CHAPTER

Empirical UX Evaluation: Preparation

23

Be prepared; that's the Boy Scouts' marching song… Don't be nervous, don't be flustered, don't be scared; Be prepared!

– Tom Lehrer

Highlights

- Empirical UX evaluation plan.
- Evaluation scope and rigor.
- Goals for empirical UX evaluation.
- Select team roles for empirical UX evaluation.
- Prepare an effective range of user tasks.
- Recruit participants.
- Prepare for the session.
- The UX evaluation session work package.
- Do final pilot testing before evaluation.

23.1 INTRODUCTION

23.1.1 You Are Here

We begin each process chapter with a "you are here" picture of the chapter topic in the context of The Wheel, the overall UX design lifecycle template (Fig. 23-1). This chapter is about how to prepare for empirical UX evaluation. Much of this will apply to planning for other kinds of evaluation, too.

Although, for completeness, we include quantitative UX data collection techniques in Chapter 24 and quantitative data analysis in Chapter 26, this is emphasized less than it used to be in previous usability engineering books

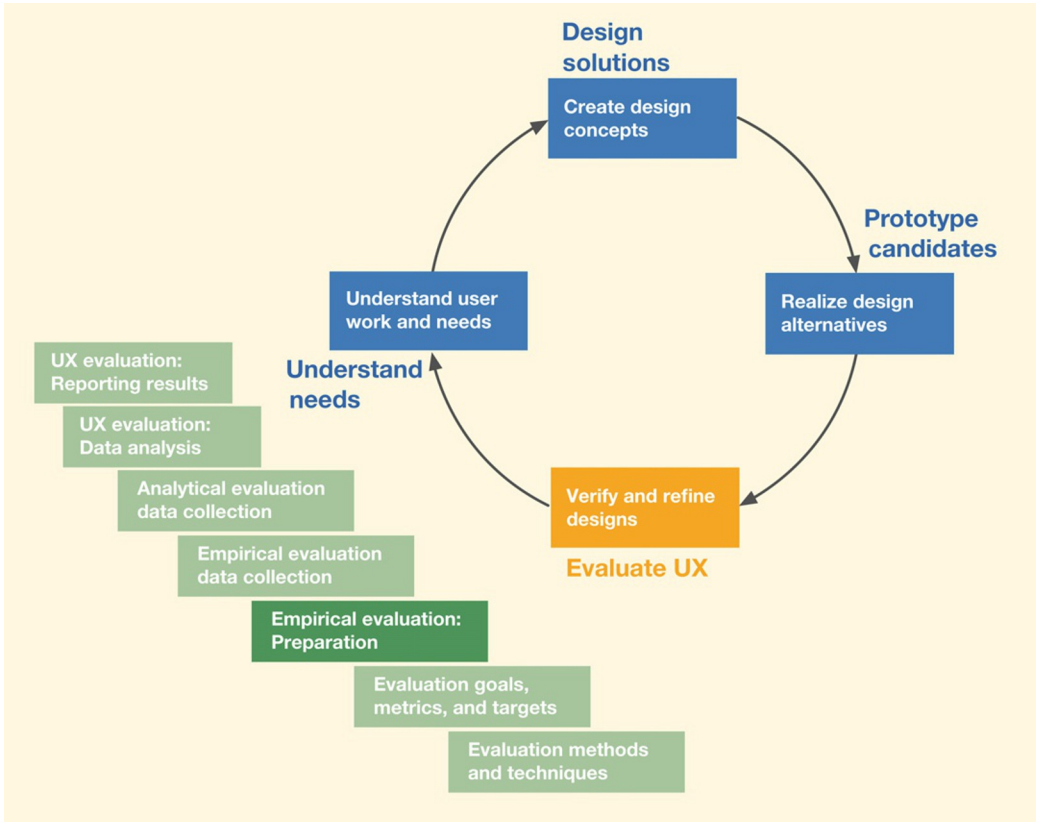

Fig. 23-1
You are here in the chapter on preparing for empirical evaluation within the Evaluate UX lifecycle activity in the context of the overall Wheel lifecycle template.

because of less focus in practice on quantitative user performance measures and more emphasis on qualitative evaluation to reveal UX problems to be fixed.

23.1.2 A Plan for the Empirical UX Evaluation Session

Empirical UX evaluation methods (Chapter 24) entail data observed in the performance of real user participants and data coming directly from participants. The purpose of your plan for empirical UX evaluation is to identify the most appropriate evaluation goals, methods, activities, conditions,

constraints, and expectations for your project. If the plan will be read by people outside your immediate project group, you might want an upfront "boilerplate" introduction with some topics such as these, described very concisely:

- Overview of plan.
- Overview of product or parts of product being evaluated (for people outside the group).
- Goals of the product user interface (i.e., what will make for a successful user experience).
- Description of the intended user population.
- Overview of approach to informed consent.
- Overview of how this evaluation fits into the overall iterative UX process lifecycle.
- Overview of the UX evaluation process in general (e.g., preparation, data collection, analysis, reporting, iteration).
- General evaluation methods and activities planned for this session.
- Estimated schedule.
- Responsible personnel.

The body of the plan should start with evaluation goals (next section) and should also include a description of your approach and mechanics:

- Description of resources and constraints (e.g., time needed/available, state of prototype, lab facilities, and equipment).
- Pilot testing plan.
- Approach to evaluation, choices of data collection techniques.
- Mechanics of the evaluation (e.g., materials used, informed consent, location of evaluation, UX goals and metrics involved, tasks to be explored, including applicable benchmark tasks).
- All instruments to be used (e.g., benchmark task descriptions, questionnaires).
- Approaches to data analysis.
- Specifics of your approach to evaluate emotional impact and, if appropriate, long-term emotional aspects of interaction.

23.2 EVALUATION SCOPE AND RIGOR

23.2.1 Evaluation Scope

You can perform UX evaluation at any scope. A large-scope approach is best for the early part of the agile UX funnel, where you would evaluate the overall product or system architecture and user workflow, and the conceptual design—addressing the ecological layer of the needs pyramid.

Participant

A participant, or user participant, is a user, potential, or user surrogate who helps evaluate UX designs for usability and user experience. These are the people who perform tasks and give feedback while we observe and measure. Because we wish to invite these volunteers to join our team and help us evaluate designs (i.e., we want them to participate), we use the term "participant" instead of "subject" (Section 21.1.3).

Emotional Impact

An affective component of user experience that influences user feelings. Includes such effects as enjoyment, pleasure, fun, satisfaction, aesthetics, coolness, engagement, and novelty and can involve deeper emotional factors such as self-expression, self-identity, a feeling of contribution to the world, and pride of ownership (Section 1.4.4).

Scope (of Delivery)

Describes how the target system or product is "chunked" (broken into what size pieces) in each iteration or sprint for delivery to the client and users for feedback and to the software engineering team for agile implementation (Section 3.3).

A small-scope approach is best for frequent iterations within the sprints of the late agile UX funnel. Small-scope evaluation is used to address task-level design in the interaction level of the pyramid.

As you might guess, you can address the emotional-needs layer of the pyramid at any scope.

23.2.2 Evaluation Rigor

You can perform UX evaluation at any level of rigor. You can perform low-rigor evaluation in early design stages when the design is changing rapidly and careful attention to preserving detail in the data would be wasted. You might also have to perform your UX evaluation at a relatively low level of rigor in the late funnel simply because of the pressure to keep up with the agile sprints.

Larger organizations used to devote huge resources to do rigorous UX evaluation and rigorous methods still have a place in projects that demand it. However, with the shift of attention toward agile methods, it is more difficult to justify the time and cost in most projects. High rigor in UX evaluation requires careful attention to detail and full preservation of the data—both data purity and data completeness. Because of the amount of detail we give for completeness, our descriptions may appear to represent a very rigorous view of the process. But, in fact, what you read here can be applied at any level of rigor.

23.3 GOALS FOR AN EMPIRICAL UX EVALUATION SESSION

One of the first things to do in an evaluation plan is to set and prioritize your evaluation goals specific to this project. Identify the most important design issues and user tasks to investigate. Decide which parts of the system or functionality you simply will not have time to look at.

Your evaluation goals can include:

- Evaluation range (parts of the system to be covered by this evaluation).
- Scope (size of chunk) at which evaluation will be applied (Section 23.2.1).
- Rigor (level of formality and completeness) to be applied (Section 23.2.2).
- Layers of the needs pyramid to be addressed (from the top: emotional, interaction, and ecological needs, Section 12.3.1).
- Types of data to collect (Section 21.1.4), especially whether it will involve quantitative data.

- UX goals, targets, and metrics, if any, to be addressed (Chapter 22).
- Matching this evaluation to the current stage of product design evolution (early design ideas to explore versus near-final prototype).

23.4 SELECT TEAM ROLES

23.4.1 Participation and Buy-In

Encourage your whole project team to participate in at least some evaluation. Broad participation begets buy-in and ownership, necessary for your results to be taken as a serious mandate to fix problems. Roles include the facilitator, the prototype "executor," all data collectors, and other supporting roles. Anyone on the team can learn as an observer.

23.4.2 Facilitator

Your facilitator is the leader of the evaluation team, the orchestrator, and the one who makes sure it all works right. The facilitator has the primary responsibility for planning and executing the evaluation sessions, and the final responsibility to make sure the laboratory is set up properly. Because the facilitator will be the principal contact for participants during a session and responsible for putting the participant at ease, you should select someone with good "people skills."

23.4.3 Prototype Executor

If you are using a low-fidelity click-through wireframe prototype, you need to select a prototype executor, a person to "execute" the prototype and move it through its paces as users interact.

The prototype executor must have a thorough technical knowledge of how the design works. So that the prototype executor responds only to participant actions, he or she must have a steady Vulcan sense of logic. The executor must also have the discipline to maintain a poker face and not speak a single word throughout the entire session.

23.4.4 Quantitative Data Collectors

If you plan to include quantitative data collection, you'll need people to do it. Depending on your UX metrics and quantitative data collection instruments, people designated to collect quantitative data may be walking around with stopwatches and counters (mechanical, electronic, or paper and pencil). These people must be ready to record the quantitative data as it occurs. Because quantitative metrics usually involve simple descriptive statistics (e.g., averages), the data collector may wish to enter performance and other data directly into a spreadsheet.

23.4.5 Qualitative Data Collectors

Select as many team members as possible to serve as qualitative data collectors and recorders. No evaluation team member should be idle during a session. Thoroughness will improve with more people doing the job. Everyone should be ready to collect qualitative data, especially critical incident data.

23.4.6 Supporting Actors

Inspection (UX)

An analytical evaluation method in which a UX expert evaluates an interaction design by looking at it or trying it out, sometimes in the context of a set of abstracted design guidelines. Expert evaluators are both participant surrogates and observers, asking themselves questions about what would cause users problems and giving an expert opinion predicting UX problems (Section 25.4).

Sometimes you need someone to interact with the participant as part of the task setting or to manage the props needed in the evaluation. For example, for task realism you may need someone to call the participant on a telephone in the participant room or, if your user participant is an "agent" of some kind, you may need a "client" to walk in with a specific need involving an agent task using the system. Select team members to play supporting roles and handle props.

23.5 PREPARE AN EFFECTIVE RANGE OF USER TASKS

If evaluation is to be task based, including task-driven UX inspection methods, select appropriate tasks to support evaluation. Select different kinds of tasks for different evaluation purposes.

23.5.1 Benchmark Tasks to Generate Quantitative Measures

Benchmark Task

A task description devised for a participant to perform during UX evaluation so that UX measures such as time on task and error rates can be obtained and compared to a baseline value across the performances of multiple participants (Section 22.6).

Benchmark tasks portray representative, frequent, and critical tasks that apply to the key work role and user class represented by each participant. If you have defined a benchmark task to generate quantitative measures, you should now have the corresponding task description and UX target metrics ready and waiting to guide your data collection and to compare with observed results.

Also, the benchmark task description should be printed and ready to use by participants to generate data to be measured. Make sure each task description says only *what* to do, with no hints about *how* to do it. Also, don't use any language that telegraphs any part of the design (e.g., names of user interface objects or user actions, or words from labels or menus).

23.5.2 Unmeasured Tasks

You might also like to have descriptions for unmeasured tasks, tasks for which participant performance will not be measured quantitatively. Evaluators can use these representative tasks to add breadth to qualitative evaluation by addressing aspects of the design not covered in some way by the benchmark tasks.

In early stages, you might employ only unmeasured tasks, the sole goal of which is to observe critical incidents and identify initial UX problems to root out and fix at least the most obvious and most severe problems before any measured user performance data is useful.

Just as for benchmark tasks created for evaluation UX attributes, you should print out representative unmeasured task descriptions, which should be just as specific as the benchmark task descriptions to give to the participant to perform in the evaluation sessions.

23.5.3 Exploratory Free "Use"

In addition to strictly specified benchmark and unmeasured tasks, the evaluator may also find it useful to observe the participant in informal interaction with the design, a free-play period without the constraints of predefined tasks. This does not necessarily mean that they are even doing tasks, maybe just exploring.

To engage a participant in free use, the evaluator might simply say "play around with the interface for a while, doing anything you would like to, and talk aloud while you are playing." Free use is valuable for revealing participant expectations and system behavior in situations not anticipated by designers, often situations that can break a poor design.

23.5.4 User-Defined Tasks

Sometimes tasks that users come up with will address unexpected aspects of your design (Cordes, 2001). You can include user-defined tasks by giving your participants a system description in advance of the evaluation sessions and ask them to write down some tasks they think are appropriate to try. Otherwise, you can wait until the session is under way and ask each participant extemporaneously to come up with tasks to try.

If you want a more uniform task set over your participants but still wish to include user-defined tasks, you can ask a different set of potential users to come up with a number of candidate task descriptions before starting any evaluation session. This is a good assignment for a focus group. You can vet, edit, and merge these into a set of user-defined tasks to be given to each participant as part of each evaluation session.

Focus Group (in UX Practice)

A small discussion group of representative users or stakeholders aimed at identifying broad themes and issues in a work practice (Section 7.4.4.3).

23.6 RECRUIT PARTICIPANTS

The next step in preparing for empirical UX evaluation is selection and recruitment of participants—finding representative users, usually outside your team and often outside your project organization, to help with evaluation.

This section is mainly focused on participants for empirical UX evaluation, but also applies to other situations where user participants are needed.

23.6.1 Establish Budget and Schedule for Recruiting User Participants Upfront

Finding and recruiting evaluation participants might be part of the process where you are tempted to cut corners and save a little on the budget or might be something you think to do at the last minute. But, to protect the larger investment already made in the UX lifecycle process and in setting up formative evaluation so far, you need to secure a reasonable amount of resources—money in the budget to pay the participants and time in the schedule to recruit the full range and number of evaluation participants you will need. If you do this kind of evaluation infrequently, you can engage the services of a professional recruiter to do your participant recruiting or even the UX evaluation consulting group to do the whole evaluation.

23.6.2 Identify the Right Kinds of Participants

Formal Summative Evaluation

A formal, statistically rigorous summative (quantitative) empirical UX evaluation that produces statistically significant results (Section 21.1.5.1).

In formal summative evaluation, the process of selecting participants is referred to as "sampling," but that term is not appropriate here because what we are doing has nothing to do with the implied statistical relationships and constraints. In fact, it's quite the opposite. You're trying to learn the most about your design with the smallest number of participants and with exactly the right selected (not random) participants. Look for participants who are "representative users," that is, participants who match your target work role's user class descriptions and who are knowledgeable of the general target system domain. If you have multiple work roles and user classes, you should try to recruit participants representing *each* category. If you want to be certain your participants are representative, you can prepare a short written demographic survey to administer to participants to confirm that each one meets the requirements of your intended work activity role's user class characteristics.

In fact, participants must match the user class attributes in any UX targets they will help evaluate. So, for example, if initial usage is specified, you need participants unfamiliar with your design.

23.6.2.1 "Expert" participants

If you have a session calling for experienced usage, it's obvious that you should recruit an expert user, someone who knows the system domain and knows your particular system. Expert users are good at thinking aloud to generate qualitative data. These expert users will understand the tasks and can tell you what they don't like about the design. But you cannot necessarily depend on them to tell you how to make the design better.

Recruit a UX expert if you need a participant with broad UX knowledge and who can speak to design flaws in terms of design guidelines. As participants, these experts may not know the system domain as well and the tasks might not make as much sense to them, but they can analyze user experience, find subtle problems (e.g., small inconsistencies, poor use of color, confusing navigation), and offer suggestions for solutions.

Or you can consider recruiting a so-called "double expert," a UX expert who also knows your system very well, perhaps the most valuable kind of participant.

23.6.3 Determine the Right Number of Participants

The question of how many participants you need is entirely dependent on the kind of evaluation you are doing and the conditions under which you are doing it. There are some rules of thumb, such as the famous "three to five participants is enough" maxim, which is quoted so often out of context as to be almost meaningless. However, it is a good starting point until you learn more about what you need. For further discussion about the "three to five users" rule and its limitations, see Section 28.6.

The good news is that your experience and intuition will be effective touchstones for knowing when you have gotten the most out of an iteration of UX evaluation and when to move on. One telltale sign of having used enough participants is the lack of many new critical incidents or UX problems being discovered with additional participants.

You have to decide for yourself every time you do empirical UX evaluation—how many participants you can or want to afford. Sometimes it is just about achieving your UX targets, regardless of how many participants and iterations it takes. More often it is about getting in, getting some insight, and getting out.

23.6.4 Consider Recruiting Methods and Screening

Now the question arises as to where to find participants. Inform your customer early on about how your evaluation process will proceed so you will have the best chance of getting representative users from the customer organization at appropriate times.

Here are some hints for successful participant recruiting:

- Try to get the people around you (coworkers, colleagues elsewhere in your organization, spouses, children, and so on) to volunteer their time to act as participants, but be sure their characteristics fit your key work role and the corresponding user class needs.
- Newspaper ads and emails can work to recruit participants, but these methods are usually inefficient.

- If the average person off the street fits your participant profile (e.g., for a consumer software application), hand out leaflets in shopping malls and parking lots or post notices in grocery stores or in other public places (e.g., libraries).
- Use announcements at meetings of user groups and professional organizations if the cross section of the groups matches your user class needs.
- Recruit students at universities, community colleges, or even K-12, if appropriate.
- Consider temporary employment agencies as another source for finding participants.

A possible pitfall with temporary employment agencies is that they usually know nothing about UX evaluation, nor do they understand why it is so important to choose appropriate people as participants. The agency goal, after all, is to keep their pool of temporary workers employed, so it will be up to you to screen their candidates against your user class characteristics.

23.6.5 Use a Participant Recruiting Database

If you are going to be doing evaluation often, you should maintain a recruiting database of contact information for your potential participants. Because all the participants you have used in the past should be in this database, you can draw on the good ones for repeat performances.

You can also sometimes use your own customer base or your customer's contact lists as a participant recruiting source. Perhaps your marketing department has its own contact database.

23.6.6 Decide on Incentives and Remuneration

Generally, you should not ask your participants to work for free, so you will usually have to advertise some kind of remuneration. You will usually pay a modest hourly fee (e.g., about a dollar above minimum wage for an off-the-street volunteer). Expert participants cost more, depending on your specialized requirements. Don't try to get by too cheaply; you might get what you pay for.

Instead of, or in addition to, money, you can offer various kinds of premium gifts, such as coffee mugs with your company logo, gift certificates for local restaurants and shops, T-shirts proclaiming they survived your UX tests, free pizza, or even chocolate chip cookies! Sometimes just having a chance to learn about a new product before it is released or to help shape the design of some new technology is motivation enough.

23.6.7 Don't Give Up on Difficult-To-Find User Participants

Be creative in arranging for hard-to-find participant types. Sometimes, the customer—for whatever reasons—simply will not let the developer organization have access to representative users. The navy, for example, can be rightfully

hesitant about calling in its ships and shipboard personnel from the high seas to evaluate a system being developed to go onboard.

Specialized roles (such as an ER physician) have demands on their time that make if difficult, or impossible, to schedule them in advance. Sometimes you can have an "on call" agreement through which they call you if they have some free time and you do your best to work them in.

Sometimes when you cannot get a representative user, you can find a user representative, someone who is not exactly in the same role but who knows the role from some other angle. A domain expert is not necessarily the same as a user, but might serve as a participant, especially in an early evaluation cycle. We once were looking for a particular kind of agent of an organization who worked with the public, but had to settle, at least at the beginning, for supervisors of those agents.

23.6.8 Recruit for Codiscovery

Consider recruiting pairs of participants specifically for codiscovery evaluation. Your goal is to find people who will work well together during evaluation and, as a practical matter, who are available at the same time. We have found it best not to use two people who are close friends or who work together on a daily basis; such close relationships can lead to too much wisecracking and acting out.

Look for people whose skills, work styles, and personality traits complement each other. Sometimes this is a good place to give them the Myers-Briggs test (Myers, McCaulley, Quenk, & Hammer, 1998) for collaborative personality types.

Codiscovery

A qualitative data collection technique employing two or more participants interacting in a team approach to evaluation, usually with a think-aloud data collection technique. Two people can verbalize more naturally, yielding multiple viewpoints expressed within conversational interplay (Sections 21.4.2.3 and 24.2.3.3).

23.6.9 Manage Participants as Any Other Valuable Resource

Once you have gone through the trouble and expense to recruit participants, don't let the process fail because a participant forgot to show up. Devise a mechanism to manage participant contact to keep in touch, remind in advance of appointments, and to follow up, if useful, afterward.

You need a standard procedure and a foolproof way to remind you to follow it for calling your participants in advance to remind them of their appointment, just as they do in doctor's offices. No-show participants cost money in unused lab facilities, evaluator frustration, wasted time, and schedule delays.

23.6.10 Select Participants for Subsequent Iterations

A question that commonly arises is whether you should use the same participants for more than one cycle of formative evaluation. Of course you would not use a repeat participant for tasks addressing an "initial use" UX attribute.

But sometimes reusing a participant (maybe one out of three to five) can make sense. This way, you can get a reaction to design changes from the previous cycle in addition to a new set of data on the modified design from the two new participants. Calling on a previously used participant tells them you value their help and gives them a kind of empowerment, a feeling that they are helping to make a difference in your design.

23.7 PREPARE FOR THE SESSION

23.7.1 Lab and Equipment

If you are planning lab-based evaluation, the most obvious aspect of preparation is to have the lab available and configured for your needs. If you plan to collect quantitative UX data, prepare by having on hand the right kind of timers for tasks and counters for errors, from simple stopwatches to instrumented software for automatically extracting timing data.

If you think you should evaluate outside the lab or with special props or environmental conditions, see Section 22.6.4.4 for more on this topic.

Example: A Modern UX Lab at Bloomberg LP

Bloomberg LP, a leader in financial informatics, employs a modern UX evaluation lab with two areas—a participant room and an observation room—each with an independent entrance and separated by a one-way mirror. The participant room has a multimonitor workstation (Fig. 23-2) on which Bloomberg's desktop applications are evaluated.

Fig. 23-2
Desktop evaluation in the Bloomberg UX evaluation lab.

On the other side of this participant room, there is another station designed for evaluations with paper prototypes (Fig. 23-3) or mobile devices (Fig. 23-4).

The left part of Fig. 23-4 shows this station being used during an evaluation of Bloomberg's mobile application. The right part shows a close up of the mobile device holder with a mounted camera, allowing the participant to hold and move

Fig. 23-3
Paper prototype evaluation in the Bloomberg UX evaluation lab.

Fig. 23-4
Mobile devices evaluation in the Bloomberg UX evaluation lab.

the mobile device as she interacts while the mounted camera captures the user interface and her actions.

Fig. 23-5 is a view of the observation room, which is kept dark to prevent people in the participant room from seeing through. In this image you can see the participant room showing through the one-way mirror. The lab is set up to stream up to five selections of the seven video sources and four screen capture sources from the participant room to the large screens seen at the top in the observation room.

Fig. 23-5
The observation room in the Bloomberg UX evaluation lab.

This lab has been instrumental in defining the UX designs of Bloomberg's flagship desktop and mobile applications. Special thanks to Shawn Edwards, the CTO; Fahd Arshad, the head of UX Design; Pam Snook; and Vera Newhouse at Bloomberg LP for providing us these lab photos.

23.7.2 Session Parameters

Evaluators must determine protocol and procedures for conducting the evaluation—exactly what will happen and for how long during an evaluation session with a participant.

23.7.2.1 Task and session lengths

The typical length of time of an evaluation session for one participant is anywhere from 30 minutes to two hours. However, it is possible that a real-world UX evaluation session can become a day-long experience for a participant. The idea is to get as much as possible from each user without burning out the participant.

If you require sessions longer than a couple of hours, it will be more difficult for participants. In such cases, you should:

- Prepare participants for possible fatigue in long sessions by warning them in advance.
- Mitigate fatigue by scheduling breaks between tasks where participants can get up and walk around, leave the participant room, get some coffee or other refreshments.
- Have some granola bars and/or fruit available in case hunger becomes an issue.
- Always have water and possibly other beverages.

23.7.2.2 Number of full lifecycle iterations

Just as a loose rule of thumb from our experience, an ideal number of full UX engineering cycle iterations per version or release is about three, but resource constraints often limit it to fewer. In many projects, you can expect only one iteration. Of course, any iterations are better than none.

23.7.3 Informed Consent

Informed consent is formal and signed permission is given to UX professionals by usage research and evaluation participants to use the data gathered within the UX lifecycle activities, usually with certain stipulated limits.

When we collect empirical data involving human subjects, we have certain legal and ethical responsibilities, even though there is very little risk of a participant being harmed in UX evaluation.

We still have professional obligations, which center on the informed consent form, a document to establish explicitly the rights of your participants and which also serves as legal protection for you and your organization. Therefore, you should always have all participants, anyone from who you collect data of any kind, sign an informed consent form.

23.7.3.1 Informed consent permission application

Your preparation for informed consent begins with an application to your institutional review board (IRB), an official group within your organization responsible for the legal and ethical aspects of informed consent. The evaluator or project manager should prepare an IRB application that typically will include:

- Summary of the evaluation plan.
- Statement of complete evaluation protocol.
- Statement of exactly how human subjects will be involved.
- Your written subject/participant instructions.
- A copy of your informed consent form.
- Any other standard IRB forms for your organization.

Because most UX evaluation does not put participants at risk, the applications are usually approved without question. The details of the approval process vary by organization, but it can take up to weeks and can require changes in the documents. The approval process is based on a review of the ethical and legal issues, not the quality of the proposed evaluation plan.

23.7.3.2 Informed consent form

The informed consent form, an important part of your IRB application and an important part of your empirical UX evaluation, is a requirement; it is not optional. The informed consent form, which is to be read and signed by each participant, should state in clear understandable language:

- That the participant is volunteering to participate in your evaluation.
- The expected length of time for the evaluation session (the evaluator should have some idea of how long a session will take after performing pilot testing).
- That the participant can withdraw anytime, for any reason, or for no reason at all.
- That you are taking data that the participant helps generate.
- That the data are taken anonymously (neither the name of the participant nor any other kind of identification will be associated with data after it has been collected).
- That the participant understands any foreseeable risks or discomforts, which should be minimal to zero for UX evaluation.
- That the participant understands any benefits (e.g., educational benefit or just the satisfaction of helping make a good design) and/or compensation to participants (if there is payment, state exactly how much; if not, say so explicitly).
- All project/evaluator contact information.
- That they can ask the evaluator questions at any time.
- Whether any kind of recording (e.g., video, audio, photographic, or holodeck) involving the participant will be made and how you intend to use it, who will view it (and not), and by what date it will be erased or otherwise destroyed.
- A statement that, if you want to use a video clip (for example) from the recording for any other purpose, you will get their additional approval in writing.

The consent form may also include nondisclosure requirements. This form must spell out participant rights and what you expect the participants to do, even if there is overlap with the general instruction sheet. The form they sign must be self-standing and must tell the whole story.

Although informed consent may not be required in the case where your participants are also organization employees, this is an area where you should err on the side of caution. In any case, you should have two copies of the consent

form ready for reading and signing by participants when they arrive. One copy is for the participant to keep.

Example: Simple Informed Consent Form

Informed Consent for Participant of Development Project

<Name of your development organization> <Date or version number of form> Title of Project: <Project title>

Project team member(s) directly involved: <Team member names> Project manager: <Project manager name>

I. THE PURPOSE OF YOUR PARTICIPATION IN THIS PROJECT As part of the <project title> project, you are invited to participate in evaluating and improving various designs of <name of system or product>, <description of system or product>.

II. PROCEDURES You will be asked to perform a set of tasks using the <name of system or product>. These tasks consist of <description of range of tasks>. Your role in these tests is to help us evaluate the designs. We are not evaluating you or your performance in any way. As you perform various tasks with the system, your actions and comments will be noted and you will be asked to describe verbally your learning process. You may be asked questions during and after the evaluation in order to clarify our understanding of your evaluation. You may also be asked to fill out a questionnaire relating to your usage of the system.

The evaluation session will last no more than four hours, with the typical session being about two hours. The tasks are not very tiring, but you are welcome to take rest breaks as needed. If you prefer, the session may be divided into two shorter sessions.

III. RISKS There are no known risks to the participants of this study.

IV. BENEFITS OF THIS PROJECT Your participation in this project will provide information that may be used to improve our designs for <name of system or product>. No guarantee of further benefits has been made to encourage you to participate (Change this if a benefit such as a payment or a gift is offered). You are requested to refrain from discussing the evaluation with other people who might be in the candidate pool from which other participants might be drawn.

V. EXTENT OF ANONYMITY AND CONFIDENTIALITY The results of this study will be kept strictly confidential. Your written consent is required for the researchers to release any data identified with you as an individual to anyone other than personnel working on the project. The information you provide will have your name removed and only a subject number will identify you during analyses and any written reports of the research.

The session may be recorded. If it is recorded, the recordings will be stored securely, viewed only by the project team members and erased after three months. If the project team members wish to use a portion of your recording for any other purpose, they will get your written permission before using it. Your signature on this form does not give them permission to show your recording to anyone else.

VI. COMPENSATION Your participation is voluntary and unpaid (Change this if a benefit such as a payment or a gift is offered).

VII. FREEDOM TO WITHDRAW You are free to withdraw from this study at any time for any reason.

VIII. APPROVAL OF RESEARCH This research has been approved, as required, by the Institutional Review Board <or the name of your review committee> for projects involving human subjects at <your organization>.

IX. PARTICIPANT RESPONSIBILITIES AND PERMISSION I voluntarily agree to participate in this study, and I know of no reason I cannot participate. I have read and understand the informed consent and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent for participation in this project. If I participate, I may withdraw at any time without penalty. I agree to abide by the rules of this project.

Signature Date

Name (please print) Contact: phone or email

23.7.4 Other Paperwork

User instructions

Overview. In conjunction with developing evaluation procedures, you, as the evaluator, should write *introductory instructional remarks* that will be read uniformly by each participant at the beginning of the session. All participants thereby start with the same level of knowledge about the system and the tasks they are to perform. This uniform instruction for each participant will help ensure consistency across the test sessions.

Specifics. These introductory instructions should explain briefly the purpose of the evaluation and tell a little bit about the system the participant will be using as well as describe what the participant will be expected to do and the procedure to be followed by the participant. For example, instructions might state that a participant will be:

- Asked to perform some benchmark tasks that will be given by the evaluator.
- Allowed to use the system freely for a while.
- Given some more benchmark tasks to perform.
- Asked to complete an exit questionnaire.

Make it clear that you are not evaluating the participant. In your general instructions to participants, make it clear that the purpose of the session is to evaluate the system, not to evaluate them. You should say explicitly *"You are helping us evaluate the system—we are not evaluating you!"* Some participants may be fearful that somehow their performance might not be up to "expectations" or that participation in this kind of test session could reflect poorly on them or even be used in their employment performance evaluations (if, for example, they work for the same organization that is designing the interface they are helping evaluate). They should be reassured that this is not the case. This is where it is important for you to reiterate your guarantee of confidentiality with respect to individual information and anonymity of data.

Prepare the participant for thinking aloud. The instructions may inform participants that you want them to think aloud while working. Explain what this is and how to do it and offer a very brief trial run for learning.

Print out and copy the general instructions so that you can give one to each participant.

23.7.4.1 Nondisclosure agreements (NDAs)

Sometimes an NDA is required by the developer or customer organizations to protect the intellectual property contained in the design. If you have an NDA, print out copies for reading, signing, and sharing with the participant.

23.7.4.2 Questionnaires and surveys

If your evaluation plan includes administration of one or more participant questionnaires, make sure that you have a good supply available. It is best to keep blank questionnaires in the control room or away from where a newly arriving participant could read them in advance.

23.7.4.3 Data collection forms

If appropriate, make up a simple data collection form in advance. Your data collection form(s) should contain fields suitable for all types of quantitative data you collect and, probably separate, data collection forms for recording critical incidents and UX problems observed during the sessions. The latter should include spaces for the kind of supplementary data you like to keep, including associated tasks, effect on user (e.g., minor or task-blocking), guidelines involved, potential cause of problems in design, relevant designer knowledge (e.g., how it was supposed to work), etc. Keep your data collection forms simple and easy to use on the fly. Consider a spreadsheet form on a laptop.

23.7.5 Training Materials

Use training materials for participants only if you anticipate that a user's manual, quick reference cards, or any sort of training material will be available and needed by users of the final system.

23.7.6 The UX Evaluation Session Work Package

To summarize, as you do the evaluation preparation and planning described in this chapter, you need to gather your evaluation session work package, all the materials you will need in each evaluation session.

Examples of package contents can include:

- The evaluation configuration plan, including diagrams of rooms, equipment, and people in evaluation roles.
- General instruction sheets.
- Informed consent forms, with participant names and date entered.
- Any nondisclosure agreements.
- All questionnaires and surveys, including any demographic survey.
- All printed benchmark task descriptions, one task per sheet of paper (Section 22.6.4.1).
- All printed unmeasured task descriptions (these can be listed several to a page).
- Any special instructions to watch out for particular parts of the design, evaluation scripts, things to do before each participant session (e.g., to reset browser caches so that no auto complete entries from previous participant's session interfere with the current session), etc.
- For each evaluator, a printout (or laptop version) of the UX targets associated with the day's sessions.
- All data collection forms, on paper or on laptops.
- Any props needed to support tasks.
- Any training materials to be used as part of the evaluation.
- Any compensation to be given out (e.g., money, gift cards, T-shirts, coffee mugs, used cars).

Exercise 23-1: Empirical UX Evaluation Preparation for Your System

Goal: To get some practice in preparation for a simple empirical evaluation.
Activities: If you are working with a team, get together with your team.

Decide roles for team members. Include at least a facilitator and a prototype executor, plus a quantitative data recorder and one or more critical incident recorders.

In addition, if you are doing this exercise in a classroom with other teams, assign two team members as participants to trade to another team when you start data collection in the next exercise.

The prototype executor should get out the wireframe prototype deck you made in a previous exercise and become familiar with the navigation.

This activity works well for a team of about four. If you have more or fewer members in your team, it is easy to make adjustments. If there are only two of you, for example, one person can be the executor and the other person can record critical incidents and time the benchmark tasks. If there are four or five of you, the extra people will be valuable in helping record critical incidents. If you have been working alone on all the previous exercises, you may want find a couple of other people to help you run the evaluation. In addition and in any case, you need to recruit two people to serve as participants to evaluate your prototype.

Get out the UX target table you made in a previous exercise.

Have at least two benchmark tasks that you created in a previous exercise, each written on a separate piece of paper.

Assuming you used a questionnaire for subjective data in your evaluation session, get out copies of the questionnaire, one for each participant you will be using, and circle the questions you want participants to answer.

Review your evaluation protocols.

Deliverables: Just have everything ready for the next exercise, data collection.
Schedule: It should not take too long to get ready for evaluation.

Subjective UX Evaluation Data

Data based on opinion or judgment, of evaluator or user (Section 21.1.4.2).

23.7.7 Do Final Pilot Testing: Fix Your Wobbly Wheels

If your UX evaluation plan involves using a prototype, give it a final shakedown to be sure it is robust enough to support evaluation without breaking. This step really applies to any level of fidelity if your prototype will be seen by people beyond your UX team.

In addition to shaking down your prototype, think of your pilot testing as a dress rehearsal to be sure of your lab equipment, benchmark tasks, procedures, and personnel roles.

You don't want to "use up" a user participant by getting them started only to discover the prototype dies and prevents benchmark task performance.

Simulate user experience evaluation conditions by having one member of your team "execute" the prototype while another member plays "user" and tries out all benchmark tasks. The user person should go through each task in as many ways as anyone thinks possible to head off unexpected problems. Don't assume error-free performance by your users; try to have appropriate error messages where user errors might occur.