

NISTIR 7742

Customized Common Industry Format Template for Electronic Health Record Usability Testing

Robert M. Schumacher
User Centric. Inc,

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

U.S. Department of Commerce
Gary Locke, Secretary

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

NIST
**National Institute of
Standards and Technology**
U.S. Department of Commerce

NISTIR 7742

Customized Common Industry Format Template for Electronic Health Record Usability Testing

Robert M. Schumacher
User Centric. Inc,

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

November 2010



U.S. Department of Commerce
Gary Locke, Secretary

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

15-Nov-10

Customized Common Industry Format Template for Electronic Health Record Usability Testing

INTENDED AUDIENCE

This document provides a template for the modified version of *Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Common Industry Format (CIF) for usability test reports* (ISO/IEC 25062:2006(E)), the Common Industry Format (CIF) usability test report. This version of the CIF has been customized for use in usability testing of Electronic Health Records (EHRs) by usability administrator(s) and data logger(s). The template enables usability engineers to effectively communicate the results of EHR usability testing.

INTENDED PURPOSE

The intention of the CIF is to help vendors demonstrate evidence of usability in their final product in a format that allows both independent evaluation of a single product and comparison across multiple products. This document has been prepared as a template to guide EHR usability test administrators meet the usability processes approach put forth by the National Institute of Standards and Technology (NIST). The following customized CIF template is intended to assist EHR vendors, healthcare providers, and researchers in reporting the results of usability testing for each system tested.

USING THIS DOCUMENT

This document is not intended to be a tutorial on usability or usability testing.¹ To work with this document you should have expertise with common usability industry practices

¹ Excellent starting points for information are www.usability.gov and www.upassoc.org as well as Dumas, J., Redish, J. (1994) *A Practical Guide to Usability Testing*. Norwood, NJ: Ablex. Dana Chisnell & Jeffrey Rubin *Handbook of Usability Testing: How to Plan, Design, and Conduct*

and with standard ISO/IEC 25062:2006.² The ISO document is intended for the reporting of summative (i.e., quantitative) studies. The modifications here allow for the reporting of qualitative findings (i.e., formative) but strongly recommend and encourage the collection of quantitative measures of user performance.

Reports delivered using this template should conform to the major headings and content areas outlined below. Minor deviations from the outline and format are acceptable, but the reports should follow the template in all material aspects. This template includes the following sections:³

- Executive Summary
- Introduction
- Method
- Results
- Appendices

In addition to these sections, the modified CIF must also include a title page; a sample title page is included in the template example.

When completing the modified CIF template, it is highly recommended that EHR usability test administrator(s) and their data logger(s) refer to the instructions and guidance in order to properly complete this template.

The sample data provided in this template is an example or placeholder of the types of content that may be useful in completing the modified CIF template. Gray background text (bounded in square brackets) needs to be replaced by the EHRs' supplied information. It is important to note that this sample content is not to be taken literally or as a starting point.

Effective Tests (2nd ed.) Wiley, 2008.

Schumacher (2009). *Handbook of Global User Research*. Burlington, MA: Morgan Kaufman.

² This document can be purchased from:

http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=43046.

³ Each of these sections has a corresponding section in the ISO/IEC 25062.

EHR Usability Test Report of Creoks GoHealth 2.0

Report based on ISO/IEC 25062:2006 Common Industry Format for Usability Test Reports

Creoks GoHealth 2.0

Date of Usability Test: 11-25-18, 11-27-18
 Date of Report: 12-1-18
 Report Prepared By: Premiere Testing Lab
 Rhonda Bowlin
 918-200-3420
 s.rbowe5@yahoo.com
 2800 N 23rd St, Broken Arrow,
 OK. 74014

Table of Contents

1	EXECUTIVE SUMMARY	7
2	INTRODUCTION	10
3	METHOD	10
3.1	PARTICIPANTS	10
3.2	STUDY DESIGN	11
3.3	TASKS	12
3.4	PROCEDURE	13
3.5	TEST LOCATION	14
3.6	TEST ENVIRONMENT	15
3.7	TEST FORMS AND TOOLS	15
3.8	PARTICIPANT INSTRUCTIONS	16
3.9	USABILITY METRICS	17
4	RESULTS	19
4.1	DATA ANALYSIS AND REPORTING	19
4.2	DISCUSSION OF THE FINDINGS	20
5	APPENDICES	22
5.1	APPENDIX 1: SAMPLE RECRUITING SCREENER	23
5.2	Appendix 2: PARTICIPANT DEMOGRAPHICS	26
5.3	Appendix 3: NON-DISCLOSURE AGREEMENT AND INFORMED CONSENT FORM	27
5.4	Appendix 4: EXAMPLE MODERATOR'S GUIDE	29
5.5	Appendix 5: SYSTEM USABILITY SCALE QUESTIONNAIRE	36
5.6	APPENDIX 6: INCENTIVE RECEIPT AND ACKNOWLEDGMENT FORM	36

EXECUTIVE SUMMARY

A usability test of Creoks, GoHealth 2.0 was conducted on 11-25-18 – 11-27-18 in various locations by Premiere Testing Labs. The purpose of this test was to test and validate the usability of the current user interface, and provide evidence of usability in the EHR Under Test (EHRUT).

During the usability test, 8 healthcare providers and 2 other users matching the target demographic criteria served as participants and used the EHRUT in simulated, but representative tasks.

This study collected performance data on 58 tasks typically conducted on an EHR:

Creating and Editing a client record

Record lab, radiology and problems to a client record

Importing and reconciling CCDA data to a client record

Medication and Allergy management

During the 60 minute one-on-one usability test, each participant was greeted by the administrator and asked to review and sign an informed consent/release form (included in Appendix 3); they were instructed that they could withdraw at any time. Most participants had prior experience with the EHR.⁴ The administrator introduced the test, and instructed participants to complete a series of tasks (given one at a time) using the EHRUT. During the testing, the administrator timed the test and, along with the data logger(s) recorded user performance data on paper and electronically. The administrator did not give the participant assistance in how to complete the task.

⁴ If training or help materials were provided, describe the nature of it. The recommendation is that all participants be given the opportunity to complete training similar to what a real end user would receive prior to participating in the usability test.

Participant screens, head shots and audio were recorded for subsequent analysis.

The following types of data were collected for each participant:

- Number of tasks successfully completed within the allotted time without assistance
- Time to complete the tasks
- Number and types of errors
- Path deviations
- Participant's verbalizations
- Participant's satisfaction ratings of the system

All participant data was de-identified – no correspondence could be made from the identity of the participant to the data collected. Following the conclusion of the testing, participants were asked to complete a post-test questionnaire.

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*, were used to evaluate the usability of the EHRUT. Following is a summary of the performance and rating data collected on the EHRUT.

Task	Measure	N	Task Successes	Path Deviation	Task Time		Errors	Task Ratings 5=Easy
					Mean (SD)	Deviations (Observed / Optimal)		
1. Authorized user should be able to add new medication order to patients.		A1.1	Mean:100% (SD: 17.36548301)	0/0	Mean: 77.5 (SD: 16.92483382)	0/0	0	5
2. Authorized user should access the list of medications ordered for patient.		A1.2	100% (SD: 3.667424164)	0/0	Mean: 22.1 (SD:3.3)	0/0	0	5
3. Authorized user should be able cancel a medication ordered to patient.		A1.3	100% (SD: 7.523961722)	0/0	Mean: 30.5 (SD: 7.392563831)	0/0	0	5

4. Authorized user should be able to update the medication order.	A1.4	100% (SD: 6.9426219 83)	0/0	Mean: 38.5 (SD: 6.264982 043)	0/0	0	5
5. Authorized user should be able to see the changes made to the medication order.	A1.5	100% (SD: 4.1231056 26)	0/0	Mean: 27.8 (SD: 3.370459 909)	0/0	0	5
6. Authorized user should be able to add Lab order to patient.	A2.1	Mean:100 % (SD: 12.776541)	0/0	Mean: 37.9 (SD: 13.01883 251)	0/0	0	5
7. Authorized user must access the list of Lab test's ordered for patient.	A2.2	Mean:100 % (SD: 5.5145262 72)	0/0	Mean: 9.7 (SD: 5.311308 69)	0/0	0	5
8. Authorized user must be able to update the Lab order.	A2.3	Mean:100 % (SD: 14.422551 79)	0/0	Mean: 30.3 (SD: 13.48369 386)	0/0	0	5
9. Authorized user should be able to see the changes made to the Lab order.	A2.4	Mean:100 % (SD: 2.6907248 09)	0/0	Mean:7.9 (SD: 2.118962 01)	0/0	0	5
10. Authorized user should be able to add Radiology order to patient.	A3.1	Mean:100 % (SD: 9.2892410 88)	0/0	Mean: 29.7 (SD: 8.967162 316)	0/0	0	5
11. Authorized user must access the list of Radiology test's ordered for patient.	A3.2	Mean:100 % (SD:2.9)	0/0	Mean:7.7 (SD:2.1)	0/0	0	5
12. Authorized user must be able to update the Radiology order.	A3.3	Mean:100 % (SD: 11.679469 17)	0/0	Mean: 27.4 (SD: 11.47344 761)	0/0	0	5
13. Authorized user should be able to see the changes made to the Radiology order.	A3.4	Mean:100 % (SD:1.8)	0/0	Mean: 7.2 (SD:1.6)	0/0	0	5
14.							
15.							
16. Application should perform drug-drug and drug-allergy interaction checks and display alerts for a user while reviewing patient's current medications and allergies.	A4.1	Mean:100 % (SD: 21.670486 84)	0/0	Mean: 74.2 (SD: 21.05136 575)	0/0	0	5
17. Application should perform drug-drug and drug-allergy interaction checks and display alerts for a user while prescribing a new	A4.2	Mean:100 % (SD: 60.041735 48)	0/0	Mean: 193.3 (SD: 59.79138 734)	0/0	0	5

medication to patient.							
18. The user must create a new patient with following demographics data: Preferred language, date of birth, birth sex, race, ethnicity, sexual orientation, gender identity.	A5.1	Mean:100 % (SD: 48.81639069)	0/0	Mean:95.01 (SD: 48.52102637)	0/0	0	5
19. The user must be able to modify newly created patient's following information: Preferred language, date of birth, birth sex, race, ethnicity, sexual orientation, gender identity.	A5.2	Mean:100 % (SD: 31.10691242)	0/0	Mean:45.06 (SD: 30.68615323)	0/0	0	5
20. The user must be able to see the above updated demographics changes in application.	A5.3	Mean:100 % (SD: 2.051828453)	0/0	Mean:13.08 (SD: 2.675817632)	0/0	0	5
21. The user must be able to add a new problem to patient.	A6.1	Mean:100 % (SD: 37.01405139)	0/0	Mean:63.05 (SD: 36.75119046)	0/0	0	5
22. The user must be able to mark a patient problem as resolved.	A6.2	Mean:100 % (SD: 6.960603422)	0/0	Mean:32.0 (SD: 7.113367698)	0/0	0	5
23. The user must be able to access the active problem list of the patient.	A6.3	Mean:100 % (SD: 5.123475383)	0/0	Mean:11.03 (SD: 4.859012245)	0/0	0	5
24. The user must be able to access the historical problem list of the patient.	A6.4	Mean:100 % (SD: 2.645751311)	0/0	Mean:8.5 (SD: 1.284523258)	0/0	0	5
25.							
26. The user must be able to add a new medication to patient.	A7.1	Mean:100 % (SD: 21.09123041)	0/0	Mean:75.03 (SD: 19.85472236)	0/0	0	5
27. The user must be able to update medication of patient.	A7.2	Mean:100 % (SD: 12.45833055)	0/0	Mean:32.06 (SD: 11.49956521)	0/0	0	5
28. The user must be able to delete a patient's medication.	A7.3	Mean:100 % (SD: 6.327716808)	0/0	Mean:30.03 (SD: 6.783067153)	0/0	0	5
29. The user must be able to stop the medication of patient.	A7.4	Mean:100 % (SD: 6.340346994)	0/0	Mean:28.09 (SD: 6.503076195)	0/0	0	5

30. The user must be able to access the active medication list of the patient.	A7.5	Mean:100 % (SD: 2.968164416)	0/0	Mean:21.03 (SD: 2.60959767)	0/0	0	5
31. The user must be able to access the historical medication list of the patient.	A7.6	Mean:100 % (SD: 3.280243893)	0/0	Mean:25.01 (SD: 3.360059523)	0/0	0	5
32. The user must be able to add a new medication allergy to patient.	A8.1	Mean:100 % (SD: 24.45424299)	0/0	Mean:64.03 (SD: 24.61320784)	0/0	0	5
33. The user must be able to update medication allergy of patient.	A8.2	Mean:100 % (SD: 7.525955089)	0/0	Mean:33.06 (SD: 7.282856582)	0/0	0	5
34. The user must be able to delete a patient's medication allergy.	A8.3	Mean:100 % (SD: 7.71038261)	0/0	Mean:27.04 (SD: 6.606057826)	0/0	0	5
35. The user must be able to stop the medication allergy of patient.	A8.4	Mean:100 % (SD: 6.554387843)	0/0	Mean:27.03 (SD: 6.79779376)	0/0	0	5
36. The user must be able to access the active medication allergy list of the patient.	A8.5	Mean:100 % (SD: 7.258787778)	0/0	Mean:20.06 (SD: 6.770524352)	0/0	0	5
37. The user must be able to access the historical medication allergy list of the patient.	A8.6	Mean:100 % (SD: 4.85386444)	0/0	Mean:20.08 (SD: 3.893584467)	0/0	0	5
38. The user should select a patient who has a Clinical Decision Support criteria matching and view the alert.	A9.1	Mean:100 % (SD: 2.865309756)	0/0	Mean:13.04 (SD: 1.959591794)	0/0	0	5
39. The user should be able to access the list of full Clinical Decision Support rules for the system.	A9.2	Mean:100 % (SD: 2.578759392)	0/0	Mean:11.01 (SD: 2.071231518)	0/0	0	5
40. The user should be able to access the resource information using InfoButton for the Clinical Decision Support rules by executing the rule.	A9.3	Mean:100 % (SD: 4.029888336)	0/0	Mean:23.06 (SD: 3.382306905)	0/0	0	5
41. The user should be able to create a new Clinical Decision Support rule.	A9.4	Mean:100 % (SD: 38.67764212)	0/0	Mean:91.01 (SD: 39.43868659)	0/0	0	5
42. The user should select a patient who triggers the newly added Clinical Decision support rule.	A9.5	Mean:100 % (SD:)	0/0	Mean:19.04 (SD:)	0/0	0	5

		5.895761189)		6.248199741)			
43. User should be able to query the FDA database for and add the UDI number.	A14.1	Mean:100 % (SD: 12.71377206)	0/0	Mean:20.9 (SD: 12.2102416)	0/0	0	5
44. The user should be able to change the UDI status.	A14.2	Mean:100 % (SD: 3.12409987)	0/0	Mean:9.8 (SD: 2.039607805)	0/0	0	5
45. The user should be able to get description of the Implantable Device list.	A14.3	Mean:100 % (SD: 13.75645303)	0/0	Mean:23.1 (SD: 13.1791502)	0/0	0	5
46. The user should be able to add a new Implantable Device to the records of a patient.	A14.4	Mean:100 % (SD: 8.304215797)	0/0	Mean:24.9 (SD: 8.45517593)	0/0	0	5
47. The user must be able to mark the implantable device as inactive.	A14.5	Mean:100 % (SD: 2.491987159)	0/0	Mean:8 (SD: 3.033150178)	0/0	0	5
48. The user should be able to import an CCDA XML file for a patient.	B2.1	Mean:100 % (SD: 17.27570548)	0/0	Mean:60.2 (SD: 16.69610733)	0/0	0	5
49. User should be able to Incorporate the imported CCDA file.	B2.2	Mean:100 % (SD: 6.69701426)	0/0	Mean:14.8 (SD: 6.38435588)	0/0	0	5
50. The user should be able to reconcile allergies, medications and problems of their choice to the patient's record and save them.	B2.3	Mean:100 % (SD: 32.1498056)	0/0	Mean:43.3 (SD: 32.22436966)	0/0	0	5
51. The user should be able to view the newly added allergies, medications and problems in the patient's chart and generate CCDA which includes new data.	B2.4	Mean:100 % (SD: 7.359347797)	0/0	Mean:20.8 (SD: 6.660330322)	0/0	0	5
52. The user should create a new prescription for the patient.	B3.1	Mean:100 % (SD: 84.01220149)	0/0	Mean:148.1 (SD: 84.21454744)	0/0	0	5
53. The user should be able to change SIG(Dosage and duration etc) of added medication.	B3.2	Mean:100 % (SD: 9.50210503)	0/0	Mean:33 (SD: 8.402380615)	0/0	0	5
54. The user must view the newly added prescription.	B3.3	Mean:100 % (SD: 16.07638019)	0/0	Mean:60 (SD: 16.2111073)	0/0	0	5

55. The user should be able to cancel the newly added prescription.	B3.4	Mean:100 % (SD: 4.450842617)	0/0	Mean:51.5 (SD: 4.984977432)	0/0	0	5
56. The user should be able to Refill the newly added prescription.	B3.5	Mean:100 % (SD: 13.41044369)	0/0	Mean:76.3 (SD: 13.03111661)	0/0	0	5
57. The user should be able to access the Fill Status of a newly added prescription.	B3.6	Mean:100 % (SD: 6.902173571)	0/0	Mean:27.3 (SD: 6.496922348)	0/0	0	5
58. The user should be able to respond to a change request for a previously added prescription.	B3.7	Mean:100 % (SD: 17.29768771)	0/0	Mean:61.1 (SD: 17.72258446)	0/0	0	5

The results from the System Usability Scale scored the subjective satisfaction with the system based on performance with these tasks to be: [86].¹

In addition to the performance data, the following qualitative observations were made:

- Major findings
 - o None Reported
- Areas for improvement
 - o Move DOB, Gender to above the address section. PIN entry field is hard to see on some screens. Icons should be in alphabetical order,

¹ See Tullis, T. & Albert, W. (2008). Measuring the User Experience. Burlington, MA: Morgan Kaufman (p. 149). Broadly interpreted, scores under 60 represent systems with poor usability; scores over 80 would be considered above average.

INTRODUCTION

The EHRUT tested for this study was **Creoks GoHealth 2.0**. Designed to present medical information to healthcare

providers in **behavioral health settings**, the EHRUT consists of **electronic health record management**. The usability testing attempted to represent realistic exercises and conditions.

The purpose of this study was to test and validate the usability of the current user interface, and provide evidence of usability in the EHR

Under Test (EHRUT). To this end, measures of effectiveness, efficiency and user satisfaction, such as **adding lab and radiology orders**, were captured during the usability testing.

METHOD

PARTICIPANTS

A total of **10** participants were tested on the EHRUT(s). Participants in the test were **behavioral health providers**. Participants were recruited by **Premiere Testing Labs** and were compensated \$50 for their time. In addition, participants had no direct connection to the development of or organization producing the EHRUT(s). Participants were not from the testing or supplier organization. Participants were given the opportunity to have the same orientation and level of training as the actual end users would have received.

Recruited participants had a mix of backgrounds and demographic characteristics conforming to the recruitment screener. The following is a table of participants by characteristics, including demographics, professional experience, computing experience and user needs for assistive technology. Participant names were replaced with Participant IDs so that an individual's data cannot be tied back to individual identities.

ID	Gender	Age	Education	Occupation/Role	Professional Experience	Computer Experience	Product Experience	Assistive Technology Needs
ID01	Male	50-59	Bachelor's Degree	Crisis Unit Director	130	120	72	No
ID02	Female	30-39	Bachelor's Degree	Case Manager	36	36	36	No
ID03	Female	50-59	Trade/technical/vocational training	LPN	252	72	72	No
ID04	Male	20-29	Bachelor's Degree	Case Manager	60	156	60	No
ID05	Male	50-59	Bachelor's Degree	Case Manager	84	84	84	No
ID06	Female	30-39	Master's Degree	Licensed Clinician	120	144	24	No
ID07	Female	30-39	Trade/technical/vocational training	LPN	108	240	72	No
ID08	Male	30-39	Bachelor's Degree	Case Manager	72	168	36	No
ID09	Female	20-29	High school graduate, diploma or the equivalent (for example: GED)	Full-time mom	12	72	3	No
ID10	Female	20-29	Associate's Degree	Graphic Designer	72	180	1	No

11 participants (matching the

demographics in the section on Participants) were recruited and 10 participated in the usability

test. 1 participants failed to show for the study.

Participants were scheduled for 60 sessions with 30 in between each session for debrief by the administrator(s) and data logger(s), and to reset systems to proper test conditions. A spreadsheet was used to keep track of the participant schedule, and included each participant's demographic characteristics as provided by the recruiting firm.

STUDY DESIGN

Overall, the objective of this test was to uncover areas where the application performed well – that is, effectively, efficiently, and with satisfaction – and areas where the application failed to meet the needs of the participants. The data from this test may serve as a baseline for future tests with an updated version of the same EHR and/or comparison with other EHRs provided the same tasks are used. In short, this testing serves as both a means to record or benchmark current usability, but also to identify areas where improvements must be made.

During the usability test, participants interacted with 1 EHR. Participants used the system in various locations, and were provided with the same instructions. The system was evaluated for effectiveness, efficiency and satisfaction as defined by measures collected and analyzed for each participant:

- ▢ Number of tasks successfully completed within the allotted time without assistance
- ▢ Time to complete the tasks
- | Number and types of errors
- | Path deviations
- ▢ Participant's verbalizations (comments)
- | Participant's satisfaction ratings of the system

Additional information about the various measures can be found in Section 3.9 on UsabilityMetrics.

TASKS

A number of tasks were constructed that would be realistic and representative of the kinds of activities a user might do with this EHR, including:

1. Add and modify patient information,
2. Add and modify radiology/lab orders,
3. Import and reconcile CCDS data,
4. Add and manage medication allergies.,
5. Add and manage patient medications

Tasks were selected based on their frequency of use, criticality of function, and those that may be most troublesome for users.⁶ Tasks should always be constructed in light of the study objectives.

PROCEDURES

Upon arrival, participants were greeted; their identity was verified and matched with a name on the participant schedule. Participants were then assigned a participant ID.⁷ Each participant reviewed and signed an informed consent and release form (See Appendix 3). A representative from the test team witnessed the participant's signature.

To ensure that the test ran smoothly, two staff members participated in this test, the usability administrator and the data logger. The usability testing staff conducting the test was experienced usability practitioners with 16 years of experience.

The administrator moderated the session including administering instructions and tasks. The administrator also monitored task times, obtained post-task rating data, and took notes on participant comments. A second person served as the data logger and took notes on task success, path deviations, number and type of errors, and comments.

Participants were instructed to perform the tasks (see specific instructions below):

- As quickly as possible making as few errors and deviations as possible.

⁶ Constructing appropriate tasks is of critical importance to the validity of a usability test. These are the actual functions, but most tasks contain larger and more fleshed out context that aligns with the sample data sets available in the tested EHR. Please consult usability references for guidance on how to construct appropriate tasks.

⁷ All participant data must be de-identified and kept confidential.

- Without assistance; administrators were allowed to give immaterial guidance and clarification on tasks, but not instructions on use.
- Without using a think aloud technique.

For each task, the participants were given a written copy of the task. Task timing began once the administrator finished reading the question. The task time was stopped once the participant indicated they had successfully completed the task. Scoring is discussed below in Section 3.9.

Following the session, the administrator gave the participant the post-test questionnaire (e.g., the System Usability Scale, see Appendix 5), compensated them for their time, and thanked each individual for their participation.

Participants' demographic information, task success rate, time on task, errors, deviations, verbal responses, and post-test questionnaire were recorded into a spreadsheet.

Participants were thanked for their time and compensated. Participants signed a receipt and acknowledgement form (See Appendix 6) indicating that they had received the compensation.

TEST LOCATION

The test facility included a waiting area and a quiet testing room with a table, computer for the participant, and recording computer for the administrator. Only the participant and administrator were in the test room. All observers and the data logger worked from a separate room where they could see the participant's screen and face shot, and listen to

the audio of the session. To ensure that the environment was comfortable for users, noise levels were kept to a minimum with the ambient temperature within a normal range. All of the safety instruction and evacuation procedures were valid, in place, and visible to the participants.

TEST ENVIRONMENT

The EHRUT would be typically be used in a healthcare office or facility. In this instance, the testing was conducted in various remote locations. For testing, the computer used Microsoft Windows operating system. The participants used a mouse and keyboard when interacting with the EHRUT.

The Creoks GoHealth 2.0 used 20" monitor, 1366x768 resolution. The application was set up by the Premiere Testing Labs according to the vendor's documentation describing the system set-up and preparation. The application itself was running on a web-based platform using a SQL training database on a WAN connection. Technically, the system performance (i.e., response time) was representative to what actual users would experience in a field implementation. Additionally, participants were instructed not to change any of the default system settings (such as control of font size).

TEST FORMS AND TOOLS

During the usability test, various documents and instruments were used, including:

1. Informed Consent
2. Moderator's Guide

3. Post-test Questionnaire

The participant's interaction with the EHRUT was captured and recorded digitally with screen capture software running on the test machine. A **web** camera recorded each participant's facial expressions synced with the screen capture, and verbal comments were recorded with a microphone.⁸

PARTICIPANT INSTRUCTIONS

The administrator reads the following instructions aloud to the each participant (also see the full moderator's guide in Appendix [B4]):

Thank you for participating in this study. Your input is very important. Our session today will last about [60 minutes]. During that time you will use an instance of an electronic health record. I will ask you to complete a few tasks using this system and answer some questions. You should complete the tasks as quickly as possible making as few errors as possible. Please try to complete the tasks on your own following the instructions very closely. Please note that we are not testing you we are testing the system, therefore if you have difficulty all this means is that something needs to be improved in the system. I will be here in case you need specific help, but I am not able to instruct you or provide help in how to use the application.

Overall, we are interested in how easy (or how difficult) this system is to use, what in it would be useful to you, and how we could improve it. I did not have any involvement in its creation, so please be honest with your opinions. All of the information

⁸ There are a variety of tools that record screens and transmit those recordings across a local area network for remote observations.

that you provide will be kept confidential and your name will not be associated with your comments at any time. Should you feel it necessary you are able to withdraw at any time during the testing.

Following the procedural instructions, participants were shown the EHR and as their first task, were given time (10 minutes) to explore the system and make comments. Once this task was complete, the administrator gave the following instructions:

For each task, I will read the description to you and say "Begin." At that point, please perform the task and say "Done" once you believe you have successfully completed the task. I would like to request that you not talk aloud or verbalize while you are doing the tasks.⁹ I will ask you your impressions about the task once you are done.

Participants were then given 58 tasks to complete. Tasks are listed in the moderator's guide in Appendix [B4].

USABILITY METRICS

According to the *NIST Guide to the Processes Approach for Improving the Usability of Electronic Health Records*, EHRs should support a process that provides a high level of usability for all users. The goal is for users to interact with the system effectively, efficiently, and with an acceptable level of satisfaction. To this end, metrics for effectiveness, efficiency and user satisfaction were captured during the usability testing.

The goals of the test were to assess:

1. Effectiveness of Creoks GoHealth 2.0 by measuring participant success rates and errors
2. Efficiency of Creoks GoHealth 2.0 by measuring the average task time and path deviations

⁹ Participants should not use a think-aloud protocol during the testing. Excessive verbalization or attempts to converse with the moderator during task performance should be strongly discouraged. Participants will naturally provide commentary, but they should do so, ideally, after the testing. Some verbal commentary may be acceptable between tasks, but again should be minimized by the moderator.

3. Satisfaction with Creoks GoHealth 2.0 by measuring ease of use ratings

DATA SCORING

The following table (Table [x]) details how tasks were scored, errors evaluated, and the time data analyzed.¹⁰

Measures	Rationale and Scoring
<p>Effectiveness: Task Success</p>	<p>A task was counted as a “Success” if the participant was able to achieve the correct outcome, without assistance, within the time allotted on a per task basis.</p> <p>The total number of successes were calculated for each task and then divided by the total number of times that task was attempted. The results are provided as a percentage.</p> <p>Task times were recorded for successes. Observed task times divided by the optimal time for each task is a measure of optimal efficiency.</p> <p>Optimal task performance time, as benchmarked by expert performance under realistic conditions, is recorded when constructing tasks. Target task times used for task times in the Moderator’s Guide must be operationally defined by taking multiple measures of optimal performance and multiplying by some factor [e.g., 1.25] that allows some time buffer because the participants are presumably not trained to expert performance. Thus, if expert, optimal performance on a task was [x] seconds then allotted task time performance was [x * 1.25] seconds. This ratio should be aggregated across tasks and reported with mean and variance scores.</p>
<p>Effectiveness: Task Failures</p>	<p>If the participant abandoned the task, did not reach the correct answer or performed it incorrectly, or reached the end of the allotted time before successful completion, the task was counted as an “Failures.” No task times were taken for errors.</p> <p>The total number of errors was calculated for each task and then divided by the total number of times that task was attempted. Not all deviations would be counted as errors.¹¹ This should also be expressed as the mean number of failed tasks per participant.</p> <p>On a qualitative level, an enumeration of errors and error types should be collected.</p>
<p>Efficiency: Task Deviations</p>	<p>The participant’s path (i.e., steps) through the application was recorded. Deviations occur if the participant, for example, went to a wrong screen, clicked on an incorrect menu item, followed an incorrect link, or interacted incorrectly with an on-screen control. This path was compared to the optimal path. The number of steps in the observed path is divided by the number of optimal steps to provide a ratio of path deviation.</p>

¹⁰ An excellent resource is Tullis, T. & Albert, W. (2008). Measuring the User Experience. Burlington, MA: Morgan Kaufman. Also see www.measuringusability.com

¹¹ Errors have to be operationally defined by the test team prior to testing.

	It is strongly recommended that task deviations be reported. Optimal paths (i.e., procedural steps) should be recorded when constructing tasks.
Efficiency: Task Time	Each task was timed from when the administrator said "Begin" until the participant said, "Done." If he or she failed to say "Done," the time was stopped when the participant stopped performing the task. Only task times for tasks that were successfully completed were included in the average task time analysis. Average time per task was calculated for each task. Variance measures (standard deviation and standard error) were also calculated.
Satisfaction: Task Rating	<p>Participant's subjective impression of the ease of use of the application was measured by administering both a simple post-task question as well as a post-session questionnaire. After each task, the participant was asked to rate "Overall, this task was:" on a scale of 1 (Very Difficult) to 5 (Very Easy). These data are averaged across participants.¹²</p> <p>Common convention is that average ratings for systems judged easy to use should be 3.3 or above.</p> <p>To measure participants' confidence in and likeability of the Creoks GoHealth 2.0 overall, the testing team administered the System Usability Scale (SUS) post-test questionnaire. Questions included, "I think I would like to use this system frequently," "I thought the system was easy to use," and "I would imagine that most people would learn to use this system very quickly." See full System Usability Score questionnaire in Appendix 5.¹³</p>

Table [x]. Details of how observed data were scored.

RESULTS

DATA ANALYSIS AND REPORTING

The results of the usability test were calculated according to the methods specified in the Usability Metrics section above.

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

¹³ The SUS survey yields a single number that represents a composite measure of the overall perceived usability of the system. SUS scores have a range of 0 to 100 and the score is a relative benchmark that is used against other iterations of the system.

The usability testing results for the EHRUT are detailed below (see Table [x])¹⁴.

The results should be seen in light of the objectives and goals outlined in Section 3.2 Study Design. The data should yield actionable results that, if corrected, yield material, positive impact on user performance.

SEE TABLE ON PAGES 8-12

Task	Measure	Task Success	Path Deviation	Task Time		Errors	Task Ratings 5=Easy
	N			Mean (SD)	Mean (SD)		
	#	Mean (SD)	Deviations (Observed / Optimal)	Mean (SD)	Deviations (Observed / Optimal)	Mean (SD)	Mean (SD)
1.[Find item on patient summary screen]							
2.[Use patient chart to find lab results]							
3.[Check vital signs]							
Etc.							

The results from the SUS (System Usability Scale) scored the subjective satisfaction with the system based on performance with these tasks to be: [86]. Broadly interpreted, scores under 60 represent systems with poor usability; scores over 80 would be considered above average.¹⁵

DISCUSSION OF THE FINDINGS

¹⁴ Note that this table is an example. You will need to adapt it to report the actual data collected.

¹⁵ See Tullis, T. & Albert, W. (2008). Measuring the User Experience. Burlington, MA: Morgan Kaufman (p. 149).

EFFECTIVENESS

The EHRUT performed as expected. With overall effectiveness presenting very well in this test.

EFFICIENCY

Based on the observations throughout this test, the EHRUT exceeded expectations in the area of efficiency.

SATISFACTION

The satisfaction level of the test users was high based on the overall effectiveness and efficiency levels of the EHTUT.

MAJOR FINDINGS

There were no major findings reported.

AREAS FOR IMPROVEMENT

Move DOB, Gender and Age after the address fields. The PIN entry field needs to be more visible. The icons should be in alphabetical order on each tab.

