Appendix A – Running Usability Evaluations

Contents
A.1. Setting Target Goals
A.2. Setting Up the Tasks
A.3. Recruiting Participants
A.4. Running the Test
A.5. Analyzing the Data
A.6. Presenting the Results

Running a usability test with real customers is essential to good design. You may know a lot about your customers, but it is hard to predict how people will actually react and interact with a Web site. Usability tests are also effective in ending those endless opinion wars, where one member of the design team says, “I think that people like X” while another says, “No, they really like Y.” The best way to sort it all out is to recruit some participants, run a quick test, and see what they say and do.

This appendix lays out the steps for running both formal and informal usability tests, from setting up the test to running the test to analyzing and presenting the results.

This appendix assumes that you want to run a usability test where both you and the participant are in the same place. You should also be aware that an alternative approach is remote usability testing, recruiting and testing many participants online without you actually having to be there. We discuss how to do this in Appendix E.

A.1. Setting Target Goals

What do you want to learn from the test? The first thing you have to nail down is what you want to get out of the test. Are people having problems with a specific part of the Web site? Do you want to see how well a proposed design works? Or do you just want to get feedback in general about the existing Web site?

How will you get this information? After working out what you want to learn, think about how you will get this information. If people are having problems with portions of the Web site, then the straightforward thing to do is to test tasks that rely on that part and see what problems people have. If you want to test a new design, then it is useful to compare it to the old design or to a competitor’s Web site. This approach is also useful for getting feedback in general about an existing Web site.

Process data versus bottom-line data. There are two kinds of data that you can get from a usability test: process data and bottom-line data. Process data is informal,

1 As an aside, in psychology and other fields, the term “subjects” is used instead of “participants”, but we have always felt that the term subjects has a slightly sinister tone to it.
qualitative observations of what people are thinking and doing, an overall feeling of what
works and what does not on a Web site. The key thing to look for here are critical
incidents, places on your Web site where participants are confused, frustrated, or even
swear. Critical incidents also include things where people were pleasantly surprised or
say something positive about the site.

In contrast, bottom-line data is formal, quantitative measurements of what happened,
such as the time it takes to complete a task, the number of errors that occurred, or the
time it takes to learn a task.

In general, you should focus on getting process data first, as it gives a good overview of
where the problems in a Web site are and because it is easier to get. Process data can also
be obtained from low-fidelity paper prototypes, making it a handy technique for the early
stages of design.

It takes more work to get bottom-line data. One reason is because you need to have lots
of participants to get statistically reliable results. Another is that bottom-line data does
not always tell you what problems need to be fixed, it just tells you that people are going
too slowly or are making too many errors. Bottom-line data is better for later phases of
design, when tuning the performance of an existing Web site. It is also better for
comparing two Web site designs, to show that one is superior to another in some
measurable aspect. This can be especially important when trying to convince
management to make either a major change or to make a change on an important page,
such as the home page of a high traffic site.

A.2. Setting Up the Tasks

The next step is to choose several representative tasks. By this, we mean tasks that your
target customers are realistically likely to do on your Web site. When choosing tasks,
choose some that are simple in difficulty, some that are medium, and some that are hard.
Ideally, these will have already been worked out in the task analysis you carried out when
trying to know who your customers are and can just be taken from there (see Chapter 3).

Simple tasks are ones that are short and commonly performed. This includes
things like “find the latest news article about parenting” or “find the phone number and
e-mail address of the help desk.” Success on simple tasks is a binary result: either the
person succeeds or fails.

Medium difficulty tasks are a little longer and harder than simple tasks. An
example of a medium difficulty task would be “purchase the cheapest printer you can
find.” Other medium difficulty tasks include “print out a list of all your previous
purchases” and “add a message to the gourmet cooking community board.” These tasks
span a few web pages, but are reasonable things that people would do. Some medium
difficulty tasks will have binary success metrics, some will be more open-ended.
Hard tasks span many web pages and are fairly involved. Examples of hard tasks would be “make the Web site show you only the stocks you are interested in” or “purchase additional ink toner cartridges for the printer you just bought.” Some other hard tasks could be “buy a digital camera for a friend that he or she will like” or “buy a toy for your friend’s one-year-old child.” Most hard tasks are freeform in nature and determining how successful participants are will take some judgment.

Tasks should be about what people want to do. Be careful not to direct people on how to do the task. For example, instead of phrasing a task as “go to ‘My profile’ and find your previous purchases,” it should be “find all of your previous purchases.” Again, the task should be worded the way people would ordinarily think about the problem, that is what, not how. Another example of careful wording is “make the Web site show you only the stocks you are interested in.” Not as realistic would be “customize your profile to show you the stocks you are interested in,” since the words “customize” and “profile” are probably not likely to be part of people’s vocabulary. Another reason is that it might lead people on, especially if there are links labeled “customize” or “profile.”

Tasks should be realistic of what people want to do. For example, “create a new user account” is something that many people do on a Web site, but it is not because they want to. People only create an account because they have to, to get something else done. In other words, creating an account is more of a secondary task that people do to accomplish a primary task.

“Buy a digital camera for a friend that he or she will like” is very open ended, but is likely to be the way people approach the problem. It is important that tasks are realistic, since you want to find out what people are thinking and see if the design provides the right cues to support them.

The tasks should form a complete story. When taken as a whole, the tasks should be complete, forming a cohesive and believable story. For example, it does not really make sense if the ordering is “find previous purchases”, then “add a message to a community board,” and then “find the privacy policy.” The tasks need to flow together. For example, an ordering that makes more sense is, “find the privacy policy,” “purchase a printer,” and then “purchase additional ink toner cartridges.”

Also watch out to make sure you are not fragmenting tasks: “purchase the best printer for under $300” makes more sense than 1) “create an account”, 2) “find and compare printers for under $300”, and 3) “purchase the printer you found.” Testing fragmented tasks may show that users can complete the sub-tasks just fine, but when they are put together in a more realistic situation the results may not be nearly as good.

The number of tasks to have depends on how extensively you want to test your Web site. Five to ten tasks is about right for most cases, enough to cover a lot of functionality without taking a lot of time for each participant.
A.3. Recruiting Participants

After defining some tasks, you should begin recruiting participants. These participants need to be representative of eventual customers in terms of vocabulary, general knowledge, and desired tasks. If the Web site is aimed towards college students, then advertise at a nearby college. If the Web site’s is for mothers of young children, then get friends of friends that are also mothers.

Avoid friends and family. One thing to avoid is getting close friends or family to help out unless you are sure that they will give honest feedback. They may be reluctant to criticize something that you have worked so hard on. Also, do not get co-workers from down the hall. They are likely to know too much about what you are doing. It is okay to use people like this as a first pass, as a way of getting quick comments on a design and to pilot your experimental procedures, but do not rely solely on their feedback. Again, get people that would realistically use the Web site.

Buy participants’ time with gifts and prizes. One way to recruit people is to compensate them for their time. You might be surprised what some people would do for a free T-shirt. Some other ways of drawing in people include giving small toys, coffee mugs, gift certificates, some money, or giving a large cash prize of $200 to $300 to the participant that “does the best.” For straight cash payments, we normally offer around $20 per hour for university students and around $50 per hour for other participants.

If you do not have the time to go out and recruit participants, there are several market research firms who can recruit participants that meet the profile that you need for around $100/participant, not including the compensation you must pay each participant. Many usability practitioners and designers go this route, although it may double your direct costs for running the tests.

Getting the right number of participants. You do not need that many participants to get process data. If you are in the early stages of design, five or six people will be fine, especially for paper prototypes. You will need more people, often around ten to twenty participants, in the later stages of design to evaluate the site. However, you will need to increase these numbers to cover large and diverse audiences, or if your Web site is very large. It may seem expensive to get this many people right now, but consider how much trouble you will be saving later on in creating a more useful and usable Web site for your customers.

Getting bottom-line data requires a lot more people. Ten to twenty people can get you initial data, but there will still be large variability in most tasks. See section A.5 below on analyzing the data for more details on the relationship between the number of people and variability in the data.

---

2 Then again, if you have been working in the computer industry, you might not!
When recruiting, get a few more people than are really needed. The first few tests you run may be a little rough, and you may have to make some changes to make the evaluation flow smoothly. Also, not everyone remembers to show up.

There are two things you should do when recruiting people. First, give them a general overview of the experiment, describing what the Web site is about, what they will do, and approximately how long the whole thing will take. Do not give out too many details because you do not want to bias the test. Second, tell them about any prizes or compensation that will be given for participating. If a person agrees to be a participant, schedule a time and place for the test, and then get their name and either their phone number or email address so that you can give them a reminder before the test.

Choosing between groups vs. within groups experimental design. One important consideration in experimental design is to decide whether each participant participates in more than one experimental condition. Say you are testing two versions of a Web site, for example, to compare them against each other. In a between groups experiment, you break your pool of test participants into two groups and each group uses only one of the Web sites. In contrast, with a within groups experiment, you have only one group of test participants and each participant uses both sites.

These two types of experimental design have tradeoffs. For example, a within groups experiment may not require as many test participants before finding statistically significant results. This can save you considerable time and money if you are trying to get bottom line data. On the other hand, a within groups experiment can raise issues of validity if there are learning effects involved. For example, if you test the same tasks on two versions of the same Web site, your participants might be quicker completing a task the second time since they learned how to do it on the first site. You can alleviate some of these problems by randomizing or counterbalancing the order of sites tested and other experimental conditions.

In general, within groups experiments work better when testing a low-level interaction technique, such as finding the best position for a particular button on the page. Use between groups experiments when you want to compare tasks on two versions of a site or between two competitive sites. Try to make sure that the participants in the two groups match up as well as possible in terms of their demographics, Internet experience, and familiarity with the problem domain.

A.4. Running the Test

There are several considerations you need to be aware of with the actual test itself, from where you run it to what you say to the participants. We discuss these issues and procedures here.
Setting Up the Test Location

If you are going to evaluate a paper prototype, you can do the test practically anywhere. All you need is a large table and places for everyone to sit. For online prototypes, the testing location just needs to be a quiet place with a networked computer.

Video cameras and audio recorders are useful to have in both cases but they are not required. Some tests can be accomplished simply by taking notes on paper, though audio and video recordings make it easy to clarify specific issues later. In contrast, some companies have special rooms to do testing, complete with expensive recording equipment, eye-tracking devices, and one-way mirrors for observers. These kinds of setups are useful for gathering bottom-line data but are not necessary for process data.

Ethical Considerations

Tests can be a grueling experience for some people. Participants have left in tears before, embarrassed by their mistakes or their inability to successfully complete the tasks. You have a responsibility to alleviate these kinds of problems. One way of doing this is by avoiding putting on pressure to participate. You need to get participants’ informed consent as to what the test is about, and then make it clear that the test is voluntary and that they can stop the test at any time for any reason (see the sample form in Appendix C). You also need to stress that you are testing the Web site and not them, and that they are really helping you by finding problems with the site. If they are having problems, then it is the Web site’s fault, not their’s.

If other people are going to see the collected data, then the data should also be anonymized as much as possible. This includes removing names and other pieces of identifying information, as well as blurring out any pictures and videos with people’s faces. There have been cases where a videotape of a person struggling with a user interface was played, with that same person in the viewing audience! Anonymizing the data will help prevent any potentially awkward situations in the future.

Test Roles

The key role in running a usability test is the Facilitator. The Facilitator greets participants, introduces any other people in the room, explains the procedure for the test, and answers any questions participants may have.

The other people act simply as Observers, watching what participants do. Their role is to take notes and to keep quiet. Observers can also be remote if the setup allows them to view things from another location or through a two-way mirror.

If you are running a test on a paper prototype, then another role you will need is the Computer. The job of the person playing Computer is to run the interface, updating the paper interface as needed. Their secondary job is to just smile at all the bad jokes about being slow and needing to upgrade the Computer.
Running a Pilot Test

Before running the tests with actual participants, you should carry out a pilot test with two or three people. In this case, it is okay if these people are your co-workers or friends. The key here is to get used to the procedure of running a test, as well as working out any bugs you may have in your procedure. A pilot also helps you figure out how long the test will take, so you know whether you need to cut or possibly add more tasks.

After you have finished the pilot tests, try analyzing the collected data. This data should not be used in the final analysis, but you should do it to make sure that you are collecting the right data. For example, one time we were evaluating a Web site and asked people to sort a list of features according to importance. While this was useful, it turned out that the data we gathered was extremely difficult to analyze properly. One person mentioned that only the top two things in the list were really important to her, while another mentioned that it was the top four items. In retrospect, a better way of doing this would have been to ask people to rate the importance of each feature from 1 (“not important”) to 7 (“very important”). This is something that we would have caught had we tried analyzing the data from the pilot test before carrying out the real test.

Testing Paper Prototypes

Paper prototypes are useful for getting process data early on, but you should not use them for bottom-line data since they are too far removed from the final implementation. Most people have not seen paper prototypes before. You will have to explain the concept to them, but most people get it pretty quickly.

Ask the participant to point at things with their finger, using it as a mouse. If they click on a link, then the person playing the Computer just switches to another piece of paper representing the next page. If they click on a drop-down menu, then the Computer can place an index card with the choices on top. It is useful if the paper prototype is larger than it would be in reality, to make it easier for everyone to see what the participant is pointing at (see Figure 1).
It is difficult to simulate highly interactive elements with paper prototypes, such as mouse rollovers and animations. In most cases, this is a good thing, as it forces design teams to focus on the core issues first. Be aware that this is a limitation of paper prototypes, and plan accordingly.

Testing Online Prototypes

Online computer-based prototypes can be used for obtaining either process data or bottom-line data. If you are testing a high-fidelity prototype, then you need to make it clear to the participants that they will be testing out an early design and not the final Web site. They might get the misconception that the Web site is “nearly done” when it is really in the early stages of design. Setting their expectations properly will help them to give you the type of high-level feedback that you will need at this stage, rather than comments on the visual, such as colors and fonts. Later, you can test again to evaluate these details.

Before starting a test, make sure to clear the web browser’s history and cache. This will make it as if the person has never been to the Web site before, making all the links unvisited.
Starting and Carrying Out the Test

**Greet the participant.** Start out by introducing yourself and the rest of the team. Then describe the purpose of the test at a high-level, and be sure to emphasize that you are testing the Web site and that you are not testing them in any way. Say something like, “We’re asking you to help us improve the Web site by helping us find problems with it. We’re testing the Web site and not you.”

You also want to make it clear that you will not provide help as they go through the tasks, because you want to see how they would go through the Web site normally. However, emphasize that it is all right for them to stop the test at any time for any reason.

This is also a good time to stick a “Do Not Disturb” sign on the door saying that there is a user study in progress. You do not want any interruptions.

**Fill out the paperwork.** After greeting the participants, get them to fill out any paperwork you may have. This can include basic demographic information, a name and address you can send a check to if you are paying them, and any consent forms. The consent forms should explain what the test is, what kinds of data will be collected, and how the data will be used. Make sure you have two copies of the consent forms, one for you and one for the participant to keep.

**Ask them to think aloud.** If you are gathering process data, then ask the participant to think aloud, to say what they are looking for and what they are trying to do. While some people are really good at this, it is a little awkward for others. The Facilitator should prompt participants every so often if they stop talking, asking things like, “So what are you looking for now?” or “What are you trying to do now?”

Do not do this if you are collecting bottom-line data, as thinking aloud may cause participants to make more errors or to go through the Web site more slowly.

**Give the participants instructions on how to start.** Ask the participants if they have any questions before starting. Then, hand them any instructions you may have, any special information, such as a fake credit card number to use, as well as the first task to complete. Ask them to read the task aloud. This will help them start thinking aloud.

<table>
<thead>
<tr>
<th>Some Common Mistakes when Running Usability Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Testing a Web site using unrealistic tasks</td>
</tr>
<tr>
<td>• Using significantly leading or biased tasks when comparing Web sites</td>
</tr>
<tr>
<td>• Recruiting participants that do not represent your expected customers</td>
</tr>
<tr>
<td>• Forgetting to clear out the web browser file cache and history list before starting</td>
</tr>
<tr>
<td>• Using only a computer that has a really fast network connection, a high resolution monitor, and fast processor (unless your customers really do have these)</td>
</tr>
</tbody>
</table>
After completing each task, you may want to have participants fill out a very short survey. You can ask questions like how easy or hard they thought the task was. You can also ask questions to make sure that they found the right piece of information. For example, if the task is to find and add a specific item to the shopping cart, you can ask them how much the item cost. This is just a redundancy check, to make sure that they really did complete the task successfully.

**Take good notes during each task.** The Observers should be taking notes about what the participant was saying and doing during the test. It also helps to record audio and video if possible. Use a digital watch or a clock to keep track of time, too. If a certain task takes far too long, then let the participant know that it is okay to just move on to the next task.

If you are measuring bottom-line data, then make sure everyone knows what to measure. For example, is it an error if someone hits the “Back” button on the browser? Is it an error if someone goes back to the home page? These need to be agreed upon beforehand. Also, what happens if someone does not finish a task? While there are not any hard rules here, a common technique is to assign a very large time and a large number of errors, just to keep everything numeric.

**Watch closely.** Yes, it will be frustrating watching someone struggle with something you put so much time into, clicking on the wrong link or not seeing the text that is right there in front of them. But bite your lip and keep your mouth closed: you are here to watch and to learn how to improve the Web site. Make sure that none of the Observers laugh, groan, or makes any other inappropriate response. These are the types of things that can unnerve your participant.

If a participant does something really interesting, ask a follow-up question. Ask open-ended questions, such as “what are you looking for?” Let the participant know things are going alright. Prompt them to keep speaking and tell you what they are thinking. Also, look out for non-verbal cues, such as a furrowed brow or a puzzled look.

It is okay to answer any general questions participants may have, but do not help them with the tasks. Also, do not help some participants more than others. Plan in advance what you will and will not help out with. For example, it is common to decide that you will help participants when they run into known bugs or functionality that has not been implemented yet. Simply get them back on track.

**Finish up with a quick survey.** Follow up with a short survey after all of the tasks have been completed. You want to get their overall impressions and comments about the Web site, seeing what they liked and disliked about the Web site. Also ask them where they felt they had problems with the site and where they thought it worked well.

**Debrief the participants after the test.** Wrap up by debriefing each participant, telling them what you were looking for, as well as discussing any interesting behavior the
participant had. People often do not remember specific actions, so it is useful to go through the Web site again or to show video segments to help prompt their memory.

Ask participants if they have any thoughts on how to fix any problems encountered. Take these comments with a grain of salt, since participants usually do not have an understanding of design or the underlying technology, but they are still useful to hear. Afterwards, finish up by asking if they have any final questions, and then thank them for their time.

A.5. Analyzing the Data

Analyzing process data. Think about what you saw and what the participants said. Did they understand the things you thought they would? Were they confused by any terms or concepts? If so, then maybe things need to be renamed or explained in greater detail. If it is a concept fundamental to the Web site, then make sure you make it clear on the homepage, because otherwise people might leave without ever bothering to figure it out.

What errors could they recover from? For example, did they click on a link but then quickly realize that it was the wrong one? It is important to minimize these kinds of problems, but these are usually just minor annoyances. A bigger problem would be if you observed systematic “ping-ponging,” that is repeated back and forth attempts from one page down unfruitful paths. This would indicate a need for more DESCRIPTIVE, LONGER LINK NAMES (K9), which will give the participant more “information scent” to find the page they are looking for.

Focus on the errors that participants could not recover from first. Did participants have problems finding items in the Web site? Did they have trouble understanding the overall structure of the Web site? What about navigation? Could they manage to make their way through the site adequately? Did they make any errors and not even notice that there was a problem? These could be fundamental problems that the site has and should be addressed first when fixing the site.

The most important question to ask yourself is why the error occurred. Was the navigation too confusing, making it difficult to go to other pages? Was the information disorganized, making it hard to find things across pages? Was the Web page too cluttered, making it hard to find anything on a page? Was it because the Web site was too slow and the participants lost track of what they wanted to do? Just like a doctor, you only get to see the symptoms, but you need to keep asking yourself if there are any fundamental issues that are causing all of these problems.

---

3 A perfect example of this is the infamous “Butterfly Ballot” used in the 2000 Presidential election in the United States. Statistical analysis of nearby counties suggests that at least several thousand citizens unintentionally voted for the wrong candidate, but the voting process lacked a verification process that would let people check their votes.
Another thing to keep in mind is that people do not give up in usability tests as often as they would in the real world. You have to realize that no matter what you do, you are still putting an implicit amount of pressure on them to try their best to successfully complete the task. People are more attentive and willing to go through a few more pages when they know they are being observed.

**Analyzing bottom-line data.** You need to be a little careful when analyzing bottom-line data. For example, suppose that our target goal was to make it so that a person new to a Web site can find and purchase an item in 20 minutes or less. When running our test, we got times of 20, 15, 45, 10, 5, and 25 for our six participants. The mean or average time for this is 20 minutes. Looks pretty good! The median for this set of numbers is 17.5, sounding even better here!

However, the problem is that there is very little certainty here because there are only six participants, and the results are highly variable. If you calculate the standard deviation, a measurement of how variable this set of numbers are, you will find that it is around 14. If we divide the standard deviation by the square root of the number of samples we have (6), we get 5.8. This is the “standard error of the mean” and it tells us how much variation we can expect in the typical value. It is plausible that the typical value is as small as the mean minus twice the standard error of the mean, resulting in a lower bound of 8.5, or as large as the mean plus twice the standard error of the mean, or 32. This latter value would clearly be far from our stated goal of 20 minutes!

We can say more precisely what we mean by plausible. The best thing to do here is to use statistical techniques. Cranking through basic statistical methods, you can calculate with 95% confidence that the actual average time will be 20 plus or minus 11 minutes. In other words, you are 95% likely to be correct in saying that the actual time will be in this range, but 5% of the time you’ll be wrong.4

Usability test data is often quite variable, which means that you need lots of participants to get good estimates of typical values. Also, it turns out that the breadth of range depends on the square root of the number of participants. In other words, if you have 4 times as many participants, you only narrow the range by an average factor of 2. Continuing the example, in general, quadrupling the number of participants from 6 to 24 will narrow the spread of the average time from 20 plus or minus 11 minutes to 20 plus or minus 6 minutes (assuming that the mean and the standard deviation stay around the same). This is where online usability evaluation methods become useful, as they make it easier to scale up the number of participants and thus tighten your confidence intervals.

Basic statistics is beyond the scope of this book, but a great introduction on the topic is *The Cartoon Guide to Statistics* by Larry Gonick Jr. This book covers the main concepts you will want to be familiar with when doing basic statistical analyses, which include mean, variance, standard deviation, correlation, regression, t-test, and ANOVA.

---

4 If you use Microsoft Excel, you can calculate this range using the *Confidence* function. If you use a more advanced tool like SPSS, then you probably already know how to calculate this.
A.6. Presenting the Results

After the data has been collected and analyzed, the results need to be presented to the design team or to the clients. This can be in the form of a written report or an oral presentation. Here is a short outline of what you should have:

- Executive Summary
- Tasks
- Participants
- Problems Found
- Participant Feedback
- Suggested Improvements
- Appendices

Start with the Executive Summary, which gives a quick overview of what you did in the test, a summary of the results, and a rundown of the recommendations for improvement. Next, in Tasks, talk about the tasks that you had participants do, describing why these tasks were chosen. Continue with Participants, a short description of the number of participants, general demographics, and any defining characteristics.

The next section, Problems Found, lists the problems encountered, prioritized by severity. Use screenshots of problem Web pages, using circles and arrows to point out critical incidents. Graphs showing how successful people were at completing tasks will also help people get it. If you are doing an oral presentation, this is a good time to show video clips to help get your message across. This is an extremely valuable technique for convincing skeptical programmers and management that there are any problems with the Web site. You can also include video clips in written reports that you will be putting online.

The Participant Feedback section contains both positive and negative feedback from participants. This can include summaries of surveys taken by participants after they finished the test, or direct quotes from them during the test.

The next section, Suggested Improvements, outlines what needs to be changed to improve the Web site. The improvements should be triaged into “must do,” “should do,” and “could do” categories. The “must do” improvements are the show stoppers, the ones that caused serious problems that people could not recover from. They also include really simple ones that only take a short time to fix, such as misspellings or broken links. The “should do” improvements represent problems that are annoying but are tolerable, ones that most people can still figure out. The “could do” improvements are changes that will take too much effort to implement for the resulting benefits. Keep these ideas on the backburner for the next iteration.

The last section, Appendices, contains any test materials used during the experiment, such as demo scripts and instructions, as well as all of the raw data in a cleaned up form.
Appendix B – Sample Web Site Evaluation Plan

Roles

**Facilitator:** reads instructions, handles transitions from one section of test to another, fields participant questions, and gets participants unstuck if necessary

**Observers:** record times, on screen events, and tallies tracked metrics

Introduction

Thank you very much for helping us to evaluate two Web sites. We are testing people’s perceptions of Web sites that are in the early stages of design. Here’s what we have planned for the next *(insert timeframe here)*:

1. First, we will start a web browser, open up a start page, and ask you to read through the introductory text.
2. Next, we will ask you to perform some tasks on the first Web site. Interspersed will be some survey questions asking you about your perceptions of the tasks and of the Web site.
3. Steps 1 and 2 will be repeated for the second Web site.
4. At the end, we will ask you for any comments you have overall.

*(Disclaimer)*

We’re asking you to help us improve the Web site by helping us find problems with it. We would like to stress that we are testing the Web sites, not you. If you have trouble with some of the tasks we ask you to perform, it is the Web site’s fault, not yours. Don’t feel bad, trouble spots are exactly what we’re looking for. And please remember that this is totally voluntary. Although we don’t know any reason why this should happen, if you become uncomfortable or find this objectionable in any way, feel free to quit at any time.

*(Hand them two consent forms, one for us, one for them as a copy. Also hand them any forms for demographic information or for getting any prizes or checks to them.)*

This consent form just says that you understand what this test is about, that you understand we will respect your privacy wishes, and that you will allow us to publish any results from this study.

*(Wait until the participant completes the forms.)*

Before we begin, I’d like to ask you to say what comes to your mind as you work. We have found that we get a great deal of information from these informal observations if we ask people to think aloud as they work through the exercises. It may be a bit awkward at
first, but it’s really very easy once you get used to it. All you have to do is speak your thoughts as you work. If you forget to think aloud, I’ll remind you to keep talking.

Do you have any questions for us before starting?

**Tasks**

Be sure to do the following for each participant:
- Give them scrap paper and pen
- Give them sample address and sample credit card number
  (“The two Web sites we will be asking you to test are e-commerce Web sites. Here is an address and fake credit card number to use. No actual purchases will be made.”)
- Start browser
- Clear out web browser cache and history
- Maximize web browser size
- Hand them the first task on a sheet of paper
- Ask them to read the first task aloud

**Debriefing**

Do you have any final comments about the Web sites, this study, or anything else?
Appendix C – Sample Consent Form

Our names are (insert your names here), and we are (describe your position and the organization you work for here). We would like you to participate in our research, which involves the evaluation of two e-commerce Web sites. This study should take about (insert timeframe here) and poses no risks to you other than those normally encountered in daily life.

All of the information that we obtain from your session will be kept confidential. The information obtained from your session will be tagged with a code number. The correspondence between your name and number will be treated with the same care as our own confidential information. We will not use your name or identifying information in any reports of our research (unless you allow it by signing the second line below).

Your participation in this research is voluntary. You are free to refuse to participate. Whether or not you choose to participate will have no bearing on your standing in relation to (insert organization name here).

If you have any questions about the research, you may call (insert contact person here) at (insert contact phone number here), or send electronic mail to (insert contact email here). You may keep the other copy of this form for future reference.

By signing this form you agree to the following statements:

I agree to participate in the evaluation of two Web sites. I know that the researchers are studying (insert description here). I realize that I will be asked to test the Web sites and discuss perceptions of those two Web sites over (insert timeframe here).

I understand that any information obtained during this study will be kept confidential.

I give (insert your names here) and their associates permission to present the results of this work in written or oral form, without further permission from me.

Date & Signature

____________________________________________

I also agree to allow my name or other identifying information, such as a picture or video, to be included in all final reports and publications resulting from my participation in this research.

Date & Signature

____________________________________________
Appendix D – Sample Observer Form

ID: _____________
Date: _____________
Time started: _____________ Time ended: _______________

**Tallies (make a mark for each incident)**

<table>
<thead>
<tr>
<th>Site A</th>
<th>Site B</th>
<th>Incident</th>
</tr>
</thead>
<tbody>
<tr>
<td>______</td>
<td>______</td>
<td>Could not figure out what to do next for more than 30 secs</td>
</tr>
<tr>
<td>______</td>
<td>______</td>
<td>Was visibly lost in the Web site</td>
</tr>
<tr>
<td>______</td>
<td>______</td>
<td>Was visibly frustrated with the Web site</td>
</tr>
<tr>
<td>______</td>
<td>______</td>
<td>Said something clearly negative about the Web site</td>
</tr>
<tr>
<td>______</td>
<td>______</td>
<td>Said something clearly positive about the Web site</td>
</tr>
<tr>
<td>______</td>
<td>______</td>
<td>Cursed out loud</td>
</tr>
</tbody>
</table>

**Tasks Completed (mark whether successfully completed or not)**

<table>
<thead>
<tr>
<th>Site A</th>
<th>Site B</th>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>______</td>
<td>______</td>
<td>1. Find the Web site’s privacy policy.</td>
</tr>
<tr>
<td>______</td>
<td>______</td>
<td>2. Find the two cheapest mp3 players on the Web site.</td>
</tr>
<tr>
<td>______</td>
<td>______</td>
<td>3. Find a gift for a friend and add it to the shopping cart.</td>
</tr>
<tr>
<td>______</td>
<td>______</td>
<td>4. Find a gift for yourself and add it to the shopping cart.</td>
</tr>
<tr>
<td>______</td>
<td>______</td>
<td>5. Check out and finalize the purchase.</td>
</tr>
<tr>
<td>______</td>
<td>______</td>
<td>6. Check the status of the purchase.</td>
</tr>
<tr>
<td>______</td>
<td>______</td>
<td>7. Subscribe yourself to the Web site’s newsletter.</td>
</tr>
</tbody>
</table>

Notes: