard training method is recommended whether the standard method is computerized, trainer-based, or via text-based manuals because it assures consistency. A wise step is to have brief pretests before the main study to ensure participants are competent in basic interactions and that they will not be outliers in their interactions with the CDS. This competency test occurs after training, when participants are asked to step through a few basic steps to assure they understand the fundamentals of the interactions. This type of testing surfaces any users who have issues that might skew results, such as vision problems or very slow reaction times.

8.2.5.6. Develop Test Materials Packet

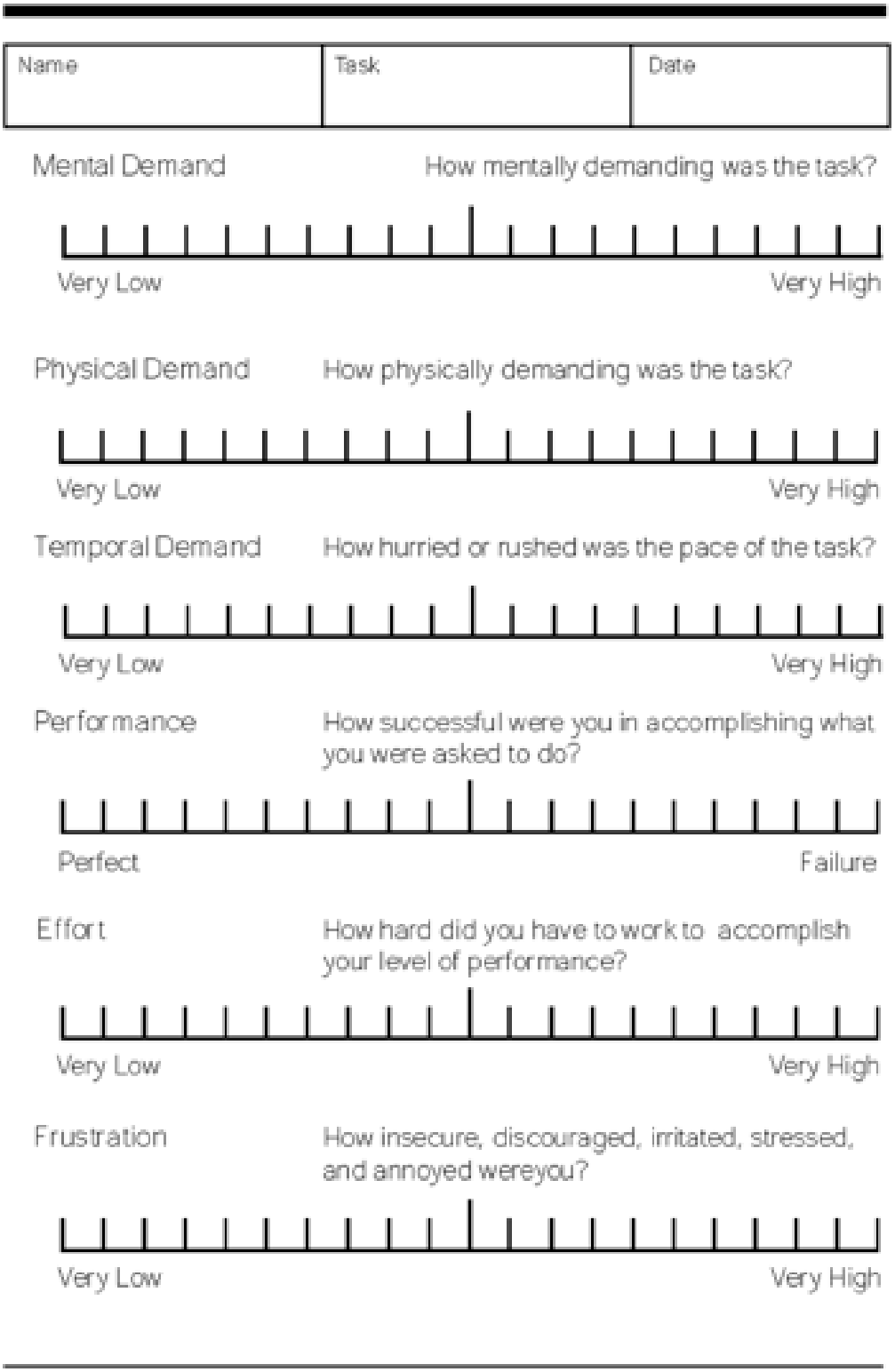
For any type of assessment, evaluators will need to gather and/or develop needed testing materials (e.g., demographic forms, data collection forms, and any existing tools needed to measure concepts of interest). These can be sequenced to display electronically or compiled into a packet (or some mixture of methods.) Evaluators can reference the developed usability testing measures from Table 8.2, “Sample Measures for Summative Usability Testing” as they create test materials.

Next, evaluators should gather or develop instruments and measures for the primary assessment goals. For example, evaluators may be interested in measuring perceived mental workload by using the NASA TLX and/or overall user satisfaction using the PSSUQ.

8.2.5.6.1. Prepare Perceived Mental Workload Instrument If Needed

Evaluators may wish to measure each participant’s perceived mental workload as mentioned in the Chapter 4, *Define Assessment Goals*. A valid and reliable instrument to measure this concept is the NASA TLX (NASA, 2017). This instrument has been used across a wide variety of complex interactions and settings including aircraft cockpits. Its use in healthcare is expanding. The instrument is available as an app or PDF. It contains six dimensions as may be seen in Figure 8.1, “NASA TLX”. Evaluators may also use pairwise comparison, although these require training and are time-consuming to administer. More detailed information is available about this instrument on the NASA TLX website [https://humansystems.arc.nasa.gov/groups/TLX/], which includes a user guide [https://humansystems.arc.nasa.gov/groups/TLX/downloads/TLX_pappen_manual.pdf] for this instrument. A summary of this method is also available at the Agency for Healthcare Research and Quality website [https://healthit.ahrq.gov/health-it-tools-and-resources/workflow-assessment-health-it-toolkit/all-workflow-tools/nasa-task-load-index](HHS AHRQ, 2017).

Figure 8.1. NASA TLX



8.2.5.6.2. Prepare a User Satisfaction Instrument If Needed

Subjective satisfaction is often measured as part of a summative usability testing session. Satisfaction is one of the main concepts in the definition of usability as readers may recall from the General Model, section 1.6.2.

A number of valid and reliable instruments exist to measure perceptions of end-user satisfaction. One of the most common is the System Usability Scale or SUS (see Appendix B.) Its use in healthcare expanded in recent years and a recent systematic review outlines its strengths (Sousa & Dunn Lopez, 2017); however, the VA is moving away from the SUS and recently recommended two other valid and reliable scales, the Post-Study System Usability Questionnaire (PSSUQ) and a related instrument, the Computer Satisfaction User Questionnaire (CSUQ).

The PSSUQ and CSUQ were both developed by IBM (Lewis, IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use, 1993), and both can be used to assess satisfaction after participants interact with a system. A brief summary about the PSSUQ is available on the Cone Trees website [http://www.conetrees.com/2010/12/ux-glossary/post-study-system-usability-questionnaire-pssuq/](Rautela, 2010). A technical report about the development and initial testing of both the PSSUQ and CSUQ is available on Lewis' website [http://drjim.0catch.com/usabqtr.pdf]. The PSSUQ was developed first and used in a preliminary usability testing session. The CSUQ was developed to expand psychometric testing. The instruments are identical and measure the same concepts; they have only minor wording differences. For example, item 3 on the PSSUQ says “I could effectively complete the tasks and scenarios using this system” while item 3 on the CSUQ says “I could effectively complete my work using this system.” A May 2017 review of usability instruments indicated that the SUS, PSSUQ and CSUO were among the strongest available satisfaction instruments based upon their assessed psychometrics, the concepts they address and the authors’ instrument quality rating (Sousa & Dunn Lopez, 2017). Readers should be aware that none of the available instruments include items on error recovery or how easy the system is to remember over time.

Evaluators will want to match the instrument selection to the purpose of their assessment and consider the importance of the concepts each instrument measures. Lewis (1993) indicates that the PSSUQ should be used for usability testing and the CSUQ can be used in field testing. For summative usability testing then, PSSUQ is likely the better choice. The instrument consists of 19 items measuring various three factors: usefulness, information quality, and interface quality (Lewis, 2002; Lewis, 1993). The actual items on the scale are in Table 8.4, “Post-Study System Usability Questionnaire (Lewis, IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use, 1993)”. The instrument employs a 7-point Likert scale ranging from 1 (strongly agree) to 7 (strongly disagree) and an N/A choice. A total of 100 points is possible. The instrument has adequate psychometric properties (Fruhling & Lee, 2005; Lewis, 2002; Lewis, 1993). Readers should be aware that the psychometric evaluations, while valid, do not yet include robust assessments of health systems.

Table 8.4. Post-Study System Usability Questionnaire (Lewis, IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use, 1993)

1. Overall, I am satisfied with how easy it is to use this system									
N/A	**Strongly Agree**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Strongly Disagree**
Comments									
2. It was simple to use this system									
N/A	**Strongly Agree**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Strongly Disagree**
Comments									
3. I could effectively complete the tasks and scenarios using this system									
N/A	**Strongly Agree**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Strongly Disagree**
Comments									
4. I was able to complete the tasks and scenarios quickly using this system									
N/A	**Strongly Agree**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Strongly Disagree**
Comments									

5. I was able to efficiently complete the tasks and scenarios using this system									
N/A	**Strongly Agree**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Strongly Disagree**
Comments									
6. I felt comfortable using this system									
N/A	**Strongly Agree**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Strongly Disagree**
Comments									
7. It was easy to learn to use this system									
N/A	**Strongly Agree**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Strongly Disagree**
Comments									
8. I believe I could become productive quickly using this system									
N/A	**Strongly Agree**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Strongly Disagree**
Comments									
9. The system gave error messages that clearly told me how to fix problems									
N/A	**Strongly Agree**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Strongly Disagree**
Comments									
10. Whenever I made a mistake using the system, I could recover easily and quickly									
N/A	**Strongly Agree**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Strongly Disagree**
Comments									
11. The information (such as on-line help, on-screen messages, and other documentation) provided with this system was clear									
N/A	**Strongly Agree**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Strongly Disagree**
Comments									
12. It was easy to find the information I needed									
N/A	**Strongly Agree**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Strongly Disagree**
Comments									
13. The information provided for the system was easy to understand									
N/A	**Strongly Agree**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Strongly Disagree**
Comments									
14. The information was effective in helping me complete the tasks and scenarios									
N/A	**Strongly Agree**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Strongly Disagree**
Comments									
15. The organization of information on the system screens was clear									
N/A	**Strongly Agree**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Strongly Disagree**

Comments									
16. The interface of this system was pleasant									
N/A	**Strongly Agree**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Strongly Disagree**
Comments									
17. I liked using the interface of this system									
N/A	**Strongly Agree**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Strongly Disagree**
Comments									
18. This system has all the functions and capabilities I expect it to have									
N/A	**Strongly Agree**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Strongly Disagree**
Comments									
19. Overall, I am satisfied with this system									
N/A	**Strongly Agree**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Strongly Disagree**
Comments									

8.2.5.6.3. Develop a Subjective Comment Form

Evaluators can develop at least a few subjective (qualitative) questions for participants to discuss verbally or in writing. If verbal, responses can be recorded for later analyses. Qualitative questions can give evaluators rich data about why CDS interventions are difficult to use, provide examples of needed improvements and features that should be retained. Sample questions adapted from NIST 7742 follow:

1. What was your overall impression of this system?
2. What aspects of the system did you like most?
3. What aspects of the system did you like least?
4. Were there any features that you were surprised to see?
5. What features did you expect to encounter but did not see? That is, is there anything that is missing in this application?
6. Compare this system to other systems you have used.
7. Would you recommend this system to your colleagues?

8.2.5.6.4. Develop a Demographic Questionnaire

Readers are referred to Section 7.2.5.5.1, “Develop a Demographic Questionnaire” on creating demographic questionnaires.

8.2.5.7. Develop a Communication Plan

Readers are referred to Practical Guide section 5.5.7 for general information on developing a communication plan (DVA, 2017c).

8.3. Conduct Assessment

Evaluators moderate sessions with participants and facilitate the data collection.

8.3.1. Pilot Test

Before conducting the assessment with the full set of scheduled participants, evaluators should strongly consider running one to three participants through the entire data collection procedure as a pilot test. This accomplishes four main goals: (a) a pilot surfaces technical issues with the technology, the training, or the instruments; (b) it identifies any unclear instructions or steps; (c) it allows the evaluators to refine the procedure before devoting resources and energy to the main assessment; and (d) it gives an estimate for the total data collection period. The latter is especially important for clinician recruitment as their time is often limited.

8.3.2. Collect Data

The evaluators will want to use structured methods to collect data. Evaluators observe individual participants as they interact with the product and score their interactions according to the assessment plan. Evaluators interact little with the participants in summative testing, but they do facilitate the conduct of the test through instrument administration. For instance, evaluators may verbally ask participants (or have written instructions) to complete a task. Evaluators would then just observe participants and only assist participants if they come to a standstill in their interactions. This assistance is recorded as part of data collection.

Evaluators can use various techniques to collect data. Simpler methods are electronic or paper data collection forms to record findings during data collection. Video recordings and Problem Step Recording capabilities are also excellent choices, although they are more complex to analyze. A sample data collection form is in Table 8.5, “Sample Data Collection Form”.

Table 8.5. Sample Data Collection Form

Task	Measure	Result
Navigate to the admission assessment	Time	2 seconds
Navigate to the admission assessment	Errors	0
Navigate to the admission assessment	Task success	Full success
Navigate to the admission assessment	Type of error	NA
Enter data into the admission assessment	Time	20 minutes
Enter data into the admission assessment	Errors	2
Enter data into the admission assessment	Type of error	1 missing field (routine vital sign and mini-mental) 1 critical error (mini-mental score)
Enter data into the admission assessment	Task success	Partial (needed coaxing to find all sections)
Access falls risk score	Time	5 seconds
Access falls risk score	Errors	5 (navigation errors)
Access falls risk score	Task success	Unsuccessful (scrolled up and down many times and could not locate falls risk score)
Access falls risk score	Type of error	1 critical 4 minor

8.4. Analyze Data

The first step in data analysis is to determine the type of data being analyzed: (a) numerical such as task times; scores on instruments or number of errors (called interval level data); (b) counts for types of data such as types of errors or severity (categorical data); and (c) comments (qualitative data). Each is analyzed differently.

Numerical data are analyzed with a statistical package or a spreadsheet using descriptive statistics to determine means (averages), standard deviations, and the data range. If evaluators want to perform statistical analyses beyond basic descriptive statistics, they will want to collaborate with a researcher or statistician. These are calculated for each task across participants. Data are aggregated versus being listed by participant as indicated in 7742; the aggregation protects participant privacy. As part of this step, evaluators will want to review instrument instructions to score the NASA TLX and the PSSUQ. NASA publishes a user guide [https://humansystems.arc.nasa.gov/groups/tlx/downloads/NASA_TLX_for_iOS_User_Guide_Final.pdf] for the NASA TLX. The PSSUQ (or CSUQ) scores are averaged across items and/or subscales (see the appendix in Lewis, 1993). Subscale scores can be calculated as well as a score for overall satisfaction:

- System use: items 1-8;
- Information Quality: items 9-15;
- Interface quality: 16-18; and
- Overall satisfaction: items 1-19.

For other analyses, counts or categorical data can be summed and averaged by task. These data are aggregated across participants and not reported for each participant due to privacy issues. Comments (qualitative data) can be formally analyzed using qualitative software, or they may be grouped into categories using a database or spreadsheet if the number of comments is not voluminous. If qualitative methods are being used, evaluators may want to collaborate with a researcher who has skills in this method.

8.5. Synthesize and Communicate Findings

Readers are referred to material in the Practical Guide, chapter 8 on synthesizing findings across objectives and generating a report.

8.5.1. Generate Report

Evaluators may find guidance in NIST 7742 helpful for generating reports after usability testing is completed. The 7742 report format is condensed and adapted in Appendix E for the summative usability testing method.

Chapter 9. Conclusion

The Assessment Protocol presented in this document detailed three main methods of assessing a CDS intervention using heuristic evaluation, user observations, and summative usability testing. As discussed in the Chapter 2, *Assumptions*, the intent of this Assessment Protocol is not to provide an all-inclusive reference for all methods to evaluate a CDS intervention. Rather, it is a resource to help guide clinicians and QI teams at the local hospital level in assessing an implemented CDS intervention.

Appendix A. Works Cited

Faulkner, L. (2003, August). Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. *Behavior Research Methods, Instruments, & Computers, 35*(3), 379-383.

Fisher, T. (2017). *How to use Steps Recorder*. Retrieved from Lifewire: https://www.lifewire.com/how-to-use-steps-recorder-2626159

Fox, N. (1998). *How to use observations in a research project.* Retrieved from Simmons College: http://web.simmons.edu/~tang2/courses/CUAcourses/lsc745/sp06/observation.pdf

Fruhling, A., & Lee, S. (2005). Assessing the reliability, validity and adaptability of PSSUQ. *AMCIS 2005 Proceedings*, (p. 378).

Georgsson, M., & Staggers, N. (2016a). Quantifying usability: An evaluation of a diabetes mHealth system on effectiveness, efficiency, and satisfaction metrics with associated user characteristics. *Journal of the American Medical Informatics Association, 23*(1), 5-11.

Georgsson, M., Staggers, N., & Weir, C. (2016b). A modified user-oriented heuristic evaluation of a mobile health system for diabetes self-management support. *Computers, Informatics, Nursing, 34*(2), 77-84.

Goodman, E., Kuniavsky, M., & Moed, A. (2012). Observing the user experience. Amsterdam, Netherlands: Morgan Kaufmann.

Guo, J., Iribarren, S., Kapsandoy, S., Perri, S., & Staggers, N. (2011). Usability evaluation of an electronic medication administration record (eMAR) application. *Applied Clinical Informatics, 2*(2), 202-224.

Harrington, L., Porch, L., Acosta, K., & Wilkens, K. (2011). Realizing electronic medical record benefits: An easy-to-do usability study. *The Journal of Nursing Administration, 41*(7-8), 331-335.

Health Information Technology Research Center (HITRC) Practice and Workflow Redesign Community of Practice. (2011). *Workflow Redesign Templates.* Retrieved from https://www.healthit.gov/sites/default/files/tools/workflow_redesign_templates_v1.pptx

Interaction Design Foundation. (2017, February). *How to conduct user observations*. Retrieved from https://www.interaction-design.org/literature/article/how-to-conduct-user-observations

Jeffries, R., Miller, J. R., Wharton, C., & Uyeda, K. (1991). User interface evaluation in the real world: A comparison of four techniques. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 119-124). New York: ACM.

Lavery, D., Cockton, G., & Atkinson, M. P. (1997). Comparison of evaluation methods using structured usability problem reports. *Behavior and Information Technology, 16*(4-5), 246-266.

Lewis, J. R. (1993). *IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use.* IBM. Retrieved from http://drjim.0catch.com/usabqtr.pdf

Lewis, J. R. (2002). Psychometric evaluation of the PSSUQ using data from five years of usability studies. *International Journal of Human–Computer Interaction, 14*(3-4), 463-488.

Lowry, S. Z., Quinn, M. T., & Ramaiah, M. (2012). *Technical evaluation, testing and validation of the usability of electronic health records.* Washington: National Institute of Standards and Technology (NIST). Retrieved from https://www.nist.gov/node/592206

National Aeronautics and Space Administration. (2017). *Task Load Index (TLX)*. Retrieved from https://humansystems.arc.nasa.gov/groups/tlx/index.php

Nielsen Norman Group. (2012, January 12). *Thinking Aloud: The #1 Usability Tool*. Retrieved from https://www.nngroup.com: https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/

Nielsen, J. (1992). Finding usability problems through heuristic evaluation. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 373-380). New York: ACM.

Nielsen, J. (1993). *Usability Engineering.* Boston, MA, USA: Academic Press.

Nielsen, J., & Mack, R. L. (Eds.). (1994). *Usability inspection methods.* New York: Wiley.

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 249-256). New York: ACM.

Office of the National Coordinator for Health Information Technology. (2012). *Fundamentals of health workflow process analysis & redesign instructor manual.* Retrieved from HealthIT.gov: https://www.healthit.gov/sites/default/files/Comp-10-Instructor-Manual.pdf

Office of the National Coordinator for Health Information Technology. (2015a, May). *Health IT-enabled quality improvement (eCQI) worksheet (inpatient, essential version).* Retrieved March 2017, from HealthIT.gov: https://www.healthit.gov/sites/default/files/cds/eCQI-Worksheet-Inpatient-Essential-05-15.pdf

Office of the National Coordinator for Health Information Technology. (2015c, May). *Health IT-enabled quality improvement (eCQI) worksheet (ambulatory, essential version).* Retrieved March 2017, from HealthIT.gov: https://www.healthit.gov/sites/default/files/cds/eCQI-Worksheet-Ambulatory-Essential-05-15.pdf.

Office of the National Coordinator for Health Information Technology. (2017c). *What is workflow redesign? Why is it important?* Retrieved from HealthIT.gov: https://www.healthit.gov/providers-professionals/faqs/ehr-workflow-redesign

Osheroff, J. A., Teich, J. M., Levick, D., Saldana, L., Velasco, F. T., Sittig, D. F., . . . Jenders, R. A. (2012). *Improving outcomes with clinical decision support: An implementer's guide* (2nd ed.). Chicago: HIMSS.

Rautela, A. (2010, December). *Post-study system usability questionnaire (PSSUQ)*. Retrieved from Cone Trees: http://www.conetrees.com/2010/12/ux-glossary/post-study-system-usability-questionnaire-pssuq/

Rubin, J., & Chisnell, D. (2008). *Handbook of usability testing: How to plan, design, and conduct effective tests.* Indianapolis, IN: Wiley Pub.

Sauro, J. (2011, February 2). *Measuring usability with the system usability scale (SUS)*. Retrieved from Measuring U: https://measuringu.com/sus/

Schumacher, R. M., & Lowry, S. Z. (2010, November 16). *Customized common industry format template for electronic health record usability testing.* Washington: National Institute of Standards and Technology (NIST). Retrieved from https://www.nist.gov/publications/nistir-7742-customized-common-industry-format-template-electronic-health-record

Schumacher, R. M., & Lowry, S. Z. (2010). *NIST guide to the processes approach for improving the usability of electronic health records.* National Institute of Standards and Technology. Retrieved from https://www.nist.gov/sites/default/files/documents/healthcare/usability/Guide_Final_Publication_Version.pdf

Sousa, V., & Dunn Lopez, K. (2017). Towards usable e-health: A systematic review of usability questionnaires. *Applied Clinical Informatics, 8*(2), 470-490.

Staggers, N., Iribarren, S., Guo, J. W., & Weir, C. (2015). Evaluation of a BCMA's electronic medication administration record. *Western Journal of Nursing Research, 37*(7), 899-921.

Tullis, T., & Albert, B. (2013). *Measuring the user experience* (2nd ed.). Amsterdam, North Holland, Netherlands: Elsevier/Morgan Kaufmann.

U.S. Department of Health and Human Services. (2017, March). *Workflow Assessment for Health IT Toolkit*. Retrieved from Agency for Healthcare Research and Quality National Resource Center Health IT Tools and Resources: https://healthit.ahrq.gov/health-it-tools-and-resources/workflow-assessment-health-it-toolkit/all-workflow-tools/nasa-task-load-index

U.S. Department of Veterans Affairs, Office of Informatics and Information Governance, Health Informatics, Office of Knowledge Based Systems. (2017b). *General model to evaluate clinical decision support (CDS) interventions.* Department of Veterans Affairs, Washington, DC.

U.S. Department of Veterans Affairs, Office of Informatics and Information Governance, Health Informatics, Office of Knowledge Based Systems. (2017c). *Practical guide for clinical decision support (CDS) evaluation.* Washington, D.C.: U.S. Department of Veterans Affairs.

U.S. Department of Veterans Affairs, Veterans Health Administration. (2009, August 14). *Informed consent for clinical treatments and procedures*. Retrieved from https://www.va.gov/vhapublications/viewpublication.asp?pub_ID=2055

University of Texas, School of Biomedical Informatics, National Center for Cognitive Informatics and Decision Making in Healthcare. (2017). *EHR usability general principles & guidelines*. Retrieved March 2017, from https://sbmi.uth.edu/nccd/ehrusability/design/guidelines/

Zhang, J., Johnson, T. R., Patel, V. L., Paige, D. L., & Kubose, T. (2003). Using usability heuristics to evaluate patient safety of medical devices. *Journal of Biomedical Informatics, 36*(1-2), 23-30.

Appendix B. System Usability Scale

The SUS (Table B.1, “System Usability Scale”) is one example of a valid and reliable tool to measure overall user satisfaction. The instrument is widely used outside of healthcare and is becoming more accepted within healthcare. Beside its extensive use, its advantages include its broad applicability to applications of all kinds and its simplicity. Sauro published an SUS scoring summary [https://measuringu.com/sus/] on the Measuring U website (Sauro, 2011).

Table B.1. System Usability Scale

Question	Strongly disagree (1)	2	3	4	Strongly agree (5)
1. I think that I would like to use this system frequently.					
2. I found the system unnecessarily complex.					
3. I thought the system was easy to use.					
4. I think that I would need the support of a technical person to be able to use this system.					
5. I found the various functions in this system were well integrated.					
6. I thought there was too much inconsistency in this system.					
7. I would imagine that most people would learn to use this system very quickly.					
8. I found the system very cumbersome to use.					
9. I needed to learn a lot of things before I could get going with this system.					
10. I felt very confident using the system.					

Appendix C. Heuristic Evaluation Report Sample

CDS Intervention Assessment Summary:

[Name of CDS and Version Tested]

Date of CDS Intervention Assessment: [Date CDS Intervention Assessment Was Conducted]

Date of Report: [Date Report Was Prepared]

Report Prepared By: [Contact Person, Title, and Affiliation]

[Phone Number]

[Email Address]

Summary of the CDS Intervention Assessment

An assessment of the CDS Intervention for [name of product, version, and EHR] was conducted on [date] in [location] by [site or laboratory]. The purpose of this test was to inspect the CDS intervention for usability problems, classify the problems using heuristic evaluation categories, and rate the problem severities.

Evaluators

[XX single- or dual-domain] experts served as evaluators. [Describe their characteristics. A sample paragraph follows: Heuristic evaluation reports include a description about the evaluators, typically in text format (versus a table as with other methods. A sample might be: The three evaluators are dual-domain evaluators in medicine and informatics. They have extensive usability experience in health IT, including CDS interventions, and with clinical aspects related to this specific CDS intervention.]

Setting and CDS Description

The assessment was completed virtually by the [XX] evaluators by accessing [XX environment]. The CDS intervention has been deployed for [describe the timeframe]. The purpose of this CDS intervention is [describe its purpose], and the target users are [describe them]. The CDS intervention is [a prototype or it has been deployed for X time]. The evaluators received training on the CDS intervention by accessing the online training module and completing the [XX-minute] training session.

Tasks

The evaluators defined representative tasks related to this CDS intervention. The specific [number of] tasks are [List the tasks. A sample list follows]:

- Navigate to the [admission assessment] form.
- [Enter provided assessment information including suicide risk and falls risk data.]
- [Find the falls risk score and needed clinical interventions.]
- [Find any abnormal lab results for today.]
- [Continue with list].

Procedure and Data Analyses

The evaluators used methods described by [name authors such as Nielsen (1994) and/or Zhang et al. (2003)]. The evaluators first used the tasks to interact with the CDS intervention individually to identify usability problems. Lists of usability problems were consolidated into a master list of problems. Any discrepancies were resolved through discussion. The evaluators then classified the usability problems into the heuristic categories identified by [name authors such as Nielsen (1994) or Zhang et al. (2003)]. Finally, the usability problems were rated [individually, as a group] for severity using Nielsen's scale of 0 (no usability problem) to 4 (usability catastrophe).

Results: Usability Problems

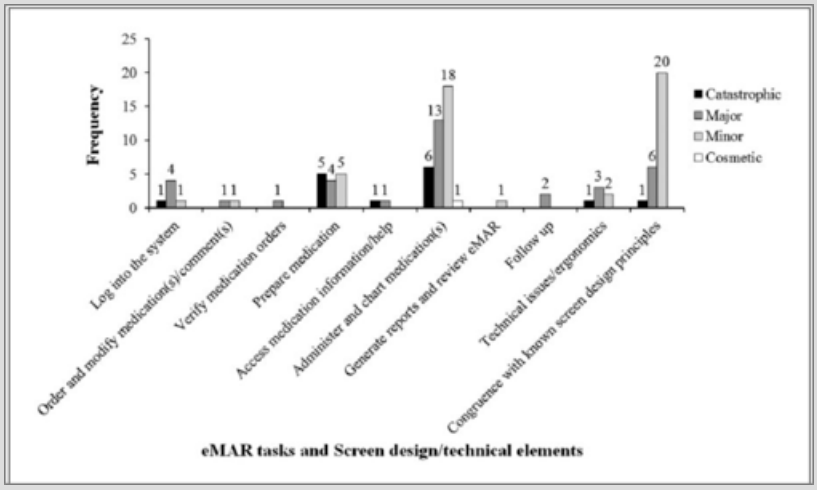
The evaluators found [XX] usability problems comprising [XX] heuristic violations across the [XX] tasks. These are summarized in Table C.1, "Sample Heuristic Violations".

Table C.1. Sample Heuristic Violations

Task	Usability Problem(s)	Example Problems	Average Severity
Navigate to admission assessment form	[Number of problems identified]	[Describe sample problems]	[List the average severity level]
Enter data into assessment form	[Number of problems identified]	[Describe sample problems]	[List the average severity level]
Find falls score, clinical interventions	[Number of problems identified]	[Describe sample problems]	[List the average severity level]
Find abnormal lab results		[Describe sample problems]	List the average severity level

The average severity of usability problems was [state the mean, CD range]. [XX] major and [XX catastrophic] usability problems were identified. The task(s) with the highest rating in severity was [describe]. Examples of catastrophic usability problems include [describe 2–3 of the most critical]. The task that was most difficult to learn was [X description] and the easiest to learn was [description]. The top recommendations for redesign included [description]. Figure C.1, "Sample Usability Problems Chart (Staggers, Iribarren, Guo, & Weir, 2015)" shows the [describe the elements such as the severity of problems by task or the type of heuristic violation by task.]

Figure C.1. Sample Usability Problems Chart (Staggers, Iribarren, Guo, & Weir, 2015)



Discussion

[Begin the discussion with a narrative summary of the major findings. One way to organize these is by the volume of violations and their severity by task. The tasks having the largest volume of usability problems were (description) and the tasks having the fewest issues were (description). The task with the highest volume of catastrophic issues was (describe). Examples of potential patient safety issues were (describe one or two).]

[Include interpretations or implications of the findings. Briefly explain which areas are critical and what the results mean for effectiveness, efficiency, patient safety, and satisfaction, especially related to workflow and cognitive impacts. An overall message or set of next steps can be included here.]

Appendix D. User Observation Report Sample

CDS Intervention Assessment Summary:

[Name of CDS and Version Tested]

Date of CDS Intervention Assessment: [Date CDS Intervention Assessment Was Conducted]

Date of Report: [Date Report Was Prepared]

Report Prepared By: [Contact Person, Title, and Affiliation]

[Phone Number]

[Email Address]

Summary of the CDS Intervention Assessment

An assessment of the CDS Intervention for [name of product, version, and EHR] was conducted on [date] in [location] by [site or laboratory]. The purpose of this test was [list the purpose such as identify usability problems, determine workflow] by employing user observations [in a naturalistic or in a controlled setting such as a quiet conference room]. The specific questions were [list them here].

Participants

[XX] healthcare clinicians [and/or other representative users] were participants. Table D.1, "Participant Characteristics (n = [XX])" summarizes participant characteristics.

Table D.1. Participant Characteristics (n = [XX])

User Characteristics	Mean	SD	Range
Gender			
Female			
Male			
Age			
Education			
High School			
Bachelor's Degree			
Master's Degree			
Doctorate			
Other demographic data			

Setting and CDS Description

The assessment was completed in [add description]. The CDS intervention has been deployed for [describe the timeframe]. The purpose of this CDS intervention is [describe its purpose], and the target users are [describe them]. The CDS intervention is [a prototype or it has been deployed for X time].

Scenario [if used] and Tasks

[If a scenario was used (e.g., in a controlled setting), then include information about it.] The evaluators developed and validated a scenario for this CDS intervention. The scenario included information about [Describe or list the scenario. The scenario could be listed in an appendix]. The evaluators asked participants to complete these representative tasks using and the CDS intervention. The specific [number of] tasks are:

- [List the tasks (e.g., Navigate to the [admission assessment] form).]
- [Enter provided assessment information including suicide risk and falls risk data.]
- [Find the falls risk score and needed clinical interventions.]
- [Find any abnormal lab results for today.]
- [Continue with list.]

Setting

The assessment was completed in [name the setting and provide a brief description such as a primary care clinic, a quiet conference room or computer training lab]. The CDS intervention was [name the intervention]. Its purpose is to [description], and the target users are [name the target users]. The CDS intervention is part of [name the EHR and its version number.]

Procedure

The user observations occurred as follows. Each participant was greeted by the administrator and asked to review and sign an informed consent/release form [provide a sample in an Appendix if required]. The moderator introduced the assessment and instructed participants to complete a demographic form. The moderator asked participants to complete tasks step-by-step as they talked aloud. The evaluator [took notes, audio or video-recorded the session] and verbally clarified any confusing steps. After the assessment, participants were asked for any comments about the activities they completed.

Data Analyses

Descriptive statistics were used to analyze any counts of activities (e.g., interruptions). The qualitative data were analyzed by [describe the product (e.g., listing usability problems by task type or diagramming workflow processes).] Content analysis was used to analyze the participants' comments.

Results: Usability Problems

Include a table of results, such as in Table 6.2, “Sample List of Usability Problems and Their Descriptions” or Table C.1, “Sample Heuristic Violations” (if the problems are classified into Heuristic Evaluation categories as part of the analyses.

Include the resulting workflow diagrams, (e.g., Figure 7.1, “Sample Workflow Diagram (HITRC, 2011)”)

Participant Comments

An analysis of subjective comments included:

- Positive features of the CDS [add details];
- Areas for improvement [add details];
- Patient safety or critical productivity issues;
- Participant reactions to how CDS fits into workflow; and
- Whether participants would recommend the CDS intervention to their colleagues.

Discussion

[Begin the discussion with a narrative summary of the major findings. One way to organize these is by the general model categories of efficiency, effectiveness, safety, and satisfaction. Next, include interpretations or implications of the findings. Briefly explain which areas are critical and what the results mean for workflow, productivity, and patient safety. Explain next steps.]

Appendix E. Summative Usability Testing Report Sample

CDS Intervention Assessment Summary:

[Name of CDS and Version Tested]

Date of CDS Intervention Assessment: [Date CDS Intervention Assessment Was Conducted]

Date of Report: [Date Report Was Prepared]

Report Prepared By: [Contact Person, Title, and Affiliation]

[Phone Number]

[Email Address]

Summary of the CDS Intervention Assessment

An assessment of the CDS Intervention for [name of product, version, and EHR] was conducted on [date] in [location] by [site or laboratory]. The purpose of this test was to test and validate the usability of the CDS intervention using summative usability testing methods. The specific questions were [list them here].

Participants

[XX] healthcare clinicians [and/or other representative users] were participants. Table E.1, "Participant Characteristics (n = [XX])" summarizes participant characteristics.

Table E.1. Participant Characteristics (n = [XX])

User Characteristics	Mean	SD	Range
Gender			
Female			
Male			
Age			
Education			
High School			
Bachelor's Degree			
Master's Degree			
Doctorate			
Other demographic data			

Setting and CDS Description

The assessment was completed in a quiet setting [add description]. The CDS intervention has been deployed for [describe the timeframe]. The purpose of this CDS intervention is [describe its purpose], and the target users are [describe them]. The CDS intervention is [a prototype or it has been deployed for X time].

Scenario and Tasks

The evaluators developed and validated a scenario for this CDS intervention. The scenario included information about [describe or list the scenario. The scenario could be listed in an appendix]. The evaluators defined representative tasks related to the scenario and the CDS intervention. The specific [number of] tasks are:

- Navigate to the [admission assessment] form.
- [Enter provided assessment information including suicide risk and falls risk data].
- [Find the falls risk score and needed clinical interventions].
- [Find any abnormal lab results for today].
- [Continue with list].

Measures

Various recommended metrics, in accordance with the examples set forth in the *NIST Guide to the Processes Approach for Improving the Usability of Electronic Health Records*(Schumacher & Lowry, 2010), were used to evaluate usability. The following measures were collected for each participant:

- Number of tasks completed successfully, partially completed and failed;
- Time to complete the tasks;
- Number and types of errors;
- The severity of the errors;
- Participant's overall satisfaction with the CDS using the PSSUQ;
- Participants' perceived mental workload; and
- Subjective comments about the positive aspects of the CDS and areas needing improvement.

Setting

The assessment was completed in [name the setting and provide a brief description such as a quiet conference room or computer training lab]. The CDS intervention was [name the intervention]. Its purpose is to [description], and the target users are [name the target users]. The CDS intervention is part of [name the EHR and its version number.]

Procedure

The assessment procedure was as follows. Each participant was greeted by the administrator and asked to review and sign an informed consent/release form [provide a sample in an Appendix if required]. The moderator introduced the assessment and instructed participants to complete a demographic form. Training on the CDS intervention consisted of [describe the training]. Participants read the scenario and then completed a series of tasks one at a time. Participants were asked to verbalize the start and end of tasks. The moderator timed the tasks, recorded user performance data on paper and electronically. After the assessment, participants completed the PSSUQ, the NASA TLX and they described their impressions of the CDS intervention.

Data Analyses

For data analyses, all participant data was de-identified. Descriptive statistics were used to analyze performance data, PSSUQ and NASA TLX scores. Content analysis was used to analyze the comments and other subjective data.

Results: Performance Data

Participants completed the assessments in an average of [time] minutes. Performance data are summarized in Table E.2, “Performance Data by Task Type”.

Table E.2. Performance Data by Task Type

Task	Task Success	Task Time		Errors	Error Type
	Mean (SD)	Mean (SD)	Range	Mean (SD)	Critical (Patient safety) / Navigation / Minor
[Find admission assessment]					
[Find abnormal labs for today]					
[Find fall risk score]					
Other tasks					

The task taking the longest time was [fill in data]. Participants completed [X task] in the shortest amount of time. The task having the highest average error was [x] and the lowest was [x]. The types of errors were [describe] and their severity included [describe].

PSSUQ and TLX

The results from the PSSUQ and NASA TLX are in Table E.3, “PSSUQ & TLX Results”. As may be seen, the average PSSUQ score was: [xx]. Higher scores indicate improved usability. Evaluators may want to set an acceptable score a priori. The scores for this CDS intervention indicate [poor, acceptable, above average] usability. The scores from the NASA TLX were [xx]. This indicates [high, moderate, low] perceived mental workload.

Table E.3. PSSUQ & TLX Results

CDS Intervention	Mean (SD)	Range
PSSUQ Score		
NASA TLX		

Participant Comments

In addition to the performance data, the following qualitative comments included:

- Positive features (add details);
- Areas for improvement (add details);
- Participant reactions to how CDS fits into workflow; and
- Whether participants would recommend the CDS intervention to their colleagues.

Discussion

[Begin the discussion with a narrative summary of the major findings]. One way to organize these is by the general model categories of efficiency, effectiveness, safety, and satisfaction. Summarize the qualitative findings. Next, include interpretations or implications of the findings. Briefly explain which areas are critical and what the results mean for workflow, productivity, and patient safety. Explain next steps.]