Alternative Methods for Field Usability Research

Laurie Kantner
Tec-Ed, Inc.
4300 Varsity Drive
Ann Arbor, MI 48108
1-734-995-1010

laurie@teced.com

Deborah Hinderer Sova
Tec-Ed, Inc.
608 Walnut Street
South Milwaukee, WI 53172
1-414-762-6584

deborah@teced.com

Stephanie Rosenbaum
Tec-Ed, Inc.
Post Office Box 51054
Palo Alto, CA 94303
1-650-493-1010

stephanie@teced.com

ABSTRACT
Field usability research involves observing people in their own environments—for example, workplaces, homes, and schools—to learn their normal or natural behavior. Through field research, we can gain an in-depth understanding of the goals, needs, and activities of people who use the products and documentation we design and develop. This paper introduces three field research methods—condensed contextual inquiry, ethnographic interviewing, and field usability testing—illustrated with a short case history of each method. The paper then describes when and why to use each method, that is, how to choose the appropriate method for different data-collection goals.

Categories and Subject Descriptors
H.5.2 [**Information Systems**] Interfaces and Presentation — *User Interfaces, Theory and methods*

General Terms
Human Factors

Keywords
Usability research, contextual inquiry, ethnographic interview, usability testing

1. BACKGROUND
Field usability research involves observing people in their own environments—for example, workplaces, homes, and schools—to learn their normal or natural behavior. Through field research, we can gain an in-depth understanding of the goals, needs, and activities of people who use the products and documentation we design and develop.

Until recently, field usability research has not gained wide acceptance among usability practitioners. Contextual inquiry [4, 7] and contextual design [3, 10] have been in-depth qualitative methods conducted primarily within large organizations that can invest in research for long-term product design improvements.

Similarly, the methodology for ethnographic interviewing has been developed through fairly extensive projects [11]. The multiple challenges of budget, schedule, and logistics make it difficult to convince management to support field research.

Over the past five years, the authors have developed three field research methods that are successful in the situations that usability practitioners face. We evolved the classic field research methods of contextual inquiry and ethnographic interviewing, and we adapted laboratory testing methodology for use in the field. This paper introduces the three methods—condensed contextual inquiry [9], ethnographic interviewing [1], and field usability testing—illustrated with a short case history of each method. The paper then describes when and why to use each method, that is, how to choose the appropriate method for different data-collection goals.

2. DESCRIPTIONS OF METHODS
2.1 Condensed Contextual Inquiry
Condensed contextual inquiry applies the principles of contextual inquiry, where users become partners with the research team and engage in dialogue as they perform normal activities. The observations allow researchers to collect data about ongoing rather than summary experience. Contextual inquiry is based on a clearly defined set of concerns, rather than a list of specific questions. In *condensed* contextual inquiry, we identify a more constrained set of concerns, enabling us to focus on a few key issues and gather concrete behavioral and perception data in shorter sessions with the participants.

2.2 Ethnographic Interviewing
While contextual inquiry is primarily observation of use, with inquiry from the researcher, ethnographic interviewing asks questions about use. We conduct ethnographic interviews to answer questions about issues that are broader than individual tasks, where time constraints cannot accommodate contextual inquiries, or when observing relevant user behavior is impractical because it occurs rarely or unpredictably (or cannot be interrupted, such as surgery). Holding interviews in the users' environment "grounds" the interviews in surroundings and artifacts, making the discussions more concrete and uncovering factors such as reminders, routines, work flows, and collaboration.

2.3 Field Usability Testing

Field usability testing adapts the methodology of laboratory testing [2, 5] by conducting the sessions in the participants' own environments, on their own computers (or other equipment). Especially in home-based research, participants' choices of computer, software, and ISP noticeably affect their experience and behavior with products and services. In field usability testing, we also design the tasks to address the participants' own goals, where task objects include the users' own files, bookmarks, or databases. These adaptations, while producing context-rich qualitative data about the target audience, make field usability testing more suitable for problem identification than for performance measures or quantitative comparisons.

3. CASE HISTORIES

This section describes a short case history for each of the three field research methodologies described in this paper: condensed contextual inquiry, ethnographic interviewing, and field usability testing.

3.1 Condensed Contextual Inquiry – Case History

A well-known Internet service provider wanted to learn how people look for information on the Internet from their home computers. The goal was to understand how people with low to medium search ability use search engines and browsing to find items reflecting their own interests. The portal designers were interested in differences in behavior based on the search goal itself (the extent to which the item was known) and the comprehensiveness of the search results sought.

3.1.1 Methodology

A two-person team made 18 home visits in southeastern Michigan and upstate New York to conduct one-hour usability sessions. In these sessions, participants performed from one to three "information lookup" tasks of their own devising. One usability researcher facilitated the sessions and took notes; the second researcher took digital photographs, operated the audio tape recorder, and took notes.

Participants were recruited through newspaper advertising. During the telephone screening interviews, recruiters asked open-ended questions to determine the candidate's skill and experience in searching. We also sought a balance in employment, education, household income, gender, and favored search engine.

In contextual inquiry, we collect two kinds of qualitative data: the researchers' observations of participants' behavior and participants' comments and explanations. In this study, we collected observations about the 36 lookups performed, participant commentary during tasks, participant opinions of task success, participant opinions of lookup experience, and participant reasons for ease/difficulty ratings.

Unlike many contextual inquiries, this study also gathered quantitative data. Quantitative data collected included: number of lookups, number of iterations per search, number of words per search string per iteration, type of search (known item, exploratory, existence, and comprehensive), consistency between stated search goal and item found, search style (linear or berrypicking), starting point (whole Web or site-specific); perceived and observed task success, and participant ease/difficulty ratings of various Internet activities.

3.1.2 Session Activities and Usability Findings

The session began with a warm-up period in which we asked the participant to describe a recent lookup activity on the Web. In addition to ice-breaking, the purpose of this activity was to confirm the participant's searching ability and to learn more about the participant's interests in case we needed to suggest a lookup activity later on.

The rest of the session consisted of the participant looking up something of interest, articulating the goal (which we wrote down), and thinking out loud as they used searching and browsing (often in combination) to find information about the item. At certain points in the task, the facilitator asked probing questions about opinions and rationale. When the participant completed a lookup, we asked debriefing questions. Then, if time permitted, we repeated the process. Most participants performed two lookups during their session. For two participants, one lookup consumed the entire one-hour session. A few participants performed three lookups.

While the specific findings are proprietary, we can describe some of the general findings. For example, the mean of two words per search string and participants' difficulty finding the search function if it was a link or filled-in box were consistent with other published findings [6]. Participants who employed a berrypicking approach, where they modify their search goal (usually in response to information in the search results), reported fewer instances of dissatisfaction with their search experience than participants who employed a linear approach (sticking with their goal).

Not surprisingly, participants' "equipment speed," defined as a combination of processor speed and Internet connection speed, strongly affected participants' success rate and satisfaction with their lookup tasks. The interview team used this dimension as a key facet in creating personas based on this study.

In addition to participants using their own equipment, performing this research in the field fostered more realistic searching by surrounding participants with cues, slips of paper, and other objects that reminded them of areas of interest. Pursuing personal lookup goals, users were invested in the outcomes and exhibited more natural behavior than when external goals are assigned. At the conclusion of each lookup, participants rated the outcome on an importance rating scale; for the few lookups where the interview facilitator suggested the search goal, participants rated those outcomes as "not important."

3.2 Ethnographic Interviewing – Case History

The interactive agency for a Big Three automaker wanted to learn how vehicle owners keep vehicle records (maintenance, insurance, etc.). The agency planned to redesign the vehicle owner website provided by the automaker, based in part on what kinds of information vehicle owners keep and where they keep it. In addition, the agency wanted to learn people's vocabulary for vehicle-related information and ways to organize information on the website that would make sense to vehicle owners.

3.2.1 Methodology

A two-person usability team made 19 home visits to vehicle owners in the San Francisco Bay area, southeastern Michigan, and upstate New York. During these one-hour interviews, participants answered questions, performed a card-sorting exercise, and then

answered a few more questions. The goal of the research—to inform redesign of a vehicle owner website—was not divulged to participants until near the end of the interview.

Participants were recruited through a display ad in the local newspaper, and screening interviews ensured that candidates owned or leased a vehicle made by the Big Three automaker and that they used the Internet. Candidates also needed to agree to have digital photographs taken so that the study team could provide visual representations of the various records people kept and the locations where they kept them.

The interviews were conducted with three different interview teams who all worked from the same protocol. After one team completed six interviews in California, all interview team members met by phone to discuss the interviews and refine the protocol. In addition, the teams wrote interview summaries using an agreed-on structure that ensured consistent data collection and simplified the task of consolidating all summaries into data tables for analysis.

The card-sorting exercise used 70 different terms, which participants organized into groups and then labeled the groups with their own category names. The results of the card sorting exercise were analyzed using freeware from IBM (USort and EZCalc) and with spreadsheets.

The usability team took digital photographs of the different types of vehicle records and their locations. The team made sure that no private information was legible within any photographs.

Qualitative data collected included stories participants told about vehicle ownership, opinions participants gave about topics presented in the card-sorting exercise, and opinions participants gave about Internet use. Quantitative data collected included types of vehicle records kept, patterns in participant descriptions of vehicle ownership, and patterns in the card-sorting results.

3.2.2 Session Activities and Usability Findings

The session began with an ice-breaking set of questions about acquiring the vehicle and the experience of owning it. Then the participant showed the usability team the records kept in the vehicle, as well as additional records kept away from the vehicle, such as in the house or garage.

The interview team conducted the card-sorting exercise on an available surface in the participant's home—kitchen table, coffee table, even an old desk on the porch of a student apartment house. The interview ended with more questions about keeping vehicle information on the Web.

Performing this research in the field, instead of bringing people to a facility, enabled the usability team to see many more examples of vehicle records than the participants would have recalled or brought with them to an interview away from their homes. The researchers could adopt the stance of observers and recorders instead of interrogators. We did not want to ask a series of questions like "Do you keep oil change records? Where?" to avoid putting participants who keep fewer records on the defensive. We simply asked, "What do you keep?" and watched them discover for themselves. One notable experience was a participant suddenly recalling that he was using a blackboard in his garage to keep a record of his latest oil change.

The findings reported for the vehicle study identified individual cases as well as trends or patterns based on participant characteristics and type of vehicle owned. For example, we found that owners of SUVs described vehicle ownership stages differently from owners of cars and trucks. The card sorting data informed the website designers of which topic areas participants identified as related and which topics were considered unimportant. The study data also helped add detail to the vehicle owner personas used by the agency.

3.3 Field Usability Testing – Case History

A major publisher of engineering journals wanted to learn how effectively researchers and librarians could use an electronic library product to locate desired articles and papers. The publisher was preparing to launch its new online biomedical engineering library service. The primary goals of this usability study were to evaluate how effectively typical users could use a prototype of the new library to locate full-text articles and papers of interest to them, to learn where the design of the library could be improved to enhance its usability, and to collect feedback from users about their experiences using the library.

3.3.1 Methodology

To gather information pertinent to the study goals, Tec-Ed conducted one-hour individual usability sessions at academic and corporate institutions where we could gain access to people who fit the publisher's criteria of performing biomedical research or supporting that research in a librarian role. Finding participating institutions who would make people available, and finding participants who would offer the time, was a difficult process. If we had asked people to leave their work place to participate in the sessions, fewer people would have agreed to participate because of time constraints.

The two-person usability team observed ten people, one at a time, as they searched for biomedical content specific to their own interests in the prototype biomedical engineering library and "thought out loud" about their experience. Each participant performed similar types of search tasks at their own computer, facilitated by the test administrator, who used the same task script for each participant while the observer took notes.

Tec-Ed conducted ten sessions at three different locations, including a pharmaceutical firm, a commercial biochemical research laboratory, and a university conducting biomedical research. Sessions were audiotaped at the two of the three sites (the third site denied permission for recording of any sort).

Quantitative data collected included task success, degree of administrator prompting needed, and ease of use ratings. Qualitative data included observations about participant behavior and opinions and rationale participants offered during their running "thinking aloud" commentary.

3.3.2 Test Tasks and Usability Findings

During the sessions, participants were asked to explore the home page and then find articles of interest based on types of criteria such as subject keyword, author, and category. During these activities, the study team noted where participants either expressed confusion about choices or terms in the user interface or experienced task difficulty or failure because they did not fully understand the user interface.

Tec-Ed reported 36 study findings to the engineering journal publisher in the following assessment rating categories.

Table 1. Findings Per Usability Assessment Rating Categories

Assessment Rating	Definition	Number of Findings
Show-stopper	Prevents completion of the task.	3
Impedance	Causes significant delay, effort, and frustration.	9
Annoyance	Causes minimal impact on user performance, but is an unnecessary source of frustration or irritation.	3
Deficiency	Needs to be more readable, or provide additional information, to promote usability.	6
Info Only	Needs more research to determine its impact on usability	11
Keeper	Promotes usability.	4
Total Findings		36

An example of a show-stopper finding was difficulty interpreting the choices for the Refine Search feature, which led to inadvertently achieving zero results. Another show-stopper finding was having to scroll to find the Help button.

The study team formulated detailed recommendations for the 12 findings in the show-stopper and impedance categories, as these would provide the most return on investment in improving the user interface.

4. CHOOSING APPROPRIATE METHODS FOR FIELD RESEARCH
Each of the field research methods described in this paper offers insight into how people actually use products and services within their own environments. When is each method appropriate?

4.1 When to Use Condensed Contextual Inquiry
Contextual inquiry assumes that participants are using a product or system to perform normal tasks. It is useful for examining a "continued use" situation, in contrast to exploring out-of-box or "learnability" issues, for which usability testing is an excellent method. Condensed contextual inquiry is a preferred method for:

- Exploring how people use a competitive or predecessor product, to identify features or feature improvements for a new or new-version product.

- Validating effective product use in the field for a product recently released.

A condensed contextual inquiry project can take 6 to 10 weeks, depending largely on the time required to recruit individual participants (if home users) or participating customer companies (if users are at work sites). For work-site recruiting, sales or marketing staff are often involved in making the first contact.

4.2 When to Use Ethnographic Interviewing
Ethnographic interviewing provides data that is useful for adding the human-centered touch to a technology solution. For example:

- You are designing a technology solution for an everyday task and need to understand the context in which it will be used so that it will seamlessly intersect other activities in people's lives.

- You are gaining perceptions about a product or service or identifying barriers to adopting a product or service, so you can establish appropriate marketing channels and create effective communications.

Ethnographic interviewing can take between 4 and 8 weeks, depending on the time required to locate participants with the right characteristics. Some ethnographic studies include longitudinal research, in which the study team follows up the initial interviews with continuing conversations and data collection over a period of time. These studies can take 3 months or longer.

4.3 When to Use Field Usability Testing
Field usability testing is an appropriate method for researching ease of use of a prototype or fully functional product, where we want to examine these issues in the context of the participant's own equipment or artifacts, or using the participant's own data. Appropriate uses include:

- Collecting usability feedback about a prototype of an application people are using to manage and report on large databases. When users work with their own data, we collect more in-depth feedback about the product itself because there are no distractions from simulated data.

- Exploring the usability of the system administration features of a complex (released) product. Again, by exploring these features in the field, we collect in-depth data about the product itself because the context is familiar to the user.

The tradeoff between laboratory and field usability testing is in whether you gain more from in-context exploration than what you lose from having less control over the flow of test tasks themselves. The best approach is conducting laboratory testing of development versions and supplementing that data with field research on the production version [8].

5. CONCLUSION
Many documentation and usability practitioners have successfully built laboratory testing into the product development processes at their organizations. The methods described in this paper will help practitioners gain the benefits of field research as well.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] Anschuetz, L. and Rosenbaum, S. "Ethnographic Interviews Guide Design of Website for Vehicle Buyers." In Proceedings of CHI 2003 (Fort Lauderdale FL, April 2003), 652-653.

[2] Dumas, J. and Redish, J. *A Practical Guide to Usability Testing*. (revised edition). Intellect Books, Portland OR, 1999.

[3] Holtzblatt, K. and Beyer, H. "Contextual Design Principles and Practice." *Field Methods Casebook for Software Design*. Wixon, D. and Ramey, J. (Eds.). John Wiley & Sons, Inc., New York, NY, 1996.

[4] Holtzblatt, K. and Jones, S. "Contextual Inquiry: A Participatory Technique for System Design." *Participatory Design Principles and Practices*. Schuler, D. and Namioka, A. (Eds.). Lawrence Erlbaum, Hillsdale, NJ, 1993, 177-210.

[5] Kantner, L. "Techniques for Managing a Usability Test." IEEE Transactions on Professional Communication, (September 1994), Volume 37, Number 3.

[6] Nielsen, J. "Search: Visible and Simple." In Jakob Nielsen's Alertbox, May 13, 2001, http://www.useit.com/alertbox/20010513.html.

[7] Raven, M.E. and Flanders, A. "Using Contextual Inquiry to Learn About Your Audience." ACM SIGDOC Journal of Computer Documentation, Vol. 20, No. 1.

[8] Rosenbaum, S. "Not Just a Hammer: When and How to Employ Multiple Methods in Usability Programs." In Proceedings of UPA 2000 (Ashville NC, August 2000), tab 19.

[9] Rosenbaum, S. "Streamlining Field Methods" Wixon, D., et al. "Usability in Practice: Field Methods Evolution and Revolution." In Proceedings of CHI 2002 (Minneapolis MN, April 2002), 880-884.

[10] Wixon, D., Holtzblatt, K., and Knox, S. "Contextual Design: An Emergent View of System Design." In Proceedings of CHI 1990 (Seattle WA, April 1990), ACM Press, 329-336.

[11] Wood, L. "The Ethnographic Interview in User-Centered Work/Task Analysis." *Field Methods Casebook for Software Design*. Wixon, D. and Ramey, J. (Eds.). John Wiley & Sons, Inc., New York, NY, 1996.