Reprint of paper appearing in *IEEE Transactions on Professional Communication*, Volume 37, Number 3, September 1994.

Techniques for Managing a Usability Test

Laurie Kantner

Abstract—Investing time, energy, and money in a usability test pays off when the data you collect answer your questions. Who makes sure usability tests meet their information-gathering goals? The project manager, who has to be ready to solve the many problems that will inevitably arise. This paper assumes the reader has taken a course on usability methods or has conducted a usability test with the assistance of a professional in the field.

"USABILITY is in the details." Usability testing demands both creative innovation and meticulous attention to details. Only "after the fact" does the critical-path nature of some details become obvious. Managers need to give these details the attention they require.

Usability test teams provide a new working context for the individual team members, who are generally from different groups or departments: development or engineering, human factors, documentation, quality control, technical support, customer sales/support, and current internal users. As with any collaborative endeavor, usability tests can run fairly successfully by themselves when they target agreed-on goals and follow an agreed-on process. But as is often the case with team projects, not everyone has the same goals or the same understanding of the process and schedule to follow; and when changes occur, expectations often differ about what alternatives to pursue.

Details overlooked during preparation for a usability test become gaping holes in the data collected during the actual test. With project management formally assigned to a designated team member, someone becomes responsible for paying attention to these details for fostering team communication and agreement on the test goals, processes, and schedule; and for revising goals when compromises become necessary.

The goal of this paper is to provide the project manager of a usability test with tools for managing the details and fostering teamwork. The guidelines presented here apply to all types of usability projects, from laboratory tests to group interviews to field studies, and to many industries. Specific examples generally assume individual laboratory tests of computer software products and documentation.

SKILLS BROUGHT TO THE TASK

The person best qualified to manage a usability test most likely wears multiple hats in the organization. Within a computer software manufacturing company, this person may be a usability engineer, quality-assurance specialist, user-interface designer, or technical communicator. Recognizing that there is more to usability testing than "crunching numbers," the ideal project manager:

- Identifies strongly as a user advocate.
- Possesses creative problem-solving ability.
- Provides an objective, sensitive view when facilitating goal-setting among individuals with different responsibilities and backgrounds.
- Understands the importance of scheduling tasks early and rescheduling diligently.
- Knows the value of keeping written track of the test plan, progress, and decisions.
- Understands the scientific method.
- Can keep the ultimate goals of the test in focus even when immersed in situational constraints.

Ideally, the project manager has played a key role throughout the product life cycle, helping define past studies for earlier product versions if not actually managing those studies directly. The project manager understands and promotes usability testing as an iterative process, keeping the purpose of the current test in focus by reminding team members of the results of earlier studies and the goals of future studies.

AXIOMS FOR COLLECTING THE RIGHT DATA

There are many ways to collect the wrong data during a usability test. Even when the test initially focuses on high-priority, specific questions, its definition can easily change during the preparation stage, increasing the likelihood of ambiguous and unusable results. Because preparation often constitutes the first half to three-fourths of the entire project, there are numerous opportunities for the definition to blur before any actual data collection begins.

The project manager is the gatekeeper, overseeing what goes into the test definition and what is taken out, and therefore must be prepared to explain the consequences of every suggested change in the definition. Some axioms that help ensure usable data at the end of the test are the following:

- Recruit participants who truly match the target audience for the product. If the participants don't match, the results won't reflect the experiences you can expect from your "real-world" audience [1, 2].
- Design tasks that match expected—and natural—use of the product. Only realistic tasks will enable identification of potential problems in actual use situations.
- Minimize variables and activities that do not generate (or might even mask) meaningful results. Identify how every activity addresses a major question the test is to answer [3].
- Make sessions as consistent as possible for each participant. Balance the order of presentation where you are deliberately varying participants' experiences. Extraneous data complicates analysis and can dilute the usefulness of the results.
- Test the test. Before conducting the actual sessions, conduct one or two "dry-run" sessions and pilot-test sessions to verify the flow of the tasks, the clarity of the administrator's language, the operation of the product, and the length of the session. The dry run identifies major flaws; the pilot test identifies where fine-tuning is needed.

Even when these axioms are well understood during the initial test definition and receive full support from the project team, two pressure points during the preparation stage lead to axiom amnesia:

- "While we have them here" tempts team members to add more test activities to take advantage of the investment in recruiting participants.
- "We're running out of time before next week's sessions" causes three types of compromises: (1) features to be tested are not available for the dry-run or pilot test, reducing the usefulness of those pre-test verification steps; (2) features to be tested cannot be completed in time and must be removed to keep the test on schedule; and (3) participant recruiters are tempted to relax the screening requirements because finding appropriate participants is taking longer than expected.

The project manager must keep waving the axiom flag, asking questions such as these:

- "If we study this additional feature, what will we do with the information we learn?"
- "Will the added activity weaken the data for one of our primary goals?"
- "If we remove this activity, how will what we don't learn hurt us? What is the trade-off between that and rescheduling the test two weeks later?"
- "Will using less-qualified participants still give us the data we want?"

These axioms govern the project manager's responsibility throughout all seven phases of a usability test:

- Planning the test.
- Designing the test activities. Recruiting participants.
- Preparing the test materials.
- Setting up the test environment. Conducting the test.
- Compiling the test results.

PLANNING THE TEST

The end product of planning the test is a document describing the test goals and methodology, participant- selection requirements, working procedure and schedule, and resource requirements. To allow enough time for participant recruiting, a draft of the participant selection questionnaire / script for screening candidates should also be ready as part of the initial test plan.

The project manager is the logical person to produce the test plan, based on input from others during initial planning discussions. The planning process should elicit agreement on the following information:

- Major questions the test should answer, with assigned priorities.
- Number of participants and desired characteristics.
- Completeness, maturity, or fidelity of the product (and documentation) for the test. For a software product: a paper prototype or on-line demonstration? For software documentation: handouts, a manual, or on-line support?
- Resources available to plan and conduct the test, and the desired schedule (first cut).
- Participant compensation available. Be ready to offer an honorarium if gifts or souvenirs do not attract candidates, and decide on the amount. Check with your company's legal department for record-keeping requirements if you offer an honorarium.

Defining the scope of a usability test is usually iterative. You begin by listing everything everyone would like to know, then evaluating the resources required to design and conduct a test of that magnitude, paring lower-priority items from the list, reevaluating resource requirements, and continuing this process until the right combination of content and feasibility is reached. Several rules of thumb help to shorten these iterations:

- Participant sessions can vary in length from half an hour to half a day. The primary consideration is the length of time required to analyze the data and make recommendations, generally two to four times as long as the time spent collecting it. In addition, long sessions may not be feasible for many potential participants, lessening the likelihood of recruiting a representative cross-section.
- A session longer than two hours requires that you add time for taking a formal break. In addition, 15 to 30 minutes of session time may be needed for participants to fill out an initial background questionnaire and for post-session debriefing.
- Base the session length on the least-experienced or most methodical participant. In laboratory testing, sessions running over their allotted time may need to be compressed or truncated, introducing more variability into the test results. Remember that the dry run, often with an internal, experienced participant, may be completed in one-third to one-half the time that a test session with an external, inexperienced participant will require.
- For studies designed to detect serious problems with a product or document, about 6 participants per "cell" or audience group is usually adequate [4]. If the goal is to generalize the results over a broad range of uses and users, consult with a statistician to determine an appropriate number of participants.

The table below suggests ranges of hours to allow for each test phase, to help the project manager determine the resources and calendar time required to prepare for and conduct a usability test.

Table I. Working Time Requirements for Usability Test Phases.

Project Phase	Working Time
Design	20 to 40 hours
Participant recruiting	2 to 4 hours per final participant
Materials preparation	30 to 80 hours
Session administration	Number of participants times the session length, plus time for dry-runs, pilot-testing, breaks between sessions, and modification to materials
Results reporting	50 to 150 hours (depends on study complexity, number of participants, session length, and level of detail in the final report)

Recruiting 8 to 10 participants will probably take 20 to 40 hours; however, you need to allocate two to three times as much calendar time to allow for "down time" spent awaiting returned calls and collecting more names of candidates.

Figure 1 shows a typical schedule for a laboratory test with 8 participants, 1.5-hour test sessions, and a detailed final report. When developing your schedule, assume that the date by which developers promise a testable product may slip another week or two. Make sure your plan accommodates this likelihood yet still delivers the test results on the originally specified date, since that date is unlikely to move. Also, postponing test sessions is time-consuming and increases participant cancellations.

Note that a usability test that must be completed in *less* than 8 to 12 *weeks* may mean using fewer participants, testing only a small part of the product, and providing a limited analysis (and final report) [5]. However, a smaller-scale test may be appropriate in two cases: to demonstrate the value (albeit limited) of usability testing to uninformed management; and to test mature products undergoing limited changes from one release to another.

Designing the Test Activities

It's tempting to go from the test plan directly to preparing the materials. However, a detailed test design aligns specific expectations about the test among all concerned parties, before the time-consuming activity of preparing the materials begins.

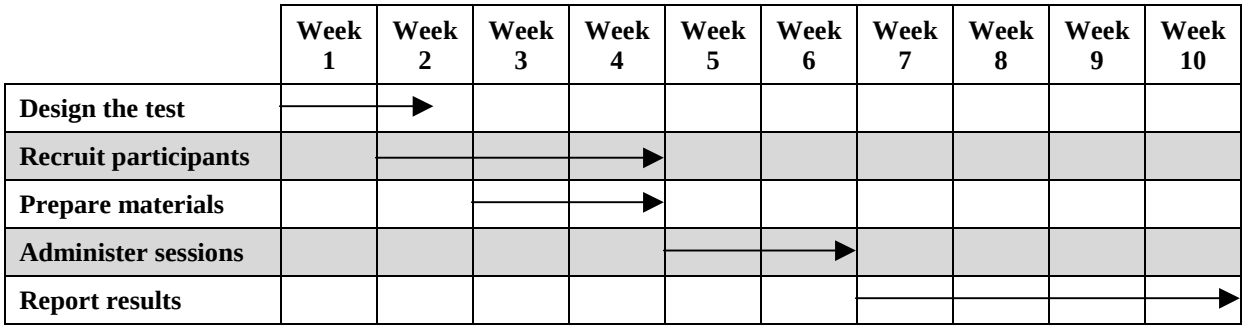

Figure 1. Typical schedule for a laboratory usability test.

The end product of the design stage is a document listing the organization and order of tasks, detailed activity descriptions, relevant issues / questions each task or activity is to address, and estimated timings for each task or activity. If the initial timings result in a session that is too long, the test designer uses the test goals and priorities to eliminate activities. The initial test design should conservatively allow for about 25% growth in length as a result of the review process.

In addition to describing activities the participants will perform, the design lists assumptions about participant skills, as well as product functionality and documentation elements to be tested. The more detailed these descriptions, the earlier you can identify any discrepancies between test activities and skills and resources to complete them.

Team members carefully review the test design for how it addresses test goals and its assumptions about participant skills and product / documentation readiness. The project manager facilitates the review discussion and negotiates differing priorities when team members suggest design changes.

At the same time, team members also approve the final participant-screening questionnaire so that screening can start right away. That approval should include agreement on the priorities of the different participant characteristics, so that any compromises made quickly during screening are consistent with test goals.

At this point, the test team's ability to produce all required elements to meet the originally defined schedule becomes clear. If necessary, the project manager revises the test schedule to reflect changes.

Although it provides more concrete details about the exact activities during the test, the design is still a concept document, written without benefit of the product or documentation to be studied. Thus, it is only as "accurate" as the development team's stated intentions and the test designer's imagination; expect significant factual changes during materials preparation.

RECRUITING PARTICIPANTS

As a parallel activity to the test design and materials development, recruiting participants requires adequate time and attention. Writing a detailed recruiting script and questionnaire enables supporting staff to conduct this activity, freeing the test designer to concentrate on test activities.

The recruiting script provides the sequence of qualifying questions the recruiter asks prospective candidates, designed to minimize the time required of both the candidate and the recruiter. The script contains notes to the recruiter indicating where an answer disqualifies the candidate from participating, at which point the script prompts the recruiter to thank the candidate and ask for referrals. It also provides standard language for describing the test, so that all participants hold roughly the same expectations.

To ensure that a consistent methodology is used to screen candidates, preferably only one person should conduct all recruiting. The recruiter should be outgoing, persuasive, and sufficiently knowledgeable about the test to set candidates at ease. The recruiter should test the script on coworkers to ensure that it reads smoothly and naturally before using it on potential participants.

People's motivations to participate vary, and so does their level of commitment when the actual day of their session arrives. Plan for inevitable drop-outs and "no shows" by recruiting 10-20% more people than you actually need. Inform backup participants of the importance of their flexibility (in being willing to be scheduled at the last minute), and be prepared to offer them the same compensation as firmly scheduled participants.

Expect the recruiter to have to make 10-15 telephone calls for each final participant recruited. Many initial calls are to persons within customer organizations who can refer the recruiter to an appropriate candidate; the script needs to provide language for these transactions. The recruiter should make sure each telephone call results in either a qualified candidate or names of other persons to contact, preferably both.

The results of recruiting are:

- Recruited participants (including extras).
- A questionnaire filled in by the recruiter for all candidates who pass the initial qualification stage.
- A log of all contacts made, organized by candidate and by status of recruiting ("accepted," "pending," "rejected," "rejected but possible future candidate"). The contact log should contain complete contact information for candidates, including a fax number for instant confirmation.
- A summary of the characteristics met by the recruited participants. Team members should review the list of recruited participants to identify any qualification problems before final confirmation packages are sent.
- A short description of the usability test for those candidates who need to present something in writing to receive their manager's approval to participate.

Regular contact and final confirmation are key ingredients for letting participants know how important their participation is. The purpose of the first telephone call is to determine whether a person qualifies; a second telephone call confirms their selection and clarifies convenient and inconvenient times for scheduling their test; and a third call confirms the place and time. These phone calls should be prompt to avoid leaving candidates wondering about their status. If the exact schedule of the test sessions is uncertain, more phone calls are required to set the stage and then make final confirmation.

Schedule the sessions to allow some break time between participants, and try to avoid more than six hours of sessions per day. The team members' fatigue can become a factor in the test results. However, be prepared to offer evening (and perhaps weekend) sessions to accommodate participants who cannot make time during the working day.

At least three working days before sessions begin, participants should receive a confirmation package containing a letter thanking them for agreeing to participate, listing the location and time of their session, providing a map to the location, and describing their compensation (gift, souvenir, honorarium).

In recruiting dry-run and pilot-test candidates, qualifications are generally somewhat less stringent. A dry-run participant need not have all the characteristics of the desired participants (most tests use internal participants-employees-for dry runs). The pilot-test participant should be an external candidate who would qualify as a real participant except in perhaps one characteristic.

PREPARING THE TEST MATERIALS

The test materials govern all activity during the test sessions. They contain several types of information:

- A script for the administrator. Using a script ensures a consistent presentation from one participant to the next and minimizes the effects of fatigue as the day progresses. Logistics notes for the administrator.
- Logistics notes help the administrator keep track of details about the environment or the order of activities to ensure consistency.
- Convenient spaces for note-taking (such as questions with check-boxes or lists of key issues to observe [3]). Note-taking forms make it easy to record data for quick compilation to report initial results after the test. This technique also ensures that key issues are noted consistently, making it possible to analyze trends without viewing videotapes.
- Removable questionnaires (if used) for the administrator to hand the participant. Questionnaires elicit participant opinions, an important element of most tests.
- (If the test includes protocol analysis) a handout or videotape instructing participants on how to "think aloud."
- Other handouts for participants, providing task instructions, pictures of objects to identify, or lists of items to rank or rate. For some studies, participants simply receive blank paper for providing requested information.

The detailed task descriptions from the test design serve as the test designer's outline for writing the test materials. Often one draft, reviewed by team members, and one revision cycle are sufficient to prepare a draft for the dry run. (The initial revisions often catch awkward script language as well as obvious logistical and procedural errors.)

In addition to the materials used during the sessions, the following administrative materials are recommended:

- Participant background questionnaire. This questionnaire asks for the same information collected during oral screening, to verify participant characteristics. It can also collect other information of interest.
- Nondisclosure forms and videotape release forms—check with your legal department about these. If you will be videotaping participants' faces, the videotape release form is more complex. Plan on at least two more cycles of revisions to the materials-after the dry run, and after each pilot test-before actual sessions begin, and space the dry-run and pilot-test sessions to allow for this revision time.

SETTING UP THE TEST ENVIRONMENT

Whether a specially equipped usability lab or an office or conference room converted for temporary use, the environment should be customized for the usability test sessions. In addition to the necessary equipment, power connections, and telephone access, the environment should provide a comfortable working space for the participant. Here are some guidelines.

- Allow sufficient working surface for any documents or paperwork the participant uses during the session.
- Consider a separate room or area in which the participant fills in forms and relaxes, with enough space to hold refreshments.
- Furnish the space with high-quality furniture and decor to resemble the participant's actual work environment [6].

Create lists of what to supply for each session, such as sufficient videotapes and audiotapes, as well as checklists of how to reset the equipment and environment between participants for smooth, error-free transitions. Ensure that sufficient copies of all documents are available: one set of test materials per participant for the administrator and one set for each observer; copies of background questionnaires, nondisclosure forms, and videotape release forms; documentation for reference during the session; and any other documents that are part of the session.

Make sure the right people are planning to be available to assist throughout the sessions, starting with yourself: clear your calendar, and tell your coworkers and colleagues that you will be unavailable during the test sessions. If a colleague will be assisting you with observing and note-taking, get that person's time commitment in advance. Make sure a technical resource person is available throughout the sessions in case of product difficulties. Make sure someone, preferably the original screener, is responsible for scheduling backup participants in case of no-shows.

Participant stress can affect how well the test data will generalize to actual users [6]. Hospitality arrangements are an important component of reducing participant stress. Be a good host: make sure someone is there to greet participants on arrival, offer them a variety of refreshments, and (when the session is finished) escort them out.

Prevent interruptions during sessions by posting "Usability Session in Progress" signs and disabling the telephone in the usability test room. Circulate a session schedule in advance to encourage observers, and label the observation room. The observation room should provide sufficient space to accommodate the expected number of observers.

CONDUCTING THE TEST

Dry-run sessions and pilot-test sessions are often conducted a few days before actual testing begins. As mentioned earlier, these pre-tests reduce the number of bugs, help verify test session length, and identify discrepancies between test activities, product functionality, and documentation. The pre-tests also enable the administrator to practice speaking from the

script, becoming more comfortable with the language and modifying it as needed.

In testing software products, developers usually make last-minute changes to the product after each dry-run and pilot-test session. The session administrator should personally run through the task sequences again before sessions with external participants start, to make sure these software changes do not have an unexpected effect on the test.

Once real test sessions begin, the time for changes has ended. Participant experiences must be extremely consistent from one session to the next, to provide a common foundation for the data collected.

Several administrative practices can ensure usable data from each test session, as well as ensure that a sufficient number of test sessions take place:

- Obtain the participants' informed consent to participate by having them sign a form stating what will happen during the usability test [5]. Make sure they know they can withdraw from the test at any time.
- After each participant fills in the background questionnaire, compare that information with the information initially received during oral screening of participants. If you find discrepancies, decide whether they will invalidate the data you will collect. Generally they will simply change your expectations. For example, a participant who reported more computer software experience during screening than on the background questionnaire will shift to your "inexperienced" category but will still provide valuable data.
- Plan on at least one "no-show" for every 5 or 6 participants scheduled. If the number of no-shows leaves too few participants for the desired amount of data, it may be necessary to extend the sessions another day or two to schedule backup participants.
- In testing software products, prototype or preliminary software can often crash during the test activities, so try to learn the proper procedure for bringing the software back up without losing valuable data. Have technical staff available who can solve problems quickly, to avoid eliminating test activities to stay within the time limit, as well as to minimize the participants' anxiety.

As sessions are completed, recorded data for each participant accumulates: a filled-in background questionnaire, signed forms (if needed), one or more sets of annotated test materials reflecting administrator and observer notes during each session, one or more labeled videotapes (if used) and audiotapes, and perhaps computer-generated data. After each session, put all paper materials and tapes for each participant in a separate, labeled, clasped envelope, and store electronic data in two places for backup.

Between sessions, make sure all materials in each envelope are labeled with the participant's name, in case they become separated. Avoid removing more than one participant's materials from an envelope at a time.

Developers or marketing staff frequently ask for additional changes to the test design after observing the first few sessions.

Assess the value of having a smaller set of participant data for these changes versus having a larger set of participant data for the initially agreed-on activities.

COMPILING THE TEST RESULTS

If you have promised "oral quick results" to the product developers, introduce them with ample caveats about their reliability if you've had to work quickly to generate them. Divide the results into short-term issues and long-term issues. Always follow up with a carefully prepared written report, however brief, to record the results and correct any misstatements. The decision of who analyzes the results, presents them, and writes the report depends on team member skills and availability.

An increasingly popular method of reporting results is to create a "highlights videotape." The time investment depends on the presentation method chosen-a compendium of revealing experiences, organized by participant, is faster to create than a tape organized by issue with participant clips as supporting evidence.

Regardless of the method chosen, observers can flag interesting participant actions and comments on their session notes to expedite videotape preparation.

Creating a detailed test report to convey final results can increase the investment in a usability project by another 20-25% over the investment in test planning, preparation, and administration, but it provides a level of information not otherwise available. The final report, which often takes three to four weeks to complete, provides a comprehensive record of what happened during the test. It explains the reasons for specific elements of the test design and describes participant activity in more depth, providing details that may be difficult to reconstruct from a quick-results report and people's memories as time goes on. It also makes possible the analysis of design decisions and usability test results over a product's life cycle.

A final test report that includes key quotations from participants can often influence product-development decisions more readily than tables of summary data. (Developers who watch all of the sessions would hear the participant comments, but developers rarely watch more than one or two sessions.) In addition, the time you spend with the data while preparing the final report gives you a chance to discover patterns not immediately discernible initially. These issues are often secondary, but they might suggest areas for further exploration.

CONCLUSION

The nature of usability tests-with tight schedules, moving targets, and human interactions-invites the unexpected. The focus of this paper has been to help usability specialists identify where the unexpected is likely to occur, and plan alternative ways to cope with these changes. Designating project management as a separate responsibility increases the likelihood that usability test goals will be met and usability tests will become an integral part of product development.

REFERENCES

[1] S. Rosenbaum, "Selecting the appropriate subjects for documentation usability testing." *Proceedings of the Third International Conference on Human-Computer Interaction*, Boston, MA, 1989, vol. I.

[2] S. Rosenbaum, "Chapter 3: Selecting appropriate participants for documentation usability testing." *Practical Approaches to Usability Testing for Technical Documentation*, " Chris Velotta (Ed.). Society for Technical Communication, in press.

[3] J. Ramey, "Usability testing of the documentation and user interface of computer systems," seminar sponsored by the Dept. of Communication, College of Engineering, University of Washington, March 1994.

[4] R. Virzi, "Refining the test phase of usability evaluation: how many subjects is enough?", *Human Factors*, vol. 34, no. 4, August 1992.

[5] J. S. Dumas, and J. C. Redish, *A Practical Guide to Usability Testing*, Norwood, NJ: Ablex Publishing, 1993.

[6] J. R. Schrier, "Reducing stress associated with participating in a usability test," *Proc. Human Factors and Ergonomics Society 36th Annual Meeting*. 1992, pp. 1210-1214.

Laurie Kantner is a technical writer and usability specialist with Tec-Ed, Inc., in Ann Arbor, MI. She manages usability projects for computer-based products and documentation, from laboratory testing to group interviews to field studies.