

# Not Just a Hammer: When and How to Employ Multiple Methods in Usability Programs

**Stephanie Rosenbaum**  
Tec-Ed, Inc.  
stephanie@teced.com

Most usability practitioners today are skilled at conducting usability testing, especially laboratory testing. And most of us understand the benefits of conducting iterative usability studies as part of iterative product design and development. However, exploratory usability testing is so frequently performed that many organizations fail to take advantage of other usability methods. Usability engineering is in peril of being regarded as a usability testing “commodity,” rather than a consulting specialty. Instead, we should assess the trade-offs among different methods, so that our organizations can make informed decisions about where, when, and how to spend usability resources. This paper explores how to combine different usability methods, depending on:

- Where you are in the product development cycle
- What questions you want to answer
- Which audiences you want to study
- Which usage scenarios are of special interest, for what reasons

## DEFINITIONS OF METHODS

Although most practicing usability specialists are familiar with several usability methods, our approaches and methodologies may vary. Here are the definitions I used when planning this paper.

### Laboratory Testing

Laboratory testing is the classic usability method, both used most frequently and rated most effective in surveys of usability practitioners (1). Dumas and Redish define usability testing as sharing these five characteristics (2):

1. The primary goal is to improve the usability of a product. For each test, you also have more specific goals and concerns that you articulate when planning the test.
2. The participants represent real users.
3. The participants do real tasks.
4. You observe and record what participants do and say.
5. You analyze the data, diagnose the real problems, and recommend changes to fix those problems.

In the surveys cited above, the authors attempted to distinguish “laboratory testing” from usability testing with a portable or no lab equipment. In this paper, laboratory testing does not necessarily assume lab equipment. Rather, “laboratory” applies to the *controlled experimental method*, in which all participants perform the same tasks under as similar conditions as possible (3).

### Contextual Inquiry

Contextual inquiries are a field study method initially developed by the Software Human Factors group at Digital Equipment Corporation (4)(5). In a contextual inquiry, an experienced interviewer observes users in the context of their actual work situation, performing their usual job tasks (not tasks designed for the study). According to Raven and Flanders, contextual inquiry is based on the following three principles:

1. Data gathering must take place in the context of the users’ work.
2. The data gatherer and the user form a partnership to explore issues together.
3. The inquiry is based on a focus; that is, it is based on a clearly defined set of concerns, rather than on a list of specific questions (as in a survey).

Conducting a contextual inquiry normally involves a team of two, an inquirer and a note-taker/observer. The inquirer and the participant are equals; each is an expert in his or her own work. After the visits, the inquiry team reviews their notes and analyzes the information to find patterns, often using affinity diagrams. Contextual inquiries yield rich data from seeing users in their real work context, and thus can identify issues not previously recognized. They avoid the possibility of misleading results from placing users in artificial situations. And it's easier to show that we're evaluating the product, not the user.

### **Ethnographic Interviews**

Ethnographic interviews use some of the techniques of anthropology to collect more specific and concrete information from participants than normally takes place in traditional journalistic interviews. By interviewing people in their own work environments, even when circumstances don't permit a full contextual inquiry, we can examine the users' workplaces and study artifacts associated with their work processes (such as notes, memos, and printouts).

The top-down interviewing strategy of ethnographic interviews is more structured than contextual inquiry. Wood describes and illustrates several categories of effective questions (6). He notes that experts tend to translate their knowledge into terms they believe the interviewer will find easier to understand; to elicit users' natural descriptions, it's better to ask *how* terms, tools, and concepts are used, rather than *what* they are or mean.

### **Heuristic Evaluation**

Heuristic evaluations are expert evaluations of products or systems. They're conducted by usability specialists, domain experts, or—ideally—by “double experts” with both usability and domain experience (7). Heuristic evaluations by two or more usability specialists can identify a majority of the usability problems in a web site or other product, with the problem-identification percentage increasing as you add evaluators. Two evaluators can identify over 50% of the problems, and three can identify about 60%. The curve flattens after five evaluators; it would take 15 evaluators to identify 90% of usability problems.

### **Usability Focus Groups**

Usability focus groups apply a method that originated in market research to obtain qualitative information from target users. In a focus group, people with similar characteristics who don't know one another meet in a group environment, where they discuss a selected topic with the assistance of a moderator (8).

The goal of a focus group is to have its members interact and spark ideas from one another. Successful focus groups produce a variety of ideas, not a consensus of opinion; the purpose of a focus group is not to make decisions but to collect information. Usability focus groups are task-based and add user tasks and activities to the traditional discussions. However, because of the group environment where the tasks are performed in teams, it's not often possible to collect individual performance data.

## **POPULAR MODELS FOR ITERATIVE USABILITY PROGRAMS**

Most iterative usability programs today follow one or both of two popular models:

- Exploratory usability testing with two to four participants after each of several iterative development cycles
- Heuristic evaluation, followed by design revisions, followed by usability testing

These models are popular for a good reason—they work! Assuming that your primary goal is to *identify* usability problems rather than quantify them, and assuming that you're not trying to compare two interfaces explicitly, iterative exploratory usability testing is very successful (9)(10):

- 8 cycles of testing with 1 or 2 participants increased user accuracy by 20%
- Maximum benefit-cost ratio for user testing comes from 3 - 5 participants
- 80% of problems can be detected with 4 or 5 participants
- Serious flaws tend to show up earlier

In another iterative approach, heuristic evaluation as the first phase of a two-phase usability effort can greatly increase the value of laboratory testing. By identifying obvious or clear-cut usability problems, heuristic evaluation “harvests the low-hanging fruit” and provides a focus for laboratory testing (11). Without prior heuristic evaluation, ten test participants may spend half their sessions struggling with the same obvious usability problem. Meanwhile, other, equally important usability problems can be “masked” by the first problem and not be found during laboratory testing.

However, usability programs consisting only of exploratory usability testing and/or heuristic evaluation have potential weaknesses that may not be immediately apparent:

- They may not evaluate different audience groups; most small-sample usability tests assume a fairly homogenous audience
- They don’t observe users in their context of work
- They don’t address longitudinal issues; most observations focus on ease of learning and the “out of box” experience

The remainder of this paper addresses the tradeoffs among different methods and how to make effective decisions about combining methods, depending on the specific situation.

## **WHEN TO USE VARIOUS METHODS**

For each of the usability methods described earlier, we can consider when to use it, based on the answers to a number of questions:

- Where are you in the product development cycle of the target product?
- How does the target product fit into the organization’s matrix of products? Is it an existing product or a new one? If a new product, is it a new version in a mature product line, or is it the first of a new sequence of products?
- What questions do you want to answer? What prior data, if any, do you have about potential user problems or “risky” areas in the interface? What design issues have arisen about which the design team disagreed or found the decisions difficult?
- Which audiences do you want to study? How many possible target audiences does the product have, and why is each of interest? How might the goals or problems of the various audiences be different? Which user audiences are most important to the success of this release of the product?
- Similarly, what usage scenarios are of special interest to you? What design changes might you consider for the product depending on the different ways people might want to use it?
- How much do you want to learn about your user audience right now? Have you only time to discover “show-stoppers” in a forthcoming product, or do you want to make long-term planning decisions about future product design strategy?

A successful multiple-method usability program should be balanced to reflect your organization’s strategic goals. Some studies should focus on immediate short-term results, while others should address longer-term concerns. There will always be pressure from both marketing and engineering for the fastest, cheapest results. But unless your organization doesn’t expect to be in business next year, we owe both management and ourselves a more balanced, broader collection of user data.

## **Guidelines for Decision-making**

Here are a few guidelines tied to the questions:

- If you’ve never performed usability work on this product line or for this organization (or division or business unit), it often makes sense to begin with an exploratory usability test. Problem identification is likely to be the highest priority goal. Also, it’s easy to arrange laboratory testing sessions with an observation area; giving product development teams the experience of watching real users may be needed to jump-start the acceptance of usability recommendations.

- Remember that “too late” in the development cycle of a product to perform usability studies can be comfortably early in the design cycle of the next version. When development teams have no resources to spare before a product release, a study scheduled for immediately after the release can get support and provide excellent data to inform the next release. Especially with the condensed “Internet time” schedules we now see in e-commerce, we should consider decoupling the schedules for specific usability studies from product release schedules. Usability data will always benefit design; we don’t need to engage in schedule wars we can’t win.
- The more a product is targeted at a horizontal audience (“every desktop will have this product,” said a recent client), the less likely the team has identified its most important audience segments. Horizontal products especially need field studies to learn about different contexts of use, as well as larger-sample usability tests representing different possible audience segments.
- The more innovative a product (that is, the less a product is like its predecessors), the less the team probably knows about its target audiences, because we don’t have existing user data. Innovative products should have more of their usability budget allocated earlier in the development cycle than should more mature products (12).

### Planning Calculations

The following guidelines for planning calculations are suggestions only, based on my personal experience. Usability specialists in different situations and contexts may find quite different ratios more appropriate; however, the process of performing the calculations will still be valuable.

1. Decide on a time window for planning, preferably at least a year.
2. Decide on an approximate frequency for usability studies for the product or product line. Many organizations plan quarterly studies, although some have the resources for more frequent iterations. Some product teams my firm has worked with can handle only one or two studies a year.
3. Decide what percentage of the projects should be field studies; try for at least one-third or at least one a year (contextual inquiries or ethnographic interviews).
4. Decide what percentage of the projects should be task-based user studies (laboratory testing or contextual inquiry); don’t let this number fall below 50%. That is, “user-free” studies like heuristic evaluations or cognitive walkthroughs, plus studies without individual tasks (focus groups, ethnographic interviews), should constitute less than half of the program.
5. Decide what percentage of the studies should address each of the key audience groups. This decision is highly dependent on the characteristics of your organization and products. Most usability practitioners are conscientious about recruiting inexperienced or novice product users, but there is often a paucity of continued-use studies with experienced users. Discussions with technical support or help desk personnel can identify audience groups that frequently experience problems.
6. Decide what issues each study should address. Many specific problems can't be identified until after the previous study is over, but it's worthwhile to assign high-level concerns to different studies during the planning for a year's work. Then, after the usual efforts to handle potential "show-stoppers," you can have some confidence that major issues or concerns will be explored sometime during the year, even if not in the current study.

### Example of a Multiple-Method Program

Here is an imaginary multiple-methods usability program illustrating the above guidelines. The organization has an existing product, Product A, and is planning major changes it will release as Product B later this year.

1. Do contextual inquiry of Product A to learn about target users' goals and tasks, and to find out how people are actually using Product A (what problems they have with Product A, where it works well for them and where it doesn't). This information drives the design for the Product B prototype.
2. Design and prototype Product B.
3. Do an exploratory laboratory test of Product B to identify problems that many members of the target audience would encounter with Product B as prototyped.
4. Revise the prototype to address the identified problems.
5. Conduct task-based usability focus groups on the revised prototype to learn how various audience segments might use Product B and to identify their preferences (and priorities) for product features and behavior. At this point in product development, the team is often making triage decisions about which features will be implemented; this study provides data for such decisions.
6. Develop the alpha version of Product B, using the identified priorities.

7. Do performance laboratory testing of Product B, with three audience segments, to identify problems tied to specific audience segments and to measure how successfully the audiences can use Product B.
8. Develop the beta version of Product B.
9. Conduct ethnographic interviews with beta customers to learn how users interact with Product B in the contexts of their own work. This study may uncover problems that were not obvious when users performed predefined controlled tasks.
10. Make minor changes to Product B based on the ethnographic interviews.
11. Release Product B.
12. Do contextual inquiry of Product B to inform the design of the Product C prototype.

## COMMUNICATING THE BENEFITS OF MULTIPLE-METHOD PROGRAMS

A key issue in justifying mixed-method usability programs to management or to our consulting clients is that we are not criticizing laboratory testing as an important method, nor are we trying to de-emphasize the value of existing usability labs or equipment. Rather, using only one method—any one method—can be damaging to product success, because a single method may miss collecting crucial data that would result in the team making quite different design decisions. For example:

- Contextual inquiries can show that hardware products are the wrong size or shape to be satisfactory in their likely locations of use
- Continued-use ethnographic interviews can uncover problems experienced users have with software that was tested only for learnability, where a different design would support both first-time and experienced users
- Usability focus groups with different target audience segments can help prioritize the importance of new features, by collecting data from a much larger population sample than is normally practical for laboratory testing
- Heuristic evaluation can identify potential usability problems before the team has moved so far down one design path that it's expensive to consider major changes in the design direction

We don't yet have enough research data on individual research methods—none on multiple-method usability programs—to draw conclusions about how much more value such programs add to product success or profitability. However, multiple-method usability programs provide valuable insurance that our usability efforts aren't missing major "show stoppers." In addition, the richer data we collect using multiple methods will improve both the quality and the quantity of all the usability recommendations we make.

## REFERENCES

- (1) Rosenbaum, S.; Rohn, J.; and Humburg, J (2000). "A Toolkit for Strategic Usability: Results from Workshops, Panels, and Surveys." *CHI 2000 Conference Proceedings*.
- (2) Dumas, J and Redish, J (1999). *A Practical Guide to Usability Testing*. 2<sup>nd</sup> edition. Portland, Oregon: Intellect Books.
- (3) Kantner, Laurie (1994). "Techniques for Managing a Usability Test." *IEEE Transactions of Professional Communication*, 37(3) pp. 143-148.
- (4) Holtzblatt, Karen, and Sandra Jones, (1993). "Contextual Inquiry: A Participatory Technique for System Design." In *Participatory Design: Principles and Practices*. Douglas Schuler and Aki Namioka (Eds.). Hillsdale, New Jersey: Lawrence Erlbaum. pp. 177-210.
- (5) Raven, M.E., and Flanders, A. (1996). "Using Contextual Inquiry to Learn About Your Audience." *ACM SIGDOC Journal of Computer Documentation*, Vol. 20, No. 1.
- (6) Wood, L., (1996). "The Ethnographic Interview in User-Centered Work/Task Analysis." In *Field Methods Casebook for Software Design*. Wixom, D. and Ramey, J. (Eds.). New York: John Wiley & Sons, Inc.
- (7) Neilsen, J. (1993). *Usability Engineering*. New York, NY: Academic Press, Inc.

- (8) Kruger, R. (1994). *Focus Groups, A Practical Guide for Applied Research*. Thousand Oaks, CA: SAGE Publications, Inc.
- (9) Nielsen J. and Landauer, T.K. (1993). "A Mathematical Model of the Finding of Usability Problems." *Proceedings of ACM INTERCHI '93*.
- (10) Virzi, R. (1992). "Refining the Test Phase of Usability Evaluation: How Many Subjects is Enough?" *Human Factors*, The Human Factors Society, Vol. 34, No. 4, pp. 457-468.
- (11) Rosenbaum, S. and Kantner, L. (1997). "Usability Studies of WWW Sites: Heuristic Evaluation vs. Laboratory Testing." *ACM SIGDOC 97 Conference Proceedings*.
- (12) Ramey, J., personal communication (1996).

## **ABOUT THE AUTHOR**

Stephanie Rosenbaum is founder and president of Tec-Ed, Inc., a 15-person firm specializing in usability research, user interface design (including Web information architecture and graphic design), user documentation, and marketing communication. Headquartered in Ann Arbor, Michigan, Tec-Ed maintains offices in Palo Alto, California and Rochester, New York.

Tec-Ed usability clients include Sun Microsystems, Intuit, Cisco Systems, Molecular Dynamics, Xerox, and many lesser-known firms—as well as “stealth” e-commerce ventures. Recent multiple-method engagements were for Philips Medical Systems, Thomas Publishing Company’s Product News Network, and Latitude Communications. Stephanie’s role in Tec-Ed’s usability activities focuses on identifying appropriate methods and justifying usability programs to client companies.

A member of the Usability Professionals’ Association, ACM SIGCHI, and the Human Factors and Ergonomics society, as well as a Fellow and Exemplar of the Society for Technical Communication, Stephanie has delivered workshops and presentations for UPA, SIGCHI, HFES, the IEEE, and companies such as Federal Express. For the past four years, she has co-presented a series of ACM SIGCHI panels and workshops on corporate strategy and usability research. Her publications include a chapter in John Carroll’s recent volume on *Minimalism Beyond the Nurnberg Funnel*.