Contextual User Research for International Software Design

David A. Siegel [1] and Susan M. Dray[1]

[1] Dray & Associates, Inc.
2007 Kenwood Parkway
Minneapolis, MN 55405
US

Abstract. This paper, which is based on our many years of experience conducting research in more than 20 countries, examines both the importance and pitfalls of doing contextual field research when developing software, websites, or interactive products for the international market. We examine the ways in which field research gives crucial information that complements what can be obtained from other methods such as usability testing. We identify a number of core challenges in doing international fieldwork and recommend ways to address them. We conclude with at least one case study.

1 Introduction

User centered design involves the use of a range of methodologies to ensure that a user focus is introduced into the process of planning and designing products and software. Increasingly, companies have begun to recognize that without a disciplined user-centered design effort, their products have too high a risk of failing in the market because product planning and design decisions will be too heavily influenced by their enthusiasm for their own ideas.

Globalization requires an intensified focus on user-centered design. The need for systematic efforts to understand the user becomes even greater when products are being designed for users in another country and culture than your own. Beyond the usual difficulties we have imagining how other people differ from ourselves, people working in one country or culture can have difficulty imagining the enormous number of ways in which people can differ internationally. As a result, not only does the risk of design error increase, but, in particular there is a greater deep level at the conceptual level of product planning, such as a basic mismatch between the concept of the product and the usage context. The size of the investment at stake when doing business internationally can also contribute to making failure exceptionally costly.

Fortunately, many companies are realizing the need for international user-centered design (UCD) work. However, this probably most often takes the form of usability testing, which by itself is insufficient. Usability testing does not directly address the issue of contextual fit. Even designing a good usability test requires knowledge about

patterns of usage and typical user goals which may differ radically from country to country.

For example, without knowledge of local usage patterns, user expectations, and goals, you can not design task scenarios that will be meaningful to people. Two problems can result from this. You may design task scenarios that assess whether people can accomplish the tasks you ask them to do but which may have little to do with how users will ultimately value the experience the product provides for them. In more extreme cases, users may not even understand what the task is asking them to do. Fortunately, you may be able to derive indirect useful information from this, because it can be a clue pointing to an area of fundamental mismatch between the product team's assumptions and the local reality.

2   A Case Example

In order to take root, technology innovations require a good fit with existing behavioral patterns, social and economic conditions, and technology infrastructure. The overall dynamic created among these factors can differ from one country to another in ways that powerfully affect the adoption of technology. One good example that many people are well aware of is related to the gap between the adoption of mobile phones and in particular SMS functionality between the US and other countries. Although this gap is decreasing, it still exists. Many visitors to the US have been surprised by this, and have tended to wonder how people in such a technologically advanced country could be so "behind" in adoption of a functionality which seems to be the wave of the future elsewhere. One can not address such questions without understanding all of the contextual factors that determine the appeal of and need for a technology in contrast to the available alternatives, the fit with people's lifestyles, and the obstacles to adoption including economics and technical infrastructure.

A similar question has arisen in regard to adoption of smart card technology. There have been many failed attempts to introduce smart cards (containing a computer chip as opposed to simply a magnetic strip) in the US. In the early 1990's, one of us assisted in a study commissioned by the US Office of Technology Assessment (a government agency whose mission was to assess potential policy implications of new technologies in order to inform relevant legislation) regarding the adoption of smart cards in the French health care system, to evaluate their potential for the US. At the time, there were a number of seemingly promising initiatives underway in France. Some projects focused on using the cards to store health insurance benefit and utilization information. Another project focused on using them for tracking dialysis information for kidney patients. The applicability of models like these to the US was in question.

Many of these factors that seemed to contribute to the fit of these approaches in France have probably changed significantly due to the evolution of the
Web since that time. However, what is relevant for our purposes is that comparing the potential attractiveness of these approaches between France and the US required

understanding the entire context. Following are some of the specific factors that seemed to create a fit in France that were not the case in the US.

The French government perceived this technology as promising and as an area where France had a lead internationally. It therefore provided a great deal of support and subsidy for these initiatives. France is a more centralized country than the US, and so it was likely national government policy would have a more direct effect on this. For example, to pursue larger policy goals the national telephone company could simply mandate the use of these cards as stored valued cards to replace coins in public phones.

The fact that use of smart cards as stored value cards was already well-established was likely to facilitate public acceptance of their use in other domains. For example, in addition to being used for phone cards, they were also used in credit cards to handle authentication and management of daily limits, without requiring authorization from a central data center. However, there were factors that made these uses appealing in France and that did not apply in the US. In France, higher telecommunications costs were an obstacle to the kind of credit authorization process used in the US, where merchants used terminals that communicate over telephone lines to get credit card authorization. Rates of credit card fraud were significantly higher in France than in the US, which made it economically worthwhile to create the infrastructure for more powerful security on credit cards, which could be supported by chips as compared to magnetic strips.

The structure of the health care system and of the mechanisms for health care financing differed radically between France and the US. In France, there were economies of scale that did not apply in the US. For example, France had a national single-payer health insurance system applying to most of the population, and the majority of clinics participated in it as health care providers. This made it reasonable to consider installing a single system of card readers in clinics nationwide, and to use a single format for information records and benefit information that would apply across the vast majority of patients and providers. In contrast, in the US, the system was much more fragmented. There were (and still are) a large number of different insurers providing health coverage. Some were national and some regional. Different clinics even in the same area participated with different subsets of the plans. Each plan requires its participating providers to go through different administrative procedures. Although medicare is a nationwide, federally-funded program for retires in which most clinics are participating providers, medicare enrollees could represent widely differing percentages of the case load for different providers.

3   Challenges and Solutions

Carrying out research internationally to understand the constellation of local factors likely to affect adoption of a technology presents a number of basic challenges. Some of these are particular to international research, while some are like the challenges of any contextual research, only rendered more difficult by the international setting of the research. In this section, we review a number of these and offer recommendations about how to address them.

3.1 Defining the Focus

Field research typically adopts a broader focus than usability evaluation, because it has to be open to a wide range of things that may influence the design. This is true regardless of whether you are trying to provide input early in the design process or are evaluating the "fit" of the product in the users' real life circumstances. The challenge is that how you plan the study will still influence what you see. This includes how you define both the people and contexts you are going to study. This requires proactive planning based on the research focus.

Some people believe that being overly specific about your focus is a mistake. Their argument is that this narrows your focus of attention excessively, based on your preconceptions, and blinds you to things that you would not have known to look for. We disagree with this. If you do not have a planned focus, your attention is more likely to be too diffused, so that you only capture vague impressionistic information. Planning the focus in detail will allow you to plan all aspects of the methodology in a targeted way, so that you have the opportunity to drill more deeply into interesting observations. One risk of failing to do this is that you are actually more prone to being influenced by stereotypes without realizing that this is happening. Another advantage of having an explicit focus is that it can give you a baseline against which unexpected observations will actually stand out more dramatically.

There are two main approaches we take to developing a useful focus. They are logically connected to each other. The first is to do an intensive brainstorming exercise, including the product team as much as possible, to identify as many of the contextual factors that may influence how people are likely to respond to the product, and could potentially affect its adoption and usage, and to plan what type of things you might look for as clues to these things. The goal of this is not to make sure you think of every possible factor that might matter. No matter how thoroughly you do this exercise, you will almost always discover relevant cultural and contextual factors that you did not anticipate. This is an example of where the conscious development of the focus helps increase the contrast between what you were looking for and what you find. This helps to more clearly reveal where the gaps in your initial mental model were.

The second method for developing a focus is to identify the team's implicit assumptions. It is almost inevitable that you will find the team making assumptions about user populations and their local context. No matter how early in the development process the team is, and no matter vague their product plan or how open they are to new information, it is extremely unusual for them not to have some mental image of their future product possibilities. This image includes assumptions about the people they are trying to appeal to, the potential benefits of the product in the lives of those people. Whether they are right or wrong, these assumptions should influence the focus of the research. Since the goal of the research is to advance the team's understanding of their users and their contexts, these assumptions should be made as explicit as possible, so that they can be treated as hypotheses to validate or invalidate.

3.2 Defining Who and What to Study

Developing a screener for international research is very challenging, because the criteria that you operationalize in the screener reflect your working assumptions about how to define your target audience. Assessing those assumptions themselves may be among the most important things you need to focus on in the research.

When doing field research in one's own country, one can be guided by a great deal of implicit knowledge about local conditions that helps one make more accurate initial guesses about what will be relevant, and where to look for that information. Although it is still challenging, one has an easier targeting the research to provide an opportunity to observe those things. When working internationally, we typically know less in advance about the factors that will turn out to be most relevant, and we therefore have to adopt a broader focus, with higher level questions. In local studies, we often can focus on understanding how certain groups of people do certain types of tasks. When we do international work, especially when we work in a culture different from our own, we often have to wrestle with whether the basic working definition of the groups and their contexts is even relevant.

Screener items tend to be simplistic efforts to capture much richer concepts using easily administered and easily scored items. For example, one may use an income level criterion in an attempt to target middle-class participants. However, the concept of "middle class" is much richer and more complex than can be captured in a simple numerical criterion. The author of the screener may imagine the product appealing to someone who fits an image of a middle class lifestyle, personal history, and experiences associated with a certain income level in his or her own country. However, this constellation of factors is itself culturally dependent.

Adjusting the income criterion to the fit the income range of the target country is not sufficient, because the entire structure of income distribution may be different in the target country from what it is in the product team's country. For example, in one case, we observed that people who fit the local definition of "middle class" often had personal servants, which would not have been true in the US. As another example, if we recruit people who have owned a personal computer for a certain length of time from an economically developed country and from a developing country, the latter person is more likely to be exceptional in a variety of ways in relation to their own context, such as having characteristics of an earlier adopter or having a special motivation to buy the computer that the criterion would not imply for someone in the developed world.

Researchers often try to deal with this by using multiple criteria to better target the people they have in mind. However, this can produce difficulties as well, because the relationship among the criteria can be different from what you expect. For example, in one project that we worked on several years ago, our client wanted to target people in a lower income range who had dial up connections on their home computers. It turned out that this definition of a category of people was filled with contradictions for that local context. People with home computers, who had telephone lines to their homes, and who were willing to use them for an internet connection despite that fact that time charges made this expensive, tended to be affluent. This does not mean that one could not find less affluent people who fit this description, but it does mean that those people were likely to be unusual in other ways.

When doing international research, we often encourage our clients to keep the recruiting criteria simple, rather than specifying them very precisely. This is because the more you specify, the more likely you are to be building your definition on mistaken assumptions about how to define what you are interested in given the local context, and to end up with a sample that is exceptional in some way that you are not aware of. It may be possible to find people who meet your tight criteria, but you will have little idea of how exceptional those people are. Simpler criteria have the advantage of allowing you more opportunity to learn how much heterogeneity there is within a particular group of interest, and what constellations of factors tend to go together meaningfully.

3.3 Balancing Depth with Breadth

Contextual field research is often described as emphasizing depth over breadth, in the sense that it uses small samples, and that you spend a significant amount of time with each informant. However, going "deep" with a participant means that your methodology allows you the possibility of understanding the processes broadly in the sense that you have the opportunity to look at the very wide range of factors that influence the user experience dynamics in the domain of interest, rather than simply focusing on practices most related to the behaviors you hope to support with your technology. For example, if you are considering delivering a service over the Web, it is not enough to focus only on how people interact with and utilize websites similar to yours. In fact, there may be no analogous web tools in their environment. Therefore, you may need to look broadly at factors that influence how they currently fulfill the life functions relevant to that website, because these things will influence whether and how people will adopt a tool like yours.

For example, we did a study related to potential use of electronic tools for coordinating family activities in a developing country. We had to go beyond trying to learning how families attempted to coordinate with each other and how they accomplish this now. We also had to try to understand the vast range of factors in the local context of social norms, rules, infrastructure, etc. that supported the existing practices. How much variability is there in people's activities across days and weeks? What introduces variability? How far in advance are different activities planned? What are the activities that families need to synchronize with each other? When do families expect to come together and what influences how they accommodate each other's schedules? What are the logistical challenges they face in coordinating? What roles do different family members assume in regard to all of the above?

3.4 Dealing with Pressures to Overload the Study

Companies' understandable desire to get the most value from their investment in international research can create pressure to broaden the scope of the study beyond what the resources of time and sample size can really support. Another factor that contributes to this is that international studies often have large numbers of

stakeholders within the company, all of whom wants to be sure that their particular agenda is addressed. As explained above, while contextual field research can provide both breadth of understanding in the sense of examining the broad context of a process, it is unrealistic to think that it can show you every major variation in the ways that process plays out in an entire country. Attempting to cover too much can spread your efforts too thin and ensure that you only bring back superficial information, which will be misleading if it is over-generalized. It is more appropriate and effective to allow each study to have a clear focus, and to assume that you will need to gain information over time from an accumulation of research. Instead of trying to answer every imaginable question within one research study, plan your studies to make sure that you get a view of conditions in the country or of the process you are interested in from contrasting vantage points with each study. This means recruiting samples that will contrast with or complement each other, or planning the study to observe different parts of the process. Planning a series of complementary studies in this way is more likely to deepen your understanding.

For example, one series of studies attempted to understand the reasons that a particular innovative product being introduced into the developing world was not selling well. Some initial studies looked at people who were recruited through store intercepts while shopping for in the general category of that technology product. These studies revealed negative reactions to the product. This raised doubts about the fundamental desirability of the product, which was disappointing because it had been planned specifically in an effort to have a particular appeal to a particular hypothetical segment in the developing world. However, we pointed out that there might have been a built-in bias in the recruiting method. The shoppers found in the stores may not actually have been representative of the target group, unless those particular stores attracted a representative sample of the target group. We did a subsequent study in which we used conventional recruiting methods to find people who were in the target group according to screener criteria and were verifiably in the category of product we were interested in. We then did an accompanied shopping study with them, in which we let them choose the venue for shopping, but also arranged to exposed them to the product of interest. We discovered that these people tended to gravitate to very different types of retail settings than those where the product was being marketed. Also, when they were exposed to the product they had a much more positive response, often choosing it over the competition. This suggested that the problem may not have been in the product concept itself, but in its marketing strategy, which was systematically failing to reach the intended audience.

3.5 Using a Multicultural Team

We strongly believe that international research requires a combination of perspectives, including those of the design team and its culture, along with the local perspective. Different types of information will be visible to people with different backgrounds. For example, a local research team member may be well aware of local conditions that help explain an observation but that might not have been discovered by an outsider during the study. The local person is probably also more likely to know whether a participant's circumstances or behavior are atypical in the local context

(although it can be a serious mistake to assume that people have objective and unbiased knowledge of their own local contexts.) Conversely, astute outsiders brings a wealth of contrasting experiences (especially if they have extensive international experience already) they can raise questions the local would not think to ask, or perceive the significance of things that the local person does not even notice simply because they are so common they are taken for granted.

However, simply sending design team members and engineers to work with local researchers does not guarantee a synergy between these perspectives. Someone on the team needs to be expert in cross-cultural research, interpretation of ethnographic data, and cross cultural communication to bridge between these perspectives and to extract the significance of observations. Whether it is a local researcher or a foreigner, what is important is important for this person to combine skills in understanding the team's mindset and identifying their underlying assumptions, which are often only implicit, and in making sense of local conditions, so that they can identify the discrepancies between these.