



© David Wasserman/Artville LLC.

Why Do Version 1.0 and Not Release It?

Conducting Field Trials of the Tablet PC

The Challenge of New Technology Design

There have been many discussions at conferences and in journals on the merits of various user-centered design (UCD) methodologies to address different types of design problems and how they fit into the design life-cycle. Generally, user-centered design professionals have promoted ethnographic work before the beginning of the design process, and iterative usability testing in the lab as the design progresses. Arguments for these UCD methods are well known to this community, and there is no question their adoption is a tremendous step forward over traditional software development approaches. However, we should guard against a too rigid division between ethnography and usability, and a too rigid sequence in practice, with ethnography only in the pre-design phase, and laboratory usability testing as the sole UCD method later in design.

In this article, we look at a situation that highlights potential limitations of this approach, one in which there were both a clear research need and a strong business case

for a hybrid approach, integrating ethnography and usability throughout the development process. We call this approach “ethnographic field trials.” In our experience, the arguments for such an approach and its cost-justification are the strongest when introducing innovative technology. We will illustrate this with the example of a series of field trials we carried out over one and a half years to study three different iterations of the Microsoft Tablet PC prototype. We hope to follow up in subsequent articles with more detailed discussion of methodological issues and examples of ways in which the studies influenced the evolving design.

A Case Study: The Microsoft Tablet PC

As introduced on the Microsoft Web site (www.microsoft.com/tabletpc/), the Tablet PC is envisioned as the next-generation mobile business personal computer. It is expected to be available from leading computer makers in the second half of 2002. The Tablet PC runs the Microsoft® Windows® XP Tablet PC



**Susan M. Dray and
David A. Siegel**
Dray & Associates, Inc.
2007 Kenwood Parkway
Minneapolis, MN 55405,
USA
+1-612-377-1980
fax: +1-612-377-0363
dray@acm.org

**Evan Feldman and
Maria T. Potenza**
Tablet PC User Research
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052

Edition and features the capabilities of current business laptops, including attached or detachable keyboards and the ability to run Windows-based applications. The biggest innovation of the Tablet PC is that it extends pen and speech capabilities to the power of a full-function laptop. Using a special stylus or pen, the user can create, manipulate, and manage handwritten electronic documents (“notes”) and add handwritten electronic annotations to imported documents. Users can also interact with Windows-based applications via the pen or speech, even without a keyboard. The Tablet PC is intended to allow knowledge workers to use the PC in a variety of new ways and settings in which the use of a keyboard would be impractical. The hope of the design team is for the Tablet PC to be adopted widely and fully integrated into people’s work. For this to happen, it must not only be highly usable, but also perceived to be extremely useful and more beneficial than current options.

Traditional User-Centered Design

In order to develop and refine the initial concept for the Tablet PC, the Microsoft team performed ethnographic studies and contextual inquiry to understand knowledge worker

tasks, document management, and note-taking practices. They also studied knowledge worker characteristics for development of personae and other elaborate user profiles. The team also used a variety of market research methods to characterize the target user. The team did many traditional usability studies in the lab, evaluating the usability of features during development, and made many significant changes in design on the basis of these results.

These research methods provided indispensable input into the development process, but they were insufficient to answer some of the most crucial questions, which were more global in nature, namely,

- * Will the Tablet PC be usable and useful in the workplace, and will users successfully adopt it?
- * Will it be integrated into the workplace and work practices of users who explore it on their own?
- * In what ways will users have to change their practices to exploit new capabilities, how motivated will they be to do so, and what obstacles will they encounter?

We also wanted to understand usability at various points in the learning curve and at various stages in the adoption process. Was the Tablet PC useful and usable enough initially to allow users to work with it immediately, while continuing to explore and use additional functionality over time? How much of this functionality was used (or needed) once users were more experienced? For each of these questions, we wanted to assess utility and usability and the complex relationship between them. In short, we needed to fully understand the user experience with the new device, in the natural setting, over time, and to evaluate separate features and design elements in the context of the overall user experience.

To answer these questions, we needed information from studies of users interacting with Tablet PC prototypes. Ethnographic data without interaction with Tablet PC was of limited use. Such studies provide essential clues about user



needs and contextual factors that are likely to influence the acceptance, usefulness, and usability of a design, but these clues are somewhat indirect. This is especially true of new technology—with functionality previously unimagined, new tasks and scenarios enabled, and a change to the overall structure of existing work patterns. These questions also could not be answered primarily based on self-report methods, in which users describe perceived frustrations with their existing tools and practices and react to the product concept. Self-report is likely to be influenced by how “cool” various features sound, but often does not accurately predict actual user behavior.

Iterative usability testing alone could also not provide the answers. Common usability methods tend to provide fragmentary information; in this case what was needed was a comprehensive evaluation of the user experience over time. Usability tests are clearly indispensable but inevitably use tasks constructed to probe specific design features and functions. Even when task scenarios are very well designed and based on user data, they do not necessarily capture spontaneous intentions and motivations of the user at the time. Also, they present tasks in isolation, rather than as part of the overall work flow and demands encountered by a user during the workday. The focus is on a limited sample of tasks, usually chosen to evaluate high priority design concerns, not to capture the full range of natural usage. In addition, usability testing gives a snapshot of ease of use at a particular point in the learning curve, typically the beginning. Although we can often extrapolate from this, it may not be enough to predict users’ experience and satisfaction over time, especially when tasks are perceived as novel and users expect to go through a learning and discovery process.

Many of these issues are inherent to the introduction of new technology. The risk of not addressing these questions directly is particularly high considering the huge investment in launching the new technology. This is especially true when actual usage of the capabilities is unknown. Furthermore, when the new technology has a broad vision or is a gen-

eral-purpose tool, like the Tablet PC, people are likely to use it in a wide variety of individualistic ways, combining its capabilities in various ways to carry out real life tasks. Thus, significant interaction effects exist in which the usability of one element influences the usability of others, and the exact combinations will be different for different users and tasks. These are hard to capture in the lab.

The success of the Tablet PC depends not just on the ability of individual users to carry out tasks in isolation, but on the fit between the uses it lends itself to and the work context, including the social dynamics, work practice, and technical infrastructure. In view of these factors, understanding the user experience calls for a more holistic evaluation process. We needed a method that could look in an integrated way at discoverability, utility, usability, and fit with the context, and we could evaluate these things in the natural setting only if we wanted to capture the range of variation we expected.

How to Proceed?

Microsoft faced a limited set of traditional choices, none of which was realistic. One alternative was to release the product and watch market reactions while getting feedback from early adopters. This is probably the most common response and the riskiest. If the product doesn’t meet users’ expectations or doesn’t fit well into their existing or modified work process, the product won’t be used. From a development perspective, redesigning the product is expensive, especially if core functionality is not appropriate. In addition, sometimes it takes many iterations of the product in the market to get closer to meeting the user’s needs and expectations. Often, however, these iterations will not happen because a product does not even get a subsequent release if it performs poorly initially. Although usability tests and traditional ethnography help minimize this risk, answers to many of the questions about adoption and use remain too elusive, awaiting the judgement of the market.

Another alternative was for the team to gather usability data from the beta release,

BUSINESS COLUMN EDITOR

Susan Dray

Dray & Associates, Inc.

2007 Kenwood Parkway

Minneapolis, MN 55405,

USA

+1-612-377-1980

fax: +1-612-377-0363

dray@acm.org

even though a beta test is almost always a bug-fix release. However, beta testers are rarely typical users of a new technology, and the information is probably too late to affect design, although this information may become the starting point for design changes for subsequent releases.

The Tablet PC team members were not satisfied with any of these traditional approaches. Instead, they wanted to get feedback and observe use over time before they released the product. They wanted to do this with real knowledge workers in real companies doing real work, to understand the technical and personal challenges that customers would face

structured interaction with the Tablet PC in order to provide a picture of naturally evolving usage patterns and usability in context. Because the study had to combine different methodologies and different degrees of structure, managing and interpreting the highly heterogeneous qualitative data would be a major challenge.

Guided by all these considerations, we worked closely together for almost two years to conduct a series of three field trials at a variety of different companies to evaluate successively redesigned prototypes of the Tablet PC. The field trials helped identify issues in usage and usability that occurred during long-term

Because the study had to combine different methodologies and different degrees of structure, managing and interpreting the highly heterogeneous qualitative data would be a major challenge.

and to fix them before launch. They knew that this called for a new approach and a significant investment in time and dollars. However, the risk of not addressing these questions directly was perceived as particularly high in the light of previous attempts with this type of technology.

Study Requirements

Only field research would be appropriate. The study would also have to be longitudinal, to allow for studying usage and usability over time. In addition, it had to be done in a way that allowed the expected wide range of variation in usage to evolve. The methodology had to be open enough to allow for shifting the focus to follow interesting trends or to address gaps in our knowledge as they became evident. It had to allow for preplanned, structured evaluation activities in order to get equivalent data on key design features across users. It also had to allow for extensive non-

use. We started with the rudimentary prototype with only a small subset of the ultimate functionality for the first field trial (as a dry run for later trials) and with the two subsequent trials using more sophisticated prototypes.

Field Trials

We hope to discuss the study process and methodology in more detail in a later article, but following are some of the main considerations.

- * **Participants.** We learned that we could not rely on traditional screeners and self-report for recruiting participants. Instead, we needed to do an in-depth study of candidates' jobs and work practices. This allowed us to understand the baseline work practices, balance the sample in a variety of ways, and ensure a rich spectrum of work styles. It's crucial to also have a large

enough sample of participants so that the individual differences and interaction styles can be thoroughly examined. The first study (the pilot) had 19 participants; the second study had 21; and the third study, which was designed to examine changes made following the second study, had only seven.

- * **Visits.** In the first trial, we visited participants twice in one week. In the second and third trials, we began with an initial selection visit with all candidates. We then visited each person five times over four weeks in the second trial and seven times over six weeks in the third trial. Each visit lasted approximately two hours. The visit protocols varied and included training, scripted usability tasks, observation of the participant using the Tablet PC in various work situations, artifact walkthroughs, and semistructured interview. In addition to the scheduled visits, we carried out brief, impromptu visits with all users at various times.
- * **Observers.** Rotating teams of Microsoft Tablet PC team members observed all visits. We had to balance the need for exposure of the team to real users and the need for real-time input into the development process with the risk of premature closure based on team members' impressions from their brief observations.
- * **Training of users.** In deciding how much support and guidance to give users, we had to balance our desire to simulate typical conditions (in which users do not necessarily read manuals or receive corporate training) with the need to make sure that users did not simply hit a dead end beyond which we would learn nothing more. We addressed this issue by providing users with a basic initial training that alerted them to the potential capabilities of the Tablet PC and prepared them to begin exploration. Beyond that, we used situations in which users needed support and guidance as opportunities for more

Some Process Hints for Field Trials

Pilot the process. If you are new to conducting field trials, we suggest that you run a pilot test before embarking on a full-blown trial. We found that this helped because the integration of a variety of methods is tricky and because it took some time to pinpoint the best combination of tools.

Recruiting is trickier than with “traditional” studies.

Because we had to be certain we were recruiting people who really fit the target market, we devised a two-step recruiting process. First, Microsoft identified target companies and worked with people in those companies to identify a pool of potential candidates. We then interviewed each candidate twice—first by phone, and then in person, to understand his or her job, technical expertise level, and interest in participating. We were able to use this information to select a subset to participate in each trial. Motivation is particularly critical in this type of study, because the participants commit a significant amount of time to the process. Therefore, it is also critical that their bosses buy in to their participation.

Decide up front how you will handle, sort, and analyze the data. You will probably have more data than you have ever had to sift through before. Plus, you will be incredibly busy collecting more data daily. Therefore, it is useful to consider carefully how you will handle the data. We used a Microsoft Access® database consisting of an exhaustive data structure that we developed before we started the research, as well as ample places to put data (and create new categories) that didn't fit any of these existing categories. This allowed us to sift relatively rapidly through thousands of factoids on the fly.

Logistics, logistics, logistics. As critical as logistics are in any ethnographic trial, they are even more critical when you are doing a field trial. Because each visit is an investment in a long-term relationship with the user, it is important that they be coordinated and that the team members be on time and briefed beforehand if there are visiting members so that the user's time is used most effectively. A central calendar and tight coordination were required in these studies, especially when we had multiple teams going to several locations on a given day.

Debrief long and often. We discovered that it was crucial to have several types of debrief. Thorough debriefing immediately after each visit allowed us to be sure we had captured all of the relevant facts, from all observers. Often this debriefing would take nearly as long as the visit itself. An additional debriefing at the end of the day allowed us both to adjust our procedure and to generate hypotheses about trends that we would watch for in future visits. A comprehensive debrief at the end of each week allowed us to combine the data from multiple teams who had participated in visits that week, and to synthesize the main findings from that week's research agenda.

detailed usability investigation, and we provided the needed assistance.


- * **Debriefing.** We followed each visit with an intensive debriefing during which we collected observations from all team members who were present. At the end of each day, we conducted a daily debriefing, during which we identified salient observations, themes, and hypotheses. We synthesized observations at least weekly, to identify trends, to provide input to the development team, and to refine the study focus.

What We Learned

In this process, we were able to discover a great deal about how people incorporate the Tablet PC into their work lives. We had an increased opportunity for what we called “opportunistic findings”—things that just happened to occur when we were present, either in our scheduled meetings or when we dropped in to see how people were doing. Throughout the studies we developed a rapport with each of the individuals and gained their trust, which made it easy for them to share their frustrations and triumphs with us. Some of the participants worked together and helped each other to learn their newly discovered “tricks.” In this way, we were able to trace the development of new work patterns and to see the collaborative learning curve firsthand.

We identified not only design changes, but also infrastructure and training issues, allowing the team to actively address them before fielding the final product.

We discovered, not surprisingly, that integration of a new technology evolves over time, confirming our decision to study users over weeks rather than hours or days. We were able to keep track of stages in the learning and adoption process and identify obstacles at different points. As people used the Tablet PC to do their real jobs, to create and save their own files, we watched them discover new ways of using the Tablet PC and saw the actual scenarios of usage as they evolved. We were then able to apply this learning directly to the product design.

The Tablet PC team has likened the field trials to shipping a first version of the Tablet PC and learning about the problems from 47 users rather than from millions of initial users. Because the team got to observe the visits—either in person or by videotape—they gained a much better sense of the users, their tasks, and their concerns. Most features were completely reengineered. Functionality and specific design elements were modified to improve usability, utility, and fit with the users’ work lives. We believe that as a result the real product that will ship later in 2002 will be more usable and more useful to the actual users. 

PERMISSION TO MAKE DIGITAL OR
HARD COPIES OF ALL OR PART OF THIS
WORK FOR PERSONAL OR CLASSROOM
USE IS GRANTED WITHOUT FEE
PROVIDED THAT COPIES ARE NOT
MADE OR DISTRIBUTED FOR PROFIT OR
COMMERCIAL ADVANTAGE AND THAT
COPIES BEAR THIS NOTICE AND THE
FULL CITATION ON THE FIRST PAGE.
TO COPY OTHERWISE, TO REPUBLISH,
TO POST ON SERVERS OR TO REDIS-
TRIBUTE TO LISTS, REQUIRES PRIOR
SPECIFIC PERMISSION AND/OR A FEE.
© ACM 1072-5220/02/0300 \$5.00