Using Participants' Real Data in Usability Testing: Lessons Learned

Todd Zazelenchuk
Intuit, Inc.
2525 Garcia Ave.
Mountain View, CA 94043 USA
todd_zazelenchuk@intuit.com

Sara Sazegari
Intuit, Inc.
180 Jefferson Dr.
Menlo Park, CA 94025 USA
sara_sazegari@intuit.com

Kari Sortland
Intuit, Inc.
180 Jefferson Dr.
Menlo Park, CA 94025 USA
kari_sortland@intuit.com

Mark Keavney
Intuit, Inc.
2525 Garcia Ave.
Mountain View, CA 94043 USA
mark_keavney@intuit.com

Alex Genov
Intuit, Inc.
7535 Torrey Santa Fe Road
San Diego, CA 92129 USA
alex_genov@intuit.com

Abstract
In usability testing, we place great importance on authentic tasks, real users, and the appropriate fidelity of prototypes, considering them carefully in our efforts to simulate people's real-life interactions with our products. We often place less importance on the data with which we ask participants to interact. Commonly, test data are fabricated, created for participants to imagine as their own. But relating to artificial data can be difficult for participants, and this difficulty can affect their behavior and ultimately call our research results into question. Incorporating users' *real data* into your usability test requires additional time and effort, along with certain considerations, but it can lead to richer and more valid usability results.

Keywords
Usability testing, engagement, methodology, user data, real data, scenarios, privacy, security.

ACM Classification Keywords
H5. Information interfaces and presentation. Human Factors.

CHI 2008, April 5 – 10, 2008, Florence, Italy
ACM 978-1-60558-012-8/08/04.

Introduction

In usability testing, we strive to make the situation as realistic as possible. First, we craft the scenarios and data to be representative of users' actual contexts. Then we carefully describe them to the test participant, sometimes with artifacts to help make them come alive. All of this is done to get the participant to role-play a user of the software. But no matter how well we set the stage, there is a difference between someone user-testing the software and that same person actually using it. This difference can result in usability problems being missed. It can also lead to *false positives*—difficulties that look like usability problems during the test, but are actually the result of participants' failure to understand the scenario or the data that we have invented for it.

One way to reduce the artificial aspect of usability studies and to make their results more valid is to use the test participant's *real data* in the study, rather than data created to simulate the real thing. In other words, rather than making up reasonable approximations of users' banking transactions, annual income statements, or medical claim details, you find ways to incorporate users' actual information into the test. In this article, we identify the key considerations for conducting *real data* studies, illustrating a few variations on the theme in three mini case studies featuring Intuit products. We point out the land mines associated with *real data* testing, and offer guidance on how to navigate them as you take your own usability testing one step closer to the real thing.

Benefits of Using "Real Data"

Incorporating users' *real data* into your usability studies often requires additional effort. Thankfully, there are clear benefits that can make it worthwhile.

The primary benefit is increased ecological validity [1]: incorporating users' *real data* better approximates the real-life situation that is under study. Additional benefits include:

- Participants are no longer required to bear the cognitive load of remembering an artificial scenario or recognizing the *invented* data presented to them.
- Participants' knowledge of the data, or lack thereof, is realistic – there is no need to speculate on what a participant "should have" or "would have" recognized if the scenario were more true-to-life.
- Participants are more emotionally engaged by their own data and, therefore, tend to immerse more fully in the test. This often results in a better understanding of the product's strengths and weaknesses, and richer feedback to the design team.

Considerations in Using "Real Data"

While there are numerous benefits to *real data*, there are also challenges to be considered. Three primary considerations include:

GETTING THE DATA

In order to run a *real data* study, you must be able to get your participants' *real data*. Depending on the nature of your product, this could be as straightforward as asking recruited participants to bring their data with them, or as difficult as extracting data files from multiple backend systems in advance of your study,

and then mapping and importing the data into your prototype. In the case studies that follow, we share examples from both ends of the spectrum, along with lessons learned for making this part of the process a success.

SCENARIOS, TASKS, AND DATA ANALYSIS

With artificial data, it is possible to predefine the 'happy path' for each task and evaluate participants' performances accordingly, but with *real data*, things are less simple. Depending on the participant, the same task may involve different amounts and types of data; in some cases, the task may not apply at all, or may need to be done multiple times. For each case study, we discuss how we constructed our scenarios and analyzed our data to account for these issues.

SECURITY AND PRIVACY

Non-disclosure agreements and consent forms for videotaping sessions are commonplace elements of most user research studies. *Real data* studies, however, may require additional processes regarding the storage and disposal of participants' sensitive personal information. In the following studies, we describe our experiences regarding the security and privacy of participants' banking, tax, and medical information.

Case #1: Online Banking Solution

Background

During 2006-2007, the Financial Institution Division of Intuit created a new online banking product, FinanceWorks [2]. This product was designed to serve customers of financial institutions (FIs) by allowing them to carry out their personal finance management directly on their FI's website.

The team had user-tested the product extensively, but wanted to move beyond the artificial nature of their user research. FinanceWorks had not yet been evaluated within the context of an FI's website, nor with users' actual banking data.

The team combined the *real data* methodology with a field study approach for added realism. Fourteen credit union customers in Salt Lake City, UT participated in the study, which took place in their homes where they typically did their online banking and paid their bills. Participants' own banking data were incorporated into a prototype designed to look and act like the credit union website, and presented to participants on a team member's laptop.

Anticipating that unique data for each participant would introduce some new challenges, the team performed two pilot sessions. The pilots generated more than the usual number of changes to the prototype. Restricting the pilot phase to one session per day allowed the team to incorporate these changes into both the prototype and test plan after each session.

Getting the Data

Getting users' *real data* was a multi-step process. The first step was securing a financial institution partner to work with. Once the credit union was on board, the team recruited its members and obtained permission to access their posted bank transactions and electronic bill pay data. Participants' data files were then extracted from various backend systems, and mapped and uploaded into the usability test prototype. This process required coordination between 5 different people in different departments!

Scenarios, Tasks, and Data Analysis
Early on, the team recognized a problem in applying a protocol where all participants received the same tasks: different participants paid bills in different ways, and their bank transaction data didn't indicate how each bill was paid. To overcome this problem, the team created a generic version of the moderator's guide, phoned each participant to remind them of the session and to understand how they paid their bills, and customized the moderator's guide accordingly.

Since the set of tasks varied with each participant, data analysis was more complicated. For example, a given participant might have completed a given task multiple times, or not at all. Therefore, an average success rate was calculated per participant per task, and then those scores were averaged to find the overall completion rate for a given task. This made data analysis more time consuming, but helped the team to accurately discern usability issues for different tasks.

Security and Privacy
Security and privacy issues were at the forefront of the team's thinking. Users' data were stored in encrypted files for the entire study. The data files were maintained only during the data analysis and destroyed after the data analysis was concluded.

Similar precautions to storing participants' data files were applied to participants' video files, which were recorded using Techsmith's Morae [3]: videos and video highlights from the study were encrypted, stored on an external hard drive and locked in a cabinet.

The team found it helpful to be very explicit in addressing participants' security and privacy concerns. In addition to the standard consent form's description of how participants' data files would be protected and used for research purposes only, participants were clearly informed how their personal data would be used in a separate consent form.

Benefits
By customizing each session with participants' posted transactions and bill payments, participants were better able to understand both the protocol and the prototype than in previous lab studies. This increased the team's confidence that it was discovering the product's true usability issues.

In addition, the real data served as a useful check on past research. Our initial contextual inquiries had found two common ways of paying bills, and a third way that was much rarer. The team had therefore designed and tested for the common bill payment methods in all previous usability tests. However, in the real data study we found that the "rare" method was actually quite common, and that users who tried to pay that way using our product experienced significant usability issues. This finding led the team to redesign the interface and subsequent test protocols to better accommodate this method.

The study also revealed a serious usability issue that had never come up in previous tests. The product allowed participants to do tasks on their bank's web site that would normally be done within desktop software such as Quicken (for example, users could enter a check that had not yet cleared to account for that money in their balance). Participants in previous studies had had no trouble with this concept, but in the real data test, many were reluctant to even attempt the task. In this

more realistic situation, participants assumed that there would be no way to do the task on their bank's web site, that they "wouldn't do that here." This discovery led the team to redesign to increase discoverability and to better educate first-time users about what the product could do.

Case #2: Personal Tax Products

Background

Experience design teams within Intuit's Consumer Tax Group (CTG) are responsible for delivering easy to use *Do It Yourself* tax software. For nearly two decades, CTG has achieved this goal with its flagship product, TurboTax [4], a tax software product designed to assist consumers with filing their personal tax returns.

In May 2007, CTG conducted a usability study to assess the improvements in the Small Business Income and Deductions section of TurboTax between the 2005 and 2006 versions of the software. The sessions were conducted in the labs at Intuit's San Diego campus. Each participant completed tasks with both TurboTax 2005 and TurboTax 2006, presented in counter-balanced order.

The team decided that it was important to use participants' actual tax scenarios and data for two reasons. First, given their interest in assessing customers' comprehension of the questions, instructions, and results presented by TurboTax, team members wanted to avoid any possible confusion or misunderstandings caused by tax-related data that was unfamiliar to participants. Second, the team wanted to eliminate the interference caused by the study moderator stepping in to clarify the scenario or data for participants.

Getting the Data

To obtain users' *real data* for this study, a requirement was added to the recruitment screener that participants bring their personal tax documents to the session. Those who agreed to participate were asked to bring enough data with them to be able to accurately complete the Small Business section of TurboTax without having to make up numbers or work from memory. Some people declined participation because of the "bring your own data" requirement, making recruitment for this study somewhat more challenging than usual. More specifically, a greater number of potential participants had to be contacted, assurance had to be given regarding the security and privacy of their details, and the team chose to pay participants more (~30%) to compensate them for their additional efforts.

Scenarios, Tasks, and Data Analysis

Participants were given a single scenario: "Imagine that you have decided to do your taxes yourself, have bought TurboTax, installed it on your computer, and are now going through the initial sections of the program." They were then given the concrete task of completing the Small Business Income and Expenses section (Schedule C) using their own situation, documents, and data. Differences in the amount or types of data between participants did not impact data analysis, since the study design was within-subjects. That is, since each participant completed both versions of the software, they acted as their own control for the purpose of the results analysis.

Security and Privacy

The team's "bring your own" method of obtaining *real data* kept security and privacy issues to a minimum:

since participants did not forward any personal data in advance, there were no issues with secure data transfer or pre-study storage of the data.

At the beginning of the study, participants were informed of the audio-video recording and were asked to sign standard Non-Disclosure and Informed Consent forms. They were assured that the video records were strictly for research purposes and would be destroyed. Those participants that were opposed to it opted out of being video-taped.

Benefits

There were several benefits to having participants use their actual financial and tax data. First, the team's goal of reducing participants' cognitive load related to artificial tax data appeared to be met: participants were completely immersed in preparing their own taxes and did not ask any clarification questions about the scenarios or tasks. A second benefit was that users' *real data* led to a higher level of cognitive engagement in the study than in previous studies. For example, when faced with a question about their income on one of the screens, participants made sure they understood the question and answered it correctly, fully aware that a wrong answer by them would lead to inaccuracies in their tax return. In contrast, in studies where participants have used fictional data, they have been more prone to make up answers and numbers and less motivated to be accurate. Finally, the participants were more emotionally engaged than in previous studies. This was evident by how they compared the tax outcomes at the end of the study with their real tax outcomes from the previous year, and by their concern with any unexpected discrepancies. As a result, the development team was more confident in the validity of the usability results.

Case #3: Managing Healthcare Finances

Background

In 2006, Intuit created a business unit dedicated to improving the consumer experience of managing healthcare finances. In the same way that Quicken [4] helps consumers manage and track their personal finances, Quicken Health (QH) [5] helps consumers manage and understand their healthcare finances.

During its first year of development, Quicken Health underwent multiple lab-based studies as part of an iterative design-test-redesign process. These studies involved fabricated scenarios and healthcare data that participants were asked to imagine as their own. Unfortunately, participants often glossed over the fictional healthcare data, unable to relate to them in the same way as their own medical information. Before long, the design team started to question the validity of their research findings. For example, in the lab users easily understood and accepted their financial obligations, but the team doubted that in real life, users would be so nonchalant with their own money and health care.

To overcome these problems, the team decided to apply the *real data* methodology to their next usability study. Seven consumers from a single health plan took part. Lab sessions were run at Intuit's facility in Mountain View, CA where participants used a prototype of Quicken Health to complete a series of tasks.

Getting the Data
Collecting and assembling users' *real data* for the study was time consuming. As part of the recruiting process, participants agreed to mail in copies of their medical claims paperwork (Explanation of Benefits, or EOB) several weeks prior to the study. They also completed a short questionnaire, sharing their health history and benefits, including their past medical conditions, family names, and insurance deductible data. The team's goal was to gather as much relevant data as possible to reflect the data that would ultimately be obtained automatically by the real product, thereby creating the most realistic experience for the study.

Collecting the data took three weeks. As participants' data were received, each individual service and medication was typed into a spreadsheet, which was later used to populate the test prototype. Once all users' data had been entered into the prototype, the Quality Assurance (QA) team reviewed the integrity of the data and overall accuracy of the prototype.

Scenarios, Tasks, and Data Analysis
The research team began by familiarizing themselves with each participant's data, and then customized each session guide per participant, allowing for more personalized scenarios than in previous fictional data studies. This advance immersion also benefited the prototyping team, allowing it to optimize its efforts by building out certain parts of the prototype more than others to take advantage of the available data for each participant.

Security and Privacy
Security and privacy were critical considerations given the sensitivity and government legislation (e.g. Health Insurance Portability and Accountability Act) surrounding personal healthcare data. Participants were clearly informed about the requirement for their personal data in the study, and were reassured of its security via a mutual non-disclosure form.

Storing participants' *real data* received special attention as well, with all data stored in offline systems or on CD-ROMs under lock and key where only certain personnel were granted access. All *real data* (paper and electronic) were destroyed within 30 days of the study's completion, and any presentations involving video clips or screen shots omitted or masked any *real data*.

Benefits
Consistent with the other case studies, the Quicken Health team found the *real data* methodology to provide a level of authenticity that went far beyond fictional data studies. This was most apparent when participants encountered highly sensitive areas such as insurance claim denials and adjustments. The negative impact associated with these topics (e.g. unexpected fees) was heightened for participants because it was their own medical service that was being denied.

In another example, participants in previous fictional studies had commonly breezed through the task of reviewing their medical claims for accuracy. In the *real data* study, however, the team discovered that participants often failed the same tasks due to the way in which physician details are processed by health insurance plans. When participants encountered these incongruences, they routinely believed there was an error in the system. By detecting this problem early, the team was able to ensure that the details ultimately displayed mapped more naturally to users'

expectations. Without the *real data* approach, this problem would have only been discovered after product launch.

Summary

Real data testing is not for every project. It requires more time and effort from the team (and sometimes even from the participants). The cost-benefit ratio can be compelling, however, when one or more of the following is true:

- the way in which the product is used is highly dependent on the user's data (e.g. financial software that supports a range of payment methods)
- the data can be obtained relatively easily, such as by asking participants to bring it with them to the study
- the product is new and the team isn't sure how users will interact with it
- incorporating real data is able to inform not only the Experience Design team, but the QA team as well by discovering issues with importing and presenting data from external sources
- the project is high-profile allowing for additional usability testing resources
- the team has begun to question the validity of fictional data usability test results.

In these cases, *real data* studies can reveal insights that would not be discovered in traditional usability studies, making the effort well worth considering.

References

[1] Brewer, M. (2000). Research Design and Issues of Validity. In Reis, H. and Judd, C. (eds) Handbook of Research Methods in Social and Personality Psychology. Cambridge:Cambridge University Press.

[2] FinanceWorks. http://www.digitalinsight.com/home/solutions.consumer.pfw

[3] Morae. http://techsmith.com/morae.asp

[4] TurboTax. http://turbotax.intuit.com/

[5] Quicken. http://quicken.intuit.com/

[6] Quicken Health. http://quickenhealth.intuit.com/

The authors would like to thank Jon Innes for his encouragement to share the ideas contained in this paper, as well as the CHI reviewers for their valuable suggestions.