Are You Sure? *Really?* A Contextual Approach to Agile User Research

Josh Evnin, ThoughtWorks, Inc. *jevnin@thoughtworks.com*
Marjorie Pries, ThoughtWorks, Inc. *mhpries@thoughtworks.com*

Abstract

The creation of excellent user experiences often appears to be a forgotten goal in the software development world. This paper discusses the use of a concrete method, Contextual Inquiry, which leads to insights that will help development teams create experiences and interfaces that match user needs and expectations. This method encourages Agile team members to see the world from the users' perspective by working directly in the users' context.

Contextual Inquiry is a valuable tool for collecting and analyzing user data, superior to relying on general feedback from user self-reports and questionnaires. Our paper starts with a historical view of the problem, then covers how we planned and executed our Contextual Inquiry study, and concludes with lessons learned from our experiences.

1. Introduction

Contextual Inquiry is part of Contextual Design[1][2], a User-Centered Design methodology that calls for the observation of software application users within their naturally occurring work context. Our Agile team successfully utilized this method with a large client in order to discover what usability issues that occurred frequently for their user base. In the process, we found that directly trusting stakeholder requirements can lead to flawed software. Agile practitioners should instead focus on real user needs in context.

2. Application Background

In 1999, ThoughtWorks was approached by a large equipment leasing firm to automate aspects of their lease origination process. At the time, our client's origination processes varied from region to region using a variety of software tools combined with manual paper-based processes. It could take several weeks for a regional office to complete a deal and forward the paperwork to headquarters for booking into the back-end billing system.

Our client recognized there would be a competitive advantage to reorganizing into one central division; provided it had a suitable software solution to replace their current patchwork processes.

Over the years, ThoughtWorks has worked with this client to incrementally build and tailor this application for their largest branch operations and dealer networks.

3. Application Evolution

The initial releases of the application went into production in 2000 and by 2004 the application was in use internally in four major countries and also rolled-out to the North American dealerships. By early 2006, our client had internal product adoption through all four countries, processing over 90% of all targeted deals through the application.

With internal operations running smoothly, our client switched gears to focus more closely on "dealership level" problems. The application had been designed around the client's highly siloed operations and recent updates had focused on technical rather than functional improvements. While these releases were necessary for application flexibility and stability, they left an impression on many dealers that their concerns were not taken into account in the big picture. Our client was aware of this increasing dissatisfaction and they declared 2007 would be "The Year of the Dealer".

4. Client Reactions

Initially, the client decided to gather information from the dealerships by distributing an open-ended questionnaire. The format of the questions allowed dealership representatives to write their perceived problems in freeform, without bound. This format led to two problems in particular:

1. **There was simply too much data to be summarized.** The open format allowed respondents to write long laundry lists of bullet points without explanatory background or

978-0-7695-3321-6/08 $25.00 © 2008 IEEE
DOI 10.1109/Agile.2008.81

prioritization. Aggregated together, there were just too many details, and too many conflicting requests, for their analysis team to organize.

2. **The wrong kinds of feedback were returned.** When prompted by open-ended questions, survey respondents tend to focus on episodic details that stand out most to them, whether or not these issues are currently impacting their work in any economically significant way. Psychological research has shown that over the long term, episodic memories become marred by inaccuracies[3]. People will focus their attention and complaints on seemingly large, but irrelevant, infrequent, or inconsequential events such as "that one time when the system went down for half an hour."

Given these responses to their questionnaire, the client was at a loss as to where they should begin to provide solutions to the seemingly numerous problems with their system.

5. Finding a Way to Solve the Problem

Our client knew they had a problem and had done some things that seemed like the right thing to do to open communication channels but they were overwhelmed by the diversity of responses. This was a new situation for them and they did not have techniques and procedures in place to address it. The problems to be solved seemed scattered. Rather than focusing on a single issue such as "performance," usability was proving difficult to scope.

Despite these issues, the client was committed to making 2007 the year of the dealer, however it was already late Fall of 2006 and they did not have a story list in hand that addressed their survey results. They needed to get something together over the holidays in time for February sign-off in order to make good on their promise.

Because of our close relationship with the client, we had some visibility into the situation and were able to suggest a Contextual Inquiry approach to gathering user & usability requirements.

Contextual Inquiry calls for the observation of software application users within their naturally occurring work environment. It would allow us to observe users on-site, handling live deals in real time and give us techniques that would help sort out typical patterns from the exceptional events.

While ThoughtWorks had not previously used this method on a project, one of our application team members had a long-running interest in usability issues and introduced these topics to the client principal (account manager). The client principal recognized that this approach had the potential to collect a lot of high-quality data with a fairly small investment.

Specifically:

1. We wanted to get beyond "reported" problems, and see with our own eyes how dealers went about their work. This would allow us to see issues that affected users regularly, in contrast to receiving reports flavored by highly memorable but rare events.
2. We wanted to make the user constituents feel "heard," and showing up at their office would be a powerful way of ensuring that. It is an Agile precept that having face-to-face contact with people is the best way to gain support and trust and build cooperation among teams and team members.

Our client had an internal usability lab and they wanted to know if they could do a traditional usability study instead of making field observations. For the following reasons, it did not seem likely that those services would be a better option:

1. Traditional usability research tends to be expensive in terms of time and dollars because of the need for a controlled lab setting. It would be costly (travel and time lost from work) to transport dealership users to the usability lab; less costly for us to travel to the users.
2. The number and range of potential problems that users were reporting was widespread. It was not likely that sufficient representative scenarios could be identified and created within the time limits.
3. Traditional usability methods require testers to assign tasks to the participants. However, formal usability testing wasn't entirely appropriate, as some of the problems were likely to go beyond our application and involve other systems and organizations in the dealer's process.

The idea of contextual research was a new venture for the client's usability lab but after discussing the options and the approach, they agreed that it could work and offered to participate.

6. Client Perceptions at the Start

At the outset of the project, it was clear that our client had already hypothesized about how to improve the application. They were prepared to hear results that confirmed their perceptions from the survey and validated their commitment to allocate most of the 2007 budget to solutions. Specifically they expected to hear the following:

1. The dealer users disliked the application, and wanted a total, or near total rewrite.
2. They wanted strict workflows and wizards built in to the application.
3. Performance was a major issue but would be out-of-scope and off the table during our visits due to a network monitoring and trouble-shooting effort already in progress.
4. The dealer information module was seldom used and further investment could be discontinued.
5. Training materials provided by the corporate office were effective and used often.
6. Dealer users felt they were part of the loop and received timely information about new features, bugs, workarounds, and special alerts.

7. Project Strategy

We recruited a member of the client's usability lab and a client manager responsible for dealer relationships to be part of our observation team. Our observations team also occasionally included "guest" observers from client management. This was very useful in obtaining buy-in and facilitating scheduling.

Our client chose representative dealerships from their five major U.S. regions: West, Texas, Southeast, New England, and Midwest. We had advised them to choose from five to seven dealerships that would give them a broad cross-section of their overall user base. The client chose the specific dealerships that we would visit based on their knowledge of dealership strength and diversity. Five to seven sites were sufficient as that covered approximately 10% of existing dealers and we could generally plan on observing two people per site. In Contextual Inquiry, the sample size can appear to be small because general patterns are usually quick to surface.

In the end, we carried out 11 observations total. This averaged out to about 2 observations per dealership, though our busiest dealership had 4 observations.

In general, we spent one workday at each dealership and either flew to the next dealer the next day or traveled to a central location to work up our team notes and observation models before visiting the next dealer.

The team focused on a single role within each dealership: the Credit Manager. This is unusual for Contextual Inquiry since the prescribed practice is to observe any and all roles that are impacted by the software or process, however minimal. In our case however, the goal was to define a prioritized action list for one application and, for security reasons, only Credit Managers and senior management use that application at dealerships.

8. Project Methodology

Our four person team was split into two distinct pairs, one member playing the facilitator role, and the other being a note-taker. Splitting the team into pairs allowed us to maximize observations at each site. The teams then re-grouped to carry out modeling activities so all members would have a view into what happened in all of the observations.

When our team arrived at a dealership, we held an initial 15 minute meet-and-greet session in the dealership's boardroom. This session was prearranged and involved the team members, as well as all concerned parties at the dealership – credit managers, territory sales managers and senior executives. During this session we explained our method, and made sure that the participants understood that they would be able to carry out their workday as they normally would. We took questions and established a sense of comfort in the participants.

After this introductory meeting, we broke up into observation pairs and went to work with our participating Credit Managers.

In each pair, the facilitator followed a few basic rules to conduct the observation:

1. **Utilize a Master-Apprentice model** – In this observation model, the participant is treated as the master of his or her domain (even if the observers are familiar with the domain). Observers are to act as if they were trainees learning the participant's role, and the participant was to teach the observers about the job as they carried out their normal work.
2. **No leading questions[4]** – Questions that suggest the facilitator or note-taker already knows how the application works are not allowed. Questions that suggest we know or expect to see a certain sequence of steps are not allowed. It is important to take care that we word our questions so that participants will not be encouraged to give us an answer they think they should be saying or that we want to hear. Participants must feel free to be open and honest; allowing them to speak their mind without guidance is vital to the method.
3. **Let the work proceed as normal** – If a phone call comes in, the participant should pick it up (assuming that is normal practice). If they need to walk to the photocopier on the other side of the office, they should feel free to do so. No matter what, the fact that they are being

observed should not hinder a participant from carrying out their job. Only by seeing the variety of steps and interruptions that occur in a user's day can we get to an understanding of how well or how poorly software applications and tools fit into their process.

The note-taker's role is equally difficult. Note taking includes recording the details of what the user did along with all their motivations, either directly stated by the participant or inferred through observing.

The note-taker has to judge when they should fade into the background to minimize distractions and when they should step forward to keep momentum going. They must be sensitive to the flow of conversation and be prepared to step in with a clarifying question or redirection if they sense the user has glossed over a point. Sometimes the interaction between facilitator and participant starts to become rote-like and veer toward a recitation of policy vs. a demonstration of what really works; a sensitive note-taker can help to re-invigorate the dynamics.

Following each observation the team would take some time to get to know the participants and client stakeholders in a non-work setting. Normally this occurred over lunch at a restaurant. Getting to know participants informally was a good way to get additional perspectives about the software and processes, bring unobserved issues to light, and put a human face on the development team.

Following each observation, the note-taker typed their notes and distributed them to the team. We used these notes extensively in the creation of our affinity diagram, which will be discussed later in this paper. After each observation, we also created models of the work that we observed. For each participant we completed five types of models, as described in Beyer & Holtzblatt's Contextual Design [1]:

1. **Flow Model** – Describes how information flows between people and highlights observed break points or interruptions. In this model, we detail *who* the participant communicated with and *what means they used* to do so.
2. **Physical Model** – Describes how the participant's workspace and area were organized. This goes beyond a simple floor plan, and describes where, exactly, information sources can be found.
3. **Sequence Model** – Describes the order in which work occurred and highlights observed break points or interruptions. This is very similar to flow charting techniques, and provides a simple timeline for the participant's work.
4. **Artifact Model** – During the observation, artifacts (or copies of artifacts) that the participant interacts with are collected. The artifact model describes in detail how each artifact was used.
5. **Cultural Model** – Describes the workplace culture present that impacts how the participant carries out their job.

We did not begin analysis until we had completed all out site visits. During analysis, we created consolidated versions of each of the models in order to see higher levels of abstraction stemming from the data. To consolidate our models, we looked for similarities in each of the types of models among all participants. By doing this, we were able to cut out events that only took place in one or two observations, and focus on events that occurred many times.

After completing our consolidated models, the team carried out an Affinity Diagramming exercise using the observation notes. The Affinity Diagram allowed us to see the totality of our data, and helped us to abstract findings based on different users in disparate settings.

This exercise was relatively simple, but required a fair amount coordination. First, the team culled through the observation notes and pulled out the significant details, line-by-line. Each of these details was then printed as an individual square on printable sticky-note paper. In all, we had several hundred individual sticky notes. Each sticky note was coded to identify the source observation and participant for the note.

With all of the sticky notes ready in a randomized pile, the team began a series of sorting passes putting similar notes together until we were able to categorize headings and sub-headings for each grouping. We went through many rounds of sorting, taking about two days in all.

The Affinity Diagram turned out to be a key element of our success, helping us to see concrete issues that could be responded to by our client and brainstorm possible responses. It also created a lot of buzz in the office because it gave our data and process high visibility.

From here, we could define a prioritized candidate story list for consideration by our client and sketch out mock-ups of possible interface changes for the development team.

We concluded our analysis phase by a prioritized story list for 2007, plus the supporting notes, consolidated models digitized affinity diagram and a final report authored by the ThoughtWorks and client leads.

9. Findings

Despite our client's original beliefs, we were surprised to discover that the application was well-liked by most of the participants. For the most part, the things that the client thought would be problematic did not reveal themselves as major issues.

As expected, a major issue for the participants was the "performance" of the system. As noted, this topic was specifically out of scope due to an ongoing program to address network and environmental issues. However, we discovered that the Credit Managers were using this word to mean something different from the traditional systems meaning.

Participants who used the word "performance" would do so in the context of pointing out inefficient navigation or redundant interactions; for example, having to click through several "Are you sure?" warning messages when trying to save a required data entry screen. With this insight, we were able to recommend relatively simple navigation changes that would improve the users' sense of the system's performance.

We were also surprised to discover the participants really liked the flexibility of the system's unstructured workflow. Unlike headquarters, the dealership Credit Managers did not work in silos, one person handled every task within the application.

This was an interesting "aha!" moment. Agile methodologies are based on the concept that the users know what they want, and they should work closely with the development team to build it. But in practice higher-level stakeholders often fill in as proxies for the end-users, assuming they have grown up from the ranks or are in daily communication with the ranks and so can speak knowledgeably for the users' interests.

Our research discovered the opposite; generally, the stakeholder views were disconnected from the end-users. Real user needs can be quite different from the things a stakeholder might specify, and our experience on this project (among others) showed us that a great way to discover these needs is to study users in context.

Simply by watching users at work for a few hours, we were able to take home findings that prevented the client from making expensive and unnecessary changes like wizards and workflow.

We were also surprised to discover that the Dealer Information module was a popular feature and the dealers ran reports from it almost daily. Instead of a candidate for removal, this module turned out to be a high priority area with several opportunities for enhancement.

Finally, all dealerships reported or revealed enough disconnection from corporate training and communication channels to merit further review of existing processes.

10. How things turned out

In general, the final report and story list were well received by the client and participants. All of the top recommendations and stories were scheduled for development and implemented in 2007-2008 with the top priority items rolling out in mid-May 2007.

Also, the results of the study encouraged the client to invite users to a two-day dealer conference where proposed changes to the application were "sneak previewed" and further feedback was received. Since then, additional communication improvements have been implemented including regular feedback panels, on-line video training materials and a hotline discussion thread.

But the transition from recommendation to implementation had some hitches. The client managers who commissioned the study handed our recommendations to their business analysts but didn't include the supporting detailed notes and models. The client BAs were thus charged with solving problems they hadn't discovered and began to draw up initial designs that made sense from their point of view "at headquarters" but were off-the-mark per the dealerships' concerns. We had to do some emergency level-setting and circulation of materials to get everyone equally informed.

11. Things to do Again

If you feel that carrying out Contextual Inquiries could help you or your clients, here are things that worked out well for planning, scheduling and executing our activities.

1. Aim for *qualitative* as opposed to *quantitative significance*. After eight or so users, you will begin to hear and see similar data. When repeating begins, the observations become less valuable.
2. Follow up with the participant and allow them to do a little venting…they will always want to complain about some things that didn't come up. But remember, in the analysis phase the "vented" information should be taken with a grain of salt. Observed findings always take precedence.
3. Do not make any promises about things you will fix. At the end of the day this will be the client's call and not following up on promises

will leave the participants jaded when future releases show that their pet problems were not everyone's big problems.

4. Display your modeling in a visible location to create excitement and interest about the research.
5. Have one or more (but not too many) client staff on the team and have them play the role of coordinator to schedule visits, host meals with the participants, and facilitate collecting metrics and reports. Sometimes, the scheduling effort can be an entire project in itself: having a client team member assigned to this task makes it immensely easier.
6. Educate a new team member with a training session (about 2 hours can be enough) where they can watch a pre-recorded observation session and then compare their notes to example models and notes. A quick affinity diagramming session is a good way to get oriented with the practice.
7. Finally, involve the client team as much as possible. As observers, client staff act as witnesses and give legitimacy to the effort within their organization. It is hard not to evangelize these methods after trying them. There is something about seeing real-world problems in person that gives a person the feeling they can drive a change for the better..

12. Gotchas and Things to Change

1. Complete notes and models immediately after each observation. The contextual information each observer witnesses slips away within minutes of leaving the participant. It is better to schedule a working day on-site (or at hotel) after doing an observation rather than rely on people to go home and mail it in.
2. Be sure an analyst and/or development lead who will be part of the project team is part of your research team. You need them to ensure that lessons learned in research persist to implementation.
3. Sketch out any major UI changes or additions when handing off story ideas to the client. And propagate descriptions and results of the study to all levels of stakeholders at the client; especially the stakeholders who will have hands-on involvement in the development work.
4. Hand-pick the team as much as possible. A traditional Usability Tester did not work out well in our case because of preference for lab-based procedures. Junior analysts are probably not a good choice either, since they are too likely to focus on the mechanics of the applications and miss motivations, goals, behaviors, and business culture.
5. Management sitting in as guest observers is good for buy-in but don't expect them to capture low-level details in their notes. It is likely they will only write about the big picture. Also, they may inadvertently introduce biased or leading questions into the observation. Be sure you clue them in to the rules and have a back-up note-taker present.

13. Summary

Given a situation where usability needs are not being met and the feedback is unclear or conflicted, Contextual Inquiries are an effective method for discovering user-centered requirements. Relying on self-reported feedback and the word of client stakeholders can lead to the creation of solutions that are not necessary or just plain wrong for everyday practice. Taking even a small amount of time to do on-site observations can yield information that will allow your team and stakeholders to understand the real problems in the field. It is an investment that yields good will, improves communication, and can save your stakeholders time and money.

14. References

[1] Beyer & Holtzblatt, *Contextual Design*, Morgan Kaufmann, San Francisco, 1998.

[2] Holtzblatt, Wendell, Wood, *Rapid Contextual Design*, Morgan Kaufmann, 2004.

[3] Loftus, E., *Eyewitness Testimony*, Harvard University Press, Cambridge, MA, 1996.

[4] Young, I., *Mental Models: Aligning Design Strategy with Human Behavior,* Rosenfeld Media, New York, 2008.