Project
User Testing Guide
Scope
Title

Author
author.email@author.author

00/00/0000

Prototype: prototype-a-link
Milestone: 1.2.0-beta

Version: v0.0.0-alpha
Release: v0.0

Prototype: prototype-b-link
Milestone: 1.2.0-alpha

Version: v0.0.0-alpha
Release: v1.2.0-alpha

Secondary: server-a-link
Commit: 12345

Depl. Env.: Localhost
Link: project-link

Depl. Server: Localhost

Main Server: Localhost
Private IP: localhost
Private Domain: localhost

Port: 1234
Public IP: localhost

Secondary Server: Localhost
Private IP: localhost
Private Domain: localhost

Port: 8448
Public IP: localhost
From: 8448

1 Introduction

This document [5] aims to describe the protocol and guidelines of the presented information. We perform a set of tests in the scope of v0.0.0-alpha, v0.0.0-alpha and v0.0.0-alpha versions from *Prototype Link A*, *Prototype Link B* and *Prototype Link C* prototypes, respectively. The repositories are part of the Some Name Link project using traditional devices (mouse and keyboard). The goal of the test is to compare each prototype, measuring the user performance, efficiency and efficacy metrics. The sessions will be recorded via video on a computer and using an interaction record, while triggering event tools. It is guaranteed the confidentiality of the recordings, which will be used only for academic purpose.

2 Description

This document describes our test plan for conducting our user tests during the development of the Project Name Link project and systems. The goals of the user testing phases include establishing a baseline of participant performance, establishing and validating participant performance measures, and identifying potential design concerns to be addressed in order to improve the efficiency, productivity, and end-user satisfaction within the development and introduction of *Scope* methods inside the participants environment room, between others, for some scope, or more precisely, for some domain.

The user test objectives are:

1. To determine design inconsistencies and issues within the UI and content areas;
   (a) **Navigation Errors**;
   (b) **Presentation Errors**;
   (c) **Control Usage Problems**;
2. Exercise the prototype under controlled test conditions with representative users;
3. Establish baseline user performance and user-satisfaction levels of the user interface for future usability evaluations;

Potential sources of error may include: (a) **Navigation Errors:** failure to locate functions, excessive keystrokes to complete a function, failure to follow recommended screen flow; (b) **Presentation Errors:** failure to locate and properly act upon desired information in screens, selection errors due to labeling ambiguities; and (c) **Control Usage Problems:** improper toolbar or entry field usage. Data will be used to access whether usability goals regarding an effective, efficient, and well-received user interface have been achieved.

3 Methodology

The hereby prototypes are both v0.0.0-alpha and v0.0.0-alpha versions of our *Prototype A Link* and *Prototype B Link* prototypes, respectively. The purpose of these prototypes is to involve an *Scope* tool (*Scope Tool Name*) for domain characteristics at a domain level[11].

3.1 Environments

This section describes the user environment over interaction. Supported by this guide, our research aims to conduct an investigation for the several environmental variables and improvements regarding the potentially enhancement that a *Scope* could take in the *Environment*.

Figure 1: Figure Title.

3.2 Participants

The participants' responsibilities will be attempting to complete a set of representative task scenarios (Section 7) presented to them in as efficient and timely a manner as possible, and to provide feedback regarding the usability and acceptability of an *Scope*. The participants will be directed to provide honest opinions regarding the user tests of the interacted systems, and to participate in post-session subjective questionnaires and debriefing.

3.3 Procedure

Participants will take part in the tests at our formed institution protocols (*e.g.*, Sample Link) with both v0.0.0-alpha and v0.0.0-alpha versions of our Prototype A Link and Prototype B Link prototypes, respectively. The interaction with the system will be used in a typical **Environment** environment. Note takers and data logger(s) will monitor the sessions for observation in the **Environment**, connected by screen recording feed. The test sessions will be recorded and further analyzed.

3.4 Briefing

A presentation of the *Scope* and its use and capabilities will be made. Participants will be presented to the available interactions and will be explained how to interact with the prototype, underlining the limitations. The facilitator will brief the participants on the *Scope* system and instruct the participant that they are evaluating the system, rather than the facilitator evaluating the participant. Participants will sign an informed consent that acknowledges: the participation is voluntary, that participation can cease at any time, and that the session will be videotaped but their privacy of identification will be granted. The facilitator will ask the participant if they have any question.

3.5 Training

The participants will receive and overview of the user test procedure, equipment and system. The facilitator will show how to interact with the system and what features are available. We choose this approach, as it provide users the most important concepts to understand and interact with our system. Also, it is of chief importance to give users information of what is and is not available analysis of our *Scope* and what it can do.

4 Roles

The roles involved in our user tests are as follows. An individual may play multiple roles, as well as the test may not require all roles.

4.1 Trainer

- Provide training overview prior to user testing phases;

4.2 Facilitator

- Provides overview of study to participants;
- Defines tasks and purpose of the user testing to participants;
- Assists in conduct of participant and observer debriefing sessions;
- Responds to participant's requests for assistance;

4.3 Data Logger

- Records participant's actions and comments;

4.4 Test Observers

- Silent observer;
- Assists the data logger in identifying problems, concerns, coding bugs and procedural errors;
- Serve as note takers;

4.5 Ethics

All persons involved with the usability test are required to adhere to the following ethical guidelines:

- The performance of any test participant must not be individually attributable;
- Individual participant's name should not be used in reference outside the testing session;
- A description of the participant's performance should not be reported to his or her superior;

5 Apparatus

For the material and apparatus, it is essential to capture the session apprehending the user interactions. In our case, we will record this interaction by using the Some Tool 00.0 (000.0.0) to obtain all interactions. We will pair this video tool with a user watch tool called Other Tool Link. All instruments will help us to capture where and when are users interacting. By looking at the test participant's reactions, we find a lot of information regarding the prototype design.

The material used in the test sessions for the user interface consists of:

- Item 1: it will allow the user to interact with the some aspects;
- Item 2: it will allow the user to interact with the some aspects;

5.1 Technical Details

To produce this traditional environment, and since we can simulate with a laptop, the mouse and keyboard interaction, we are using a Mobile Mouse 0000 together with the Computer Link (Model, 00-inch, Early 0000) with a standard integrated keyboard on the laptop.

5.2 Devices

Traditional interaction remains the most common way to interact with user interfaces in an environment. Unfortunately, most of this interaction is made by low profile equipment that makes users produce more errors and take more time interacting with those User Interfaces (UI).

5.3 User Interactions

The systems have several buttons that allows the user to interact or access to a set of user interface features. Each item of the following list represents each metaphoric icon.

The buttons are as follows:

- Button 1
- Button 2
- Button 3
- Button 4
- Button 5

5.4 Software

To track our user interactions across our system, we are using Some Tool Link. This tool is an analytic package allowing us to follow our users remotely. It also provides two critical pieces of functionality, among others, that can aid in remote user testing.

6 Evaluation

Introduction of *Scope* agents are significant factors which can naturally affect the performance of a medical workflow. While some prior studies have investigated the functionality of systems, the *Scope* acceptability has mostly been overlooked in the existing Health Informatics (HI) literature regarding a Human-Computer Interaction (HCI) research.

The following Table 1 is presenting three main *Research Questions* to have in mind during evaluation. The purpose of this questions is to facilitate systematic user studies regarding our novel *Scope* in an environment and support user stimulation for the introduction of *Scope* methods. The proposed issues, involve various aspects of workflow combined with, either need for satisfaction, nor division of attention.

Number	Research Questions
RQ1.	What is the impact of an *Scope* system for avoiding different types of errors on environment perception?
RQ2.	What are the design techniques for setting appropriate environment expectations of *Scope* systems?
RQ3.	What is the impact of expectation-setting intervention techniques on *Scope* satisfaction and acceptance?

Table 1: Research Evaluation Questions

The influence of *Scope* is an important variable for our empirical analysis. In fact, we expect that the trust of the user will increase when the user perceived that the *Scope* is giving the right inputs and that there will be a consequent increase of the user trust in our system.

List of associated *Research Questions* to respective set of *Hypotheses*:

1. **RQ1.** What is the impact of a *Scope* system for avoiding different types of errors on user perception ?
   (a) **H1.1.** A *Scope* system focused on *Some Aspect* will result in higher aspect of accuracy.
   (b) **H1.2.** A *Scope* system focused on *Some Other Aspect* will result in higher other aspects?
2. **RQ2.** What are the design techniques for setting appropriate user expectations of *Scope* systems ?
   (a) **H2.1.** A *Scope* system that directly is communicating it accuracy to users will reduce the lack between system accuracy and user perception.
   (b) **H2.2.** Providing users explanations (Some Link) will lead to higher perception of understanding how the *Scope* system works.
   (c) **H2.3.** A first user contact with the system will lead to higher perceived level of control over the *Scope* results.
3. **RQ3.** What is the impact of expectation-setting intervention techniques on satisfaction and acceptance of *Scope* ?
   (a) **H3.1.** In the mediation of an imperfect *Scope* system providing users the power of prior interventions will lead to higher acceptance and satisfaction in comparison to a lack of such interventions.

7 Tasks

During our user tests, we need to ask participants to provide a subjective assessment of their experience using our *Scope*. There are several widely used questionnaires giving us different prons-and-cons. However, in most cases, a single question instrument [16] is the right method for a quantitative usability testing. By taking less time and effort to answer, participants are pursuing to this phase after task, while it is minimally disruptive.

List of stand alone tasks for ALL phases:

Task 1.1.1: Describing Task;

Task 1.1.2: Describing Task;

Task 2.1.1: Describing Task;

Task 2.1.2: Describing Task;

Task 2.1.3: Describing Task;

8 Metrics

Our user test metrics refers to user performance measured against specific performance goals necessary to satisfy the test requirements. Scenario completion success rates, adherence to dialog scripts, error rates, and subjective evaluations will be used. Time-to-Completion (TtC) [10] of scenarios will also be collected. From the set of tasks (Section 7), each task corresponds to the set of *Phases*, *Scenarios* and *Activities* (Section 1), meaning that we first need to explain it relations.

8.1 Workload

To measure the workload, we used the NASA Task Load Index (NASA-TLX) [15] scale. The scale is a subjective workload assessment tool that will allow us to perform subjective workload assessments on our participants. For the purpose, we created a repository [7, 4] to cover this need of content.

By incorporating a multi-dimensional rating procedure, NASA-TLX derives an overall workload score based on a weighted average of ratings on six subscales:

- Mental Demand
- Physical Demand
- Temporal Demand
- Performance
- Effort
- Frustration

8.2 Usability

To measure the usability, we used the System Usability Scale (SUS) [14]. The SUS provides a "quick and dirty", reliable tool for measuring the usability. It consists of a 10 item questionnaire with ten response options for respondents; from *Strongly Agree* to *Strongly Disagree*. Originally created by John Brooke in 1986, it allows you to evaluate a wide variety of products and services, including hardware, software, mobile devices, websites and applications. For the purpose, we created a repository [6, 3] to cover this need of content.

When using SUS, participants are asked to score the following 10 items with one of ten responses that range from **Strongly Agree** to **Strongly Disagree**:

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 felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

8.3 Predictions

To measure system predictions with purpose of comparing participants acceptance, we applied our own computational method. The computational method is as follows, while we defined several variables to it, defined next to this information and further explained. Let the *Overall Accuracy* [2, 12] be $\emptyset$, a variable following the discrete uniform distribution as $\emptyset \in \mathbb{R}$. The accuracy is used by us to measure how accurate is the overall performance of our solution, considering both positive and negative classes without worrying about data imbalance. Let *Total Number of Correct Predictions* [2, 12] be τ, a variable following the discrete uniform distribution as $\tau \in \mathbb{R}$. Let *All Possible Predictions* [2, 12] be α, a variable following the discrete uniform distribution as $\alpha \in \mathbb{R}$. As follows, we report our computational method.

Computational method to measure the *Overall Accuracy* of our solution:

$$Overall\ Accuracy = \frac{Total\ Number\ of\ Correct\ Predictions}{All\ Possible\ Predictions}$$

8.4 Qualitative Evaluation

Qualitative and subjective evaluations regarding ease of use and satisfaction will be collected via *open-ended questions* [1, 13], and during debriefing at the conclusion of the session. The *open-ended questions* will utilize free-form responses and feedback, when possible. Whenever possible, it's best to ask *open-ended questions* so we can find out more than we can anticipate. We will test our questions by trying to answer them with short answers, and rewrite those to find out more about *how* and *what*. In some cases, we won't be able to accommodate free-form or write-in answers, though, and then it is necessary to limit the possibilities.

8.5 Scenario Completion

Each scenario, will require, or request, that the participant obtains, or inputs, specific data. This data would be used in course of a typical task. The scenario is completed when the participant indicates the scenario's goal has been obtained. Whether successfully or unsuccessfully. Or the scenario is completed when the participant requests and receives sufficient guidance as to warrant scoring the scenario as a critical error.

8.6 Time Completion

The time to complete (ToT) [8, 9] each scenario, not including qualitative and subjective evaluation durations, will be recorded. From this measure, it will be also possible to collect more specific metrics, such as the percentage of time that participants follow an optimal path or the number of times participants need to backtrack.

8.7 Critical Errors

Critical Errors are deviations at completion from the targets of the scenario. Obtaining or otherwise reporting of the wrong data value due to participant workflow is a Critical Error. Participants may or may not be aware that the task goal is incorrect or incomplete.

An example of a Critical Error, could be a situation where the participant is not able to open a patient. From this error, we can not even proceed to the next tasks and complete the user test. Despite of the independent completion of the scenario is the goal, we need to guarantee the execution of the test, however, when this errors occur, the facilitator must act.

Critical Errors can also be assigned when the participant initiates, or attempts to initiate, an action that will result in the goal state becoming unobtainable. In general, Critical Errors are unresolved errors preventing completion of the task or errors that produce an incorrect outcome.

8.8 Non-Critical Errors

Non-Critical Errors, are errors that are recovered from and by the participant. Or, if not detected, do not result in processing problems or unexpected results. Although Non-Critical Errors can be undetected by the participant, when they are detected they are generally frustrating to the participant.

These errors may be procedural, in which the participant does not complete a scenario in the most optimal means (*e.g.*, excessive steps and keystrokes). These errors may also be errors of confusion (*e.g.*, initially selecting the wrong function, using a UI control incorrectly such as attempting to edit an un-editable field).

Non-Critical Errors can always be recovered from during the process of completing the scenario. Exploratory behavior, such as opening the wrong menu while searching for a function, will be coded as a non-critical error.

9 Goals

The next sections will describe the goals for *Prototype A Link*, *Prototype B Link* and *Prototype C Link* prototype expectations. We will try to assess performance-related metrics such as time and correctness of participants completing *tasks* for our expectations. Our expectations are based of the Results Link obtained at the lab as pilot tests.

9.1 Completion Rate

Completion Rate is the percentage of test participants who successfully complete the task without critical errors. A critical error is defined as an error that results in an incorrect or incomplete outcome. In other words, the completion rate represents the percentage of participants who, when they are finished with the specified task, have an "output" that is correct.

*A **Completion Rate** of **90%** is the goal for each task in this usability test.*

Note: If a participant requires assistance in order to achieve a correct output then the task will be scored as a critical error and the overall completion rate for the task will be affected.

9.2 Error-Free Rate

Error-Free Rate is the percentage of test participants who complete the task without any errors (critical or non-critical errors). A non-critical error is an error that would not have an impact on the final output of the task but would result in the task being completed less efficiently.

*An **Error-Free Rate** of **80%** is the goal for each task in this tests.*

9.3 Time on Task (ToT)

The time to complete a scenario is referred to as "Time on Task" (ToT). It is measured from the time the participant begins the scenario to the time which the participant signals completion.

9.4 Subjective Measures

Subjective opinions about specific tasks, time to perform each task, features, and functionality will be surveyed. At the end of the test, participants will rate their satisfaction with the overall system. Combined with the interview/debriefing session, these data are used to assess attitudes of the participants.

Measuring subjective outcomes based on participants' experiential goals can pose challenges (Section 10) from which an *open-ended* flexible approach is catered to personally meaningful goals. On the other hand advocates of formalized User-Centred Design (UCD) goal exploration condemn such informal interviewing as ineffective and we should take it into consideration.

9.5 Case Studies

The functionality of the prototype will be best demonstrated by a series of case studies. By describing the expected workflow and capabilities of the research study at the **Environment** specific environment and changes of the workflow by using our system prototype. The study implies the evaluation of domain *Scope* features on several domain aspects. The primary goal of this case studies analysis is to generate a receiver operating characteristic to evaluate the performance and validation of our *Scope*. Let us consider a list of hypothetical use cases for the research investigation that evaluates the interaction and usability performance of the *Scope*. Therefore, the following list will show the preliminary case studies.

List of case studies to analyse our solution prototype:

- Case 1;
- Case 2;
- Case 3;
- Case 4;

10 Challenges

In addition to the challenges already highlighted in the presented document, we must accomplish the participation test issues. The difference in knowledge and expertise levels between the participants will inhibit communication and participation of participants in different ways. Moreover, the factor that posed challenges to participants are involving them to a nominal adoption of consequences in the perceptions and practice, related ethical and self conflicts in presence of results. Challenges to the test and for practitioners to improve both study and research.

11 Results

A Test Report Link will be provided at the end of this tests. It will consist of a report and/or a presentation of the results; evaluation of the metrics against the pre-approved goals, subjective evaluations, and specific issues of the system, as well as, recommendations for resolution. The recommendations will be categorically sized by development to aid in implementation strategy. The results will be translated to a spreadsheet link (view only). Also, more related information can be found at Test Link: Test Title.

12 Acknowledgements

A special thanks for the support provided by Person 1. We would like to thank Person 2 and Person 3 from the some place for the generous support and user expertise. Also, an immense thank for Person 4. My appreciation goes also to Person 5 and Person 6 for help and above all for the good companionship. Thanks to Person 7 and Person 8 for the technical inputs and network. Last but not least, thank to my advisors Person 9 and Person 10. This work was partially supported by funds through Some Name with reference AAA/BBB/00000/0000 and Other Name through the CCC/DDD/EEE/00000/0000 project, FF00/0000-GGG-HH grant. We would like to convey Some Other Name for the collaboration.

References

[1] Julia Abelson, Kathy Li, Geoff Wilson, Kristin Shields, Colleen Schneider, and Sarah Boesveld. Supporting quality public and patient engagement in health system organizations: development and usability testing of the p ublic and p atient e ngagement e valuation t ool. *Health Expectations*, 19(4):817–827, 2016.

[2] Nabeela Ashraf, Waqar Ahmad, and Rehan Ashraf. A comparative study of data mining algorithms for high detection rate in intrusion detection system. *Annals of Emerging Technologies in Computing (AETiC)*, 2(1), 2018.

[3] Francisco Maria Calisto. Mimbcd-ui/nasa-tlx: v1.0.0-alpha, September 2018.

[4] Francisco Maria Calisto. Mimbcd-ui/sus: v1.0.0-alpha, September 2018.

[5] Francisco Maria Calisto. mida-project/testing-guide-breast: v1.0.1-alpha, April 2019.

[6] Francisco Maria Calisto and Jacinto C. Nascimento. Nasa-tlx survey, 2018.

[7] Francisco Maria Calisto and Jacinto C. Nascimento. Sus survey, 2018.

[8] Cheryl Delgado and Linda Wolf. Time on task: Perceived and measured time in online courses for students and faculty. *Journal of Nursing Education and Practice*, 7(5), 2017.

[9] Jue Huang, Christine Ulke, Christian Sander, Philippe Jawinski, Janek Spada, Ulrich Hegerl, and Tilman Hensch. Impact of brain arousal and time-on-task on autonomic nervous system activity in the wake-sleep transition. *BMC neuroscience*, 19(1):18, 2018.

[10] John PA Ioannidis. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *Jama*, 279(4):281–286, 1998.

[11] Y Kobashi, Y Munetomo, A Baba, S Yamazoe, and T Mogami. Evaluation of the ossification of the cervical poste-rior longitudinal ligament utilizing x-ray, ct and mr imaging. *Orthop Res Traumatol Open J*, 2(1):35–39, 2017.

[12] Tong Li, Michael Luke Marinovich, and Nehmat Houssami. Digital breast tomosynthesis (3d mammography) for breast cancer screening and for assessment of screen-recalled findings: review of the evidence. *Expert review of anticancer therapy*, 18(8):785–791, 2018.

[13] Rajan Merchant, Rubina Inamdar, Kelly Henderson, Meredith Barrett, Jason G Su, Jesika Riley, David Van Sickle, and David Stempel. Digital health intervention for asthma: patient-reported value and usability. *JMIR mHealth and uHealth*, 6(6):e133, 2018.

[14] Konstantina Orfanou, Nikolaos Tselios, and Christos Katsanos. Perceived usability evaluation of learning management systems: Empirical evaluation of the system usability scale. *The International Review of Research in Open and Distributed Learning*, 16(2), 2015.

[15] Anjana Ramkumar, Pieter Jan Stappers, Wiro J Niessen, Sonja Adebahr, Tanja Schimek-Jasch, Ursula Nestle, and Yu Song. Using goms and nasa-tlx to evaluate human–computer interaction process in interactive segmentation. *International Journal of Human–Computer Interaction*, 33(2):123–134, 2017.

[16] J Sauro. 10 things to know about the single ease question (seq). *Measuring U, 2012*, 2012.