For   STC DC Usability SIG

Usability Tests and Heuristic Reviews
Planning and Estimation Worksheets

Scott McDaniel
Senior Interaction Designer

26 February, 2003
Eval_Design_Tool_Handout.doc

Cognetics Corporation

E-mail: info@cognetics.com   ●   Web: www.cognetics.com

51 Everett Drive #103B
Princeton Junction NJ 08550
609.799.5005

1320 Fenwick Lane, Suite 209
Silver Spring, MD 20910
301.587.7549

Usability Test Planning Tool

Designing and planning usability tests offers something of a chicken and egg problem. It can be difficult to design a test unless you know the time and resources available to you, yet on the other hand you need to know the best test design so that you can provide an estimate – or justification – of time and resources.

This tool allows you to attack the problem from both directions. If you need to work out the optimal test design first, complete the Usability Test Design section and then the Project Planning and Estimation section. If, on the other hand, you are working under schedule or budget constraints, work through the Project Planning and Estimation section first to get an idea of how to spend the available resources, and then complete the Usability Test Design section.

Usability Test Design

Every usability test requires you to address certain issues that shape the test's design. Answer these questions before creating your testing materials (scripts, prototypes, etc.) to make certain that you have covered all of the issues and discussed their impact on the test design

Goals

The stage of LUCID in which you conduct a usability test affects the test's goals. Note the stage of LUCID in which this test will take place, and then state your specific goals and the design implications for those goals.

Stage	Typical Goals	Design Implications
Envision	• Benchmark previous release • Benchmark competition	• Summative test • Measure task times, error rates, and subjective impression
Discovery	• Discover users' pain points • Evaluate concept sketches • Discover users' mental model	• Summative or formative test • Ask users to explain how they think sketches or paper prototypes work, and what they think each element means • Ask users to perform a typical task with a prototype or existing version, letting them choose the task
Design Foundation	• Evaluate UI concept • Evaluate UI navigation • Evaluate screen layout • Evaluate terminology • Evaluate key workflows • Evaluate key screens	• Formative test • Ask users to navigate to screens within the prototype • Ask users to explain the elements of a prototype screen • Ask users to perform key tasks
Design Detail	• Evaluate specific workflows	• Formative test

	Evaluate specific screens Evaluate user assistance strategy	Ask users to perform specific tasks of interest
Build	• Fine-tune specific workflows • Fine-tune specific screens • Compare usability of new version with benchmarks	• Formative or summative test • Ask users to perform tasks that require live screens and/or database • Measure task times, error rates, and subjective impression
Release	• Compare usability of new versions with benchmarks • Benchmark current release	• Summative test • Measure task times, error rates, and subjective impression

For formative tests, list the design decisions you intend to make based on the results of the test (there is space for 5, but list more or less as needed). Also list potential tasks that could provide results allowing you to make those decisions. If possible, avoid general design questions like "Is the navigation structure working?" Instead, focus on specific questions that can be clearly answered. For summative list the screens or tasks that you wish to evaluate and any usability goals.

Design Decision	Potential Tasks & Implications
Formative Example: Should we keep secondary navigation on the screen all the time or have user mouse-over the first level choices to see a menu of secondary navigation?	• Find the Account History screen • Provide some users with one strategy and some users with the other • After the test, ask user how the navigation works
Summative Example: Did we decrease the average time required to complete a sale by at least 20% as compared to the previous version?	• Complete a sale. • Test identical tasks for both versions. • Have users complete several sales to see if there are learning effects.
1)	
2)	
3)	
4)	
5)	

Participants

The number of participants to test depends on what you want to get out of the test, the budget, and the time available. Redish and Dumas recommend no less than 3 users per sub-group of the population you want to test and 2 to 4 sub-groups. If you are testing a particularly crucial product, or if you want to use statistical analysis, you may want more. You may also want more if you are conducting a comprehensive test designed to measure general usability levels or if you allow users to choose the specifics of their tasks, as Spool does.

Each test has its own set of constraints and needs, but you can use this grid as a starting point in choosing the number of users you want to test:

	Formal	**Informal**	**Quick'n'Dirty**
Formative (design tool)	3 users per sub-group	2 users per sub-group	2-3 users
Summative (usability measurement)	5 users per sub-group	3 users per sub-group	N/A

Prototype & Test Equipment

A significant part of any usability test effort is the physical production of the prototype to be tested. You can use several kinds of prototype for the usability test, depending on what you need to test. The more functional the prototype, the better the results of the usability test. The extra cost and time required to build functional prototypes, however, often makes paper prototypes or non-functional prototypes the better choice. The table below shows the strengths of the various options.

For Testing	**Paper Prototype**	**Online, Non-Functional Prototype**	**Online, Functional Prototype or Released Product**
Page Structure	☑	☑	☑
Terminology	☑	☑	☑
Navigation among pages	☑	☑	☑
Basic workflow	☑	☑	☑
Navigation within pages		☑	☑
Screens/tasks that provide immediate feedback		☑	☑
Screens/tasks that require computation or database			☑
Ability to manipulate objects in the UI			☑
Detailed workflow			☑
Error handling			☑

Analysis

Decide how you will conduct the analysis. Your needs may differ, but you can use the table below as a starting point to guide your analysis strategy.

	Formal	Informal	Quick'n'Dirty
Formative (design tool)	Descriptive statistics and observation detail	Observation details	Observation details
Summative (usability measurement)	Inferential statistics (showing statistical significance)	Descriptive Statistics	N/A

The test's goals influence the data you collect. If you need to show statistics, you will need to design the test to produce objective measurements like time to completion, error rate, indicators of confusion, number of swear words, etc. If you are not concerned with statistics, you still may want a way to categorize and rate the severity of observation details.

Reporting & Presentation

It is critical to plan not only the presentation of results to the larger team, but also the process for acting on those results. Here are typical reporting and presentation strategies, along with their strengths and weaknesses.

Strategy	Strengths	Weaknesses
Report	• Allows thorough documentation of process and results • Allows others to replicate test or compare results of later tests • Is a clear, solid deliverable	• May not get read • Often are not acted upon
Presentation	• Ensures that attendees know of results • Allows for questions and discussion	• May be difficult to get everyone together
Working Meeting	• Allows for questions and discussion • Has the best chance of incorporating results into new design iterations	• May have nothing specific to show for the test except design changes • May be difficult to get everyone together

Project Planning and Estimation

These planning methods and estimation "rules of thumb" provide you with answers to a Project Manager's questions of "How much time?" and "How much money?" for usability tests. By following the steps below, you will have solid reasoning for your figures. As you conduct more usability tests, you will be able to supplement these numbers with your own experience.

There are two estimation methods available. You can:

1. Begin with the project plan, sometimes called a Work Breakdown Structure by managers, estimating the number of hours you think are needed for each activity for each role. The total hours required for the project is the sum of each of these individual estimates. This is a bottom-up approach.

2. Begin with the estimation rule of thumb, that predicts that a usability test takes 2 hours per participant per screen tested. You can then distribute these hours among the activities in the project plan. This is a top-down approach.

Project Plan

Determine the roles and activities required for the usability test and distribute the hours accordingly.

Roles

Most usability tests require at least two people, and often a third or fourth person is helpful. A single person can conduct quick, informal tests if need be. The key roles are:

Facilitator: Interacts with users during the test, as appropriate.
Observer: Observes the user and takes notes during the test.

Additional roles that are often helpful include:

Recruiter: Contacts and screens potential users.
A/V Technician: Handles video and audio equipment before and after the test. May assist in producing highlights tapes.

Work Breakdown Structure

A work breakdown structure is a tool that managers use to estimate, organize, and track work on projects. Estimate the time needed for each activity to determine the total person hours needed for a test. If, on the other hand, you have *x* hours in which to complete a project, distribute those hours among the activities shown below to determine how much time to spend on each task.

Activity	Facilitator	Observer	Recruiter	A/V Tech
Create Evaluation Plan				
Determine product availability				
Review and learn product				
Establish goals for the usability evaluation				
Coordinate test platform				
Design evaluation tasks				
Create evaluation plan				
Recruit Participants				
Identify participants to recruit				
Create Recruitment Screeners				
Obtain contact info for participants				
Recruit participants				
Prepare Evaluation Materials				
Set up product to be tested (screens, wireframes, etc.)				
Create/assemble scripts, forms, materials				
Facilities logistics (find lab, etc.)				
Prepare lab equipment				
Conduct dry run				
Conduct Evaluation				
Sessions: x participants in y days				
Analyze evaluation data				
Present Findings				
Prepare draft report/presentation				
Review draft report/presentation				
Prepare final report/presentation				
Present final report/presentation				
Facilitate decision-making				

Estimation Rule of Thumb

Based on our experience conducting usability tests, we have arrived at a way to compute a quick starting estimate of the hours needed to conduct a usability test. Use this number as a starting point for your planning – not as the last word on the effort required for the test.

The estimated hours for a formal usability test is 10 hours per hour of testing (2 for preparation and analysis and 1 for the hour spent testing). Thus, if p is the number of participants and you want each test session to be h hours:

   Initial hours estimate = 10 x p x h

Informal tests can use a similar equation, but typically use 7 or 8 hours per hour of testing. For guidance on selecting the right number of participants, see the Usability Test Design section. Factors that may prompt you to adjust these figures include:

- **Availability of a Testable Prototype.** Is a prototype ready to go, or does one need to be built? Paper prototypes take less time to create, but they still take longer than testing an existing prototype.
- **Difficulty of Recruiting Users.** Is there a standard procedure to contact and recruit users or must you start from scratch? In particular, is there internal resistance to contacting users?
- **Specialization of the Application.** Will the Facilitator and Observer need to spend time learning the system themselves?
- **Detail of Analysis.** Do you plan to record and statistically analyze in-depth usability metrics or make simple observations and summarize them in a table?
- **Format of the Report.** Do you plan to produce a complete written report with a full review cycle, conduct only a quick results/decisions meeting, or something in between?

Estimate Cost and Schedule for the Test

Schedule

Using the activities above, and assuming 6 useful hours per workday, determine the calendar schedule for the usability test. (Note: 6 hours per day is not to suggest that we're all goofing off, but rather to acknowledge the fact that we get called into staff meetings, attend training, answer e-mail, and perform other useful – but not project related – activities.)

Cost

Estimate the labor cost of the usability test by multiplying the hourly rates of the people involved by the number of hours they are to spend on the test.

Estimate additional costs using available information. Price equipment and lab rental in your area and estimate incentives based on your area and participant income. Typical costs include usability lab rental, A/V equipment rental, and incentives for participants.

Heuristic Review Planning Tool

Heuristic Review Process

Every heuristic review requires you to address certain issues that shape the test's design. Answer these questions before beginning the review to make certain that you have covered all of the issues and discussed their impact on the review and its presentation.

Goals

The stage of LUCID in which you conduct a heuristic review affects the review's goals. Note the stage of LUCID in which this review will take place, and then state your specific goals.

Stage	Typical Goals
Envision	• Inventory usability issues in a previous release
Discovery	• Prioritize usability issues
Design Foundation	• Provide feedback on an emerging design • Evaluate navigation, architecture, and high-level concepts in the UI • Identify areas to investigate in a usability test • Substitute for an early usability test
Design Detail	• Evaluate specific workflows • Evaluate specific screens • Evaluate user assistance strategy
Build	• Fine-tune specific workflows • Fine-tune specific screens
Release	• Prepare for upcoming releases

List the review's specific objectives below (there is space for 3, but list more or less as needed). What will be done with the report or presentation?

Purpose of the Review	Process and Reporting Implications
Example: Discover and prioritize usability issues in a recent software version so management can better allocate resources for an upcoming version.	• Concentrate on high-level recommendations • Provide a high-level slide show outlining issues, examples, and recommendations • Provide complete observations in a separate, detailed report
Example: Guide the design of a still in-progress user interface.	• Hold a meeting with the UI design team to discuss each observation and decide upon a solution.
1)	
2)	
3)	

Heuristics

By definition, a heuristic review is a comparison of a user interface compliance with a given set of heuristics. The following list of heuristic sets will get you started. They are necessarily general – you should look for more specific heuristics in your particular area.

Source	Reference
Cognetics Corporation	http://www.cognetics.com/services/heuristic_guidelines.html
Jakob Nielsen	http://www.useit.com/papers/heuristic/heuristic_list.html
National Cancer Institute	http://www.usability.gov/guidelines/index.html
STC Usability SIG	http://www.stcsig.org/usability/resources/toolkit/toolkit.html
Scott McDaniel (for search user interfaces)	McDaniel, S. and McDaniel, M. (2002) The Big Dig: Mining Nuggets of Value. *User Experience*, 1:2, 20-29.

Prototype & Test Equipment

Most heuristic reviews are performed on complete, functional user interfaces. They can also be performed on prototypes. The more functional the prototype, the more complete the review can be. The table below shows the strengths of the various options.

For Reviewing	Paper Prototype	Online, Non-Functional Prototype	Online, Functional Prototype or Released Product
Page Structure	☑	☑	☑
Terminology	☑	☑	☑
Navigation among pages	☑	☑	☑
Basic workflow	☑	☑	☑
Navigation within pages		☑	☑
Screens/tasks that provide immediate feedback		☑	☑
Screens/tasks that require computation or database			☑
Ability to manipulate objects in the UI			☑
Detailed workflow			☑
Error handling			☑

Reporting & Presentation

It is critical to plan not only the presentation of results to the larger team, but also the process for acting on those results. Here are typical reporting and presentation strategies, along with their strengths and weaknesses.

Strategy	Strengths	Weaknesses
Report	• Allows thorough documentation of process and results • Is a clear, solid deliverable	• May not get read • Often are not acted upon
Presentation	• Ensures that attendees know of results • Allows for questions and discussion	• May be difficult to get everyone together
Working Meeting	• Allows for questions and discussion • Has the best chance of incorporating results into new design iterations	• May have nothing specific to show for the review except design changes • May be difficult to get everyone together

Project Planning and Estimation

These planning methods and estimation "rules of thumb" provide you with answers to a Project Manager's questions of "How much time?" and "How much money?" for a heuristic review. By following the steps below, you will have solid reasoning for your figures. As you conduct more reviews, you will be able to supplement these numbers with your own experience.

There are two estimation methods available. You can:

1. Begin with the project plan, sometimes called a Work Breakdown Structure by managers, estimating the number of hours you think are needed for each activity for each role. The total hours required for the project is the sum of each of these individual estimates. This is a bottom-up approach.

2. Begin with the estimation rule of thumb, that predicts that a heuristic review takes 6 hours per participant per screen tested. You can then distribute these hours among the activities in the project plan. This is a top-down approach.

Project Plan

Determine the roles and activities required for the heuristic review and distribute the hours accordingly.

Roles

You can complete a heuristic review alone, but generally having two or more reviewers produces better results. The key roles are:

Primary Reviewer: Reviews the UI, leads the effort, and writes report/presentation.
Secondary Reviewers: Reviews the UI and coordinates with primary reviewer.

Work Breakdown Structure

A work breakdown structure is a tool that managers use to estimate, organize, and track work on projects. Estimate the time needed for each activity to determine the total person hours needed for a review. If, on the other hand, you have *x* hours in which to complete a project, distribute those hours among the activities shown below to determine how much time to spend on each task.

Activity	Primary Reviewer	Secondary Reviewer 1	Secondary Reviewer 2	Secondary Reviewer 3
Review Kickoff				
Establish goals of review				
Establish usability goals of product				
Establish product schedule				
Introduce reviewer(s) to product				
Q & A about product				
Initial Review				
Conduct independent reviews of product				
Merge observations and questions				
Submit questions about product				
Second Review				
Meet with project team to answer questions				
Complete product review				
Heuristic Report				
Write draft report				
Review/revise report with secondary reviewers				
Review/revise report with project team				
Finish final report				
Results Presentation				
Present results and recommendations to project team				
Facilitate decision-making				
Consult with project team during implementation of recommendations				

Estimation Rule of Thumb

Based on our experience conducting heuristic reviews, we have arrived at a way to compute a quick starting estimate of the hours needed to conduct a review. Use this number as a starting point for your planning – not as the last word on the effort required for the review.

The estimated hours for a heuristic review is 6 hours per screen reviewed. Thus, if s is the number of screens to review:

Initial hours estimate = 6 x s

This estimate provides a starting point that you will have to adjust based on your specific situation. Factors that may prompt you to adjust these figures include:

- **Availability of a Reviewable Prototype.** Is a prototype ready to go, or does one need to be built? Paper prototypes take less time to create, but reviews of functional prototypes provide better results.
- **Specialization of the Field.** Will the reviewers need to spend time understanding basic concepts of the system's domain in order to understand it?
- **Specialization of the Application.** Will the reviewers need to spend time learning the system before they can review it effectively?
- **Format of the Report.** Do you plan to produce a complete written report with a full review cycle, conduct only a quick results/decisions meeting, or something in between?

Estimate Cost and Schedule for the Test

Schedule

Using the activities above, and assuming 6 useful hours per workday, determine the calendar schedule for the heuristic review. (Note: 6 hours per day is not to suggest that we're all goofing off, but rather to acknowledge the fact that we get called into staff meetings, attend training, answer e-mail, and perform other useful – but not project related – activities.)

Cost

Estimate the labor cost of the heuristic review by multiplying the hourly rates of the people involved by the number of hours they are to spend on the test. Heuristic reviews do not typically result in costs other than labor costs unless travel is involved.