Leveraging a User Research Framework to Guide Research Investments: Windows Vista Case Study

Gayna Williams

Principal Program Manager, Developer Division, Microsoft[*]

Abstract. During the development of Windows Vista we had the opportunity to invest in new methods to understand user behavior. We leveraged standard usability methods to work on feature areas during development; however, we had to invent and adapt new approaches to measure holistic experiences. In this area user research methods are evolving, due to the integration of technologies and changes in the definition of a successful experience. While considering the methods that suited our needs, a user research framework was created. This helped us manage investments in research activities. The framework is organized along two dimensions: perspective and time. Perspective refers to the breadth of the experience being considered: 'narrow' defines a focus on an individual feature area or small product area, and 'broad' defines a focus on an integrated experience. Time can indicate either a product cycle or real time. For product cycle most of the research is spent on the evaluation of the designs of the features and experiences related to predicting user behavior for a particular release of a product, whereas real time is our research investment into understanding how products are used in the wild without our intervention. Each quadrant of the two-dimensional framework highlights different research methods and purposes. It's important to realize that the value of the framework comes from the integration of findings that provides a rich holistic picture of our users to ultimately guide product decisions. This paper describes some of the methods that were evolved and created during the development of Windows Vista and their relationship to the user research framework. The methods described in the paper include user experience score-carding, measurement of desirability, and the impact of the consumer adoption program. These methods continued to be used today in the development of Windows 7.

1 Introduction

One challenge in working on an operating system is that it contributes to a computer experience in more than one significant way. It provides stand-alone experiences and it contributes substantially to extended experiences. When developing Windows Vista, the user research team had the challenge of considering how to provide deep insight in particular areas to impact product creation, while also playing a critical role in understanding the quality of the holistic experience. Many parts contribute to the ecosystem that users of Windows experience. Our role was to understand this holistically and to drive that understanding into product development.

---

[*] Previously User Research Director, Windows Vista.

Before diving into the research framework I'd like to provide some context. The Windows organization is a large organization that includes over 5000 people. Windows Vista was not the only product to be produced by this organization as the core components contribute to other products (e.g., Windows Server), or service packs (e.g., Windows XP SP2). The user experience team is a centralized organization and consists of user research, design, and user assistance. During the Windows Vista development cycle these three groups stepped up their accountability to raise the importance in the product experience. For design, there was an increase in demonstrating how design could lead product definition, and also the continuous engagement from product inception through to marketing messaging and branding. For user assistance there was a change in focus from being a team that documents help to work towards the goal of being a continuous publishing group with a data driven content strategy. And for user research we stepped up to consider how to drive accountability for user experience across teams in a holistic way, which is what this paper describes further. I was the user research director of the research team. The team was approximately 24 people in size. The team consisted of 14 user researchers who did much of the iterative work with product teams and also owned particular experiences, plus some researchers also owned particular projects or research methods that benefited the whole team, two anthropologists, one project manager, one product planner (a role focused on identifying opportunities through working with internal and external partners), two data analysts and a small development team (4 people) for building tools and managing the instrumentation projects.

The mission for the team was, "Deliver outstanding Windows client and partner experiences that build upon a deep understanding of people". It is important to understand the deliberate decision to use the word 'people' in the statement. So often in usability we are focused on "the user," defining the user as the person actually using the system in contrast to the customer, the person responsible for purchasing the system. However, we realized that we needed to understand many people within the ecosystem in order to deliver the right experience. We also realized that to succeed in delivering on what people perceive to be the Windows experience required assisting partners to also understand how to make their part of the experience better. For example, most people experience Windows when they purchase a new computer. The original equipment manufacturer (OEM) is responsible for part of the experience, and when a user sees the desktop for the first time it is a joint responsibility of Windows and the OEM.

One option when we started to work on Windows Vista was to map user researchers to particular teams within the Windows Client Organization and then manage their workload in strict alignment with the teams. However because we were organized as a central team we had the opportunity to set our own priorities and focus areas. Our position gave us a unique opportunity to have a perspective across the whole experience. We decided to leverage this position to drive product development from the perspective of holistic customer understanding. Achieving this perspective required us to approach our work differently and invest resources differently. The necessity of getting this rich view shaped the user research framework.

2   User Research Framework

There are many tried and tested research methods for improving usability during a product cycle. It is encouraging that today more user researchers succeed in implementing these methods throughout a cycle rather than being brought in at the end of a cycle to validate decisions or create recommendations where time is too short to respond to them. More recently, challenges have arisen as product teams seek to know more about the emotional experiences of users with their products. Research methods have been evolving to accommodate this need, but when we were working on Windows Vista (2002-2006) these methods were less established than they are now. At best the methods available then were for evaluating finished products, not products in development. How were we to get a sense of the overall emotional experience of Windows Vista three years ahead of the product release?

Another challenge was that the Windows engineering team is very large (i.e., a few thousand).We knew most individual tasks with Windows Vista would require people to use UI elements produced by several teams, who might be in more than one division. We decided to step up to the challenge by creating a list of tasks to serve as a common reference point for much of our research work. More about how we created and leveraged these tasks is discussed later in the paper.

So to tackle the research work for Windows Vista we needed to invest in narrowly focused but deep usability activities that aligned with the product teams, and we had to work in lock-step with the schedule of the product teams. However, to do our jobs well and deliver on an outstanding holistic experience, we knew we had to invest in broader expansive research that tackled some of the challenging new wave of requirements that target 'experience'.

These two types of investment are represented in the first row of the user research framework in Figure 1, which focuses on user research work that aims to predict user behavior with the product when it is complete. I define first a narrow perspective, the investment in mapping research activities to the requirements of product engineering. The second perspective is broad by comparison, extending across engineering insofar as it is encountered throughout the product, or across feature areas as it spans possible feature boundaries. Although the second is a critical investment if user research is to deliver 'experiences,' it had less history to guide us in successfully integrating it into the product development cycle.

To achieve the product, holistic understanding   we went outside the product design cycle to study people currently using Windows. We realized that the current use of the product was influenced by how the previous version had been created, this knowledge provides tremendous value. We invested time in instrumentation, survey techniques, and field work throughout the development cycle, and as a result were always able to learn from current user behaviors as one input to informed decision-making. We were careful not to influence the current users we were learning from by revealing information obtained from other users exposed to prototypes or other information that we were using to help assess the future behavior of Windows Vista users.

		Perspective	
		Narrow	Broad
Time	**Product Cycle** (predicting user behavior)	**Features** E.g., Cognitive walkthrus, Heuristic reviews, Iterative lab testing, Focus groups, Synthesis of research	**User Experiences** E.g., Experience Metrics Desirability, Productivity
	Real Time (understanding user behavior)	**Product** E.g., Instrumentation, Site/Beta visits, Survey panels	**Life** E.g., Ethnographies, Education programs

Fig. 1. User Research Framework

The final perspective that was used to ground our understanding of people and their behaviors was obtained through an investment in collecting ethnographies. In this area we removed the restrictions of considering our product and even our company's technologies from the research brief and focused on audiences and situations that were considered to be of future interest. This work provided a rich context of understanding the world in which our finished products would be situated (or not!).

The latter two perspectives are key elements in the lower row of the framework. This is referred to as Real Time because for the most part we are not influencing behavior when collecting observations by trialing software or scenarios with the users. The lower left cell represents a narrow product perspective, meaning we define the audience we're engaging with by the product we are interested in, whereas the lower right cell is a life perspective as we do our best to observe the situations and audience without making a priori decisions as to which products we wish to see used.

The framework allowed us to consider how to invest resources in tool and method development, and how to invest our research time. Below, I go through the framework in greater detail.

3   Applying Windows Vista User Research to the Framework

3.1   Product Cycle – Narrow: Features/Product Area

This is the part of the framework that I felt is best understood through well-established methods, such as iterative usability testing, heuristic evaluations, and paper prototyping. As a team we were significantly invested in this work, which maps most closely to how the engineering teams work; when user researchers (URs) are well integrated with the teams they work with, it is easier for the research to have an impact. The URs were assigned to work with a particular themed area (e.g., Photos & Video, or Storage) which usually mapped to a particular product team (and sometimes

to more than one as the elements of a themed area were distributed across teams). However, URs were aware that while they worked in detail with the teams on their areas, they were also accountable for driving the broader holistic goals of the product through broader user tasks. This work then started to enter the larger experience work. Occasionally the UR might be in conflict with the team with which they worked most closely in order to drive for a change that would benefit a high level task –one of the challenging aspects of being a user researcher is being able to maintain a trusted relationship with a team while driving a user issue. Product teams who may have had a dedicated user research for their work previously had to adjust to the UR driving a broader charter.

3.2 Product Cycle-Broad: User Experiences

Cross Product Experiences

An operating system supports and enables many different tasks for many different audiences (home user, enterprise, IT specialist). We needed to prioritize which areas of Windows Vista we cared about most. This required setting up criteria to evaluate different tasks. The criteria we considered included task frequency, known task difficulty (based on our previous understanding of customer challenges), and newly enabled tasks. We were able to leverage previous research work from field, lab, and ethnographies to help us in identifying these tasks. We also had to define what a task was and how this differed from focusing on features. A task is defined in user language; to complete the task the user may use several features. For example, for a user to download 50 photos and send her favorite to her friend in email involves multiple features provided by several different teams (devices, photo download, file management, email setup, email send/receive, add attachment, receive an attachment). Although we were responsible for creating the list of tasks, we also needed buy-in from the individual teams that these were indeed tasks that they wanted to address with their features. We had to work with the development teams to create success criteria that were acceptable to development and research (e.g., 80% of participants should complete the task successfully) plus we had to incorporate some leeway in success criteria that would allow for emotional evaluation of experience, and customer site visit feedback. We established a list of over 160 tasks that we tracked during the development of Windows Vista. This list of tasks provided significant benefit to the development team throughout the development cycle. Frequently a group such as the performance test team would ask for the top scenarios that Windows Vista was targeting, and our list was defined with sufficient detail to be an actionable starting point for responding to such requests.
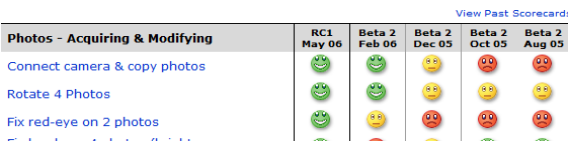

Fig. 2. Scorecard example

This list of tasks provided a critical starting point for driving accountability into engineering through the creation of a User Experience Scorecard (Figure 2). We iterated several times on creating a scoring system that teams would respond to and

that we felt reflected the experience we were on track to ship. This included allowing heuristic assessment of plans and specifications to be incorporated during the early stages of development, and evaluation as the milestones progressed in the development cycle. We used a three color rating system (red, yellow, green). Because we had detailed task success measures we used these as the primary assignment of the color rating. However if an additional data source provided insight that suggested a serious user problem we took that into account in the rating - mostly this would prevent a task that was completed successfully in a lab situation from being green if field insights suggested challenges. We made sure that anything less than green was accompanied by actionable bugs to be addressed.

At the end of most ship cycles are quality gates that must be met for a product to be released. Typically quality gates are test quality measurements, such as reliability, performance, and security. The rigorous procedures of our scorecarding method enabled us to adapt our scorecard to become part of the quality gate process. From our list of tasks we defined a subset that were considered 'ship-stoppers': critical tasks for which a failure to meet the criteria would lead to having the bugs and issues examined at a more detailed and senior review level to insure that things were fixed.

The User Experience scorecard and task list was used to drive many product changes, but identifying and eliminating task seams was a major benefit of the method.

There were challenges for the user researchers (URs) in driving the issues. Most of the URs worked closely with particular feature teams, but not with all the teams that might contribute to a particular experience, so to stay up to date on relevant feature plans required additional effort. This was one of our bets in terms of allocating resources--we decided the benefit for user experience of investing the time to track experiences across the product that mapped to user tasks would be greater than additional individual depth in particular niche areas. It was better to make the effort required to work across experiences than to leave to the users to work across siloed experiences after the product shipped. With this investment we uncovered many seams that might not have been addressed in the product had we not done this.

Emotional Connection
We were very much aware that an emotional experience is inextricably tied to satisfaction with a product, especially in the consumer market. At the time of working on Windows Vista we found methods that had been trialed to evaluate desirability, but the challenge was how to use these methods during the development phase and how to make the insights actionable. Benedek and Miner [1], members of the research team, created a desirability toolkit to help us evaluate these experiences. The tool is very simple but it provided insightful data that the URs and the designers could collaboratively turn into impactful action. After interacting with a product or prototype, a user is asked to select from a list of words, those words that they associate with the experience. The UR then discusses with the user why they selected particular words. The most important part of the assessment is the user's explanations. We used this tool during lab usability tests, benchmarks and in the field (with an automated version of the tool). Miner and Benedek were responsible for mining the themes across the studies and assessing how the particular lab study (or situation) may have influenced the selection of words and explanations. This was another example of how results pulled from the product-deep work were used to inform the broader experience of the product.

Productivity

Early in the development of Windows Vista we were asked what we could do to demonstrate improved productivity with the use of Windows Vista. As we unpacked what productivity meant in the context of Windows Vista use, we realized that it would be a difficult concept to measure for enterprise workers. After exploring the topic further with field representatives who work with our enterprise customers, we learned that they were less interested in demonstrating improved productivity than in knowing how we would assist people in climbing the learning curve as they deployed the new operating system. This insight led to a different approach to understanding how the enterprise learning experience should unfold. The feedback told us that we didn't need to build everything into the product to remove a seam—in this case, a companion experience could solve the problem. We developed an Enterprise Learning Framework (ELF) [2]. Working with enterprise users, we reviewed what should be included in the ELF. It included a time line (week before deployment, day of, day after, etc.), and what topics would be relevant to which users at that time. The topics then hooked up to the help system. In working through the topics we leveraged the insights URs had from working deep with feature teams to determine what would be useful to mention or areas in which users might have difficulties. We provided guidance to User Assistance about content to cover, something that may not otherwise have been included. To accompany the website a whitepaper was produced by Nowicki [3] which leveraged her learning from the research and creation of the framework. A triumph of the framework was in responding to enterprise customers' requests that it include both Office information and Windows Vista information, since they roll out desktops (Office and Windows), not individual pieces.

So again the investment of tackling productivity as a cross-product experience paid off rather than requesting for teams to think in an individual way about productivity.

3.3   Real Time – Narrow: Product

Customer Feedback Panel

We wanted to know a lot about user's behavior with Windows XP. To understand how a very large group of users were using Windows XP, we invested in creating the Windows Customer Feedback Panel [5]. Windows XP itself is not instrumented so we built a research platform that allowed us to upload data collecting tools to PCs over the Internet, which would then collect data from those machines on a regular basis. We recruited participants who were willing to allow us to gather instrumented data from their computers and associate it with other data sets related to them to enable us to ask follow-up questions. The advantage of leveraging a panel of known users is that we could profile characteristics of usage that applied to particular user groups. We could also survey this set of users on a needs basis. Because of the flexibility of the research platform we could adjust the data we were collecting–when new questions came up we could adjust the data collection tools to provide answers. As with all research, it was important to consider the sample bias. Although we were gathering data from more than 10K users, we knew they were slightly more technical users than average and were installing the data collection tools on home machines more often than work machines. This research platform allowed us to gather data we were not previously able to get, and was extremely good at gathering hardware,

configuration, file arrangements, and installed apps types of data. Understanding these dimensions of usage aided teams, such as the application compatibility team and the performance team, that we would not have been able to help using our regular user research.

Send-a-smile
We were now understanding what was happening on panelists' PCs. We also wanted to capture spontaneous emotional moments that arise during use. We created a tool called 'Send-a-smile' as part of the customer feedback toolset (Figure 3). A green smiley and red frown face were situated in the system tray (icons near the clock). When a user had a good moment she could click on the green face, or after a bad moment click on the red face. These would pop up a window with a text field for entering a comment and a screenshot of what was visible on the desktop. Comments and images were returned to us through the feedback tool. It was a very engaging tool to use, but as with all verbose feedback tools it was challenging to review all the feedback and turn into actionable suggestions or bugs to be entered into the bug data base [6, January 2007].

We used this Send-a-smile to gather feedback on the use of Windows XP and Vista, but it was product agnostic and was also leveraged by other teams at Microsoft.
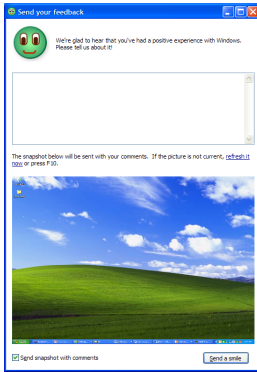

Fig. 3. Send-a-smile

Customer Adoption Panel
Windows has extensive beta programs, but most people who participate in them, especially in operating system betas, tend to be people who are relatively technically-minded. We knew it was important to include less-technical home users in the beta programs to get a rounded view of bugs and feedback on experiences. The research team owned the consumer adoption program for Windows Vista and had participants from throughout the US and overseas [6, January 2007]. The research program called "Living with Windows Vista" was an opportunity to provide all the usual bug feedback required from betas while also leveraging our research toolkit to evaluate additional dimensions of experience and use. This panel was relatively small (approximately 30 families) but we had deep engagement with them. The panel was invaluable because not only did it generate unique bugs but also we used our observations to change features, and several default settings based on problems encountered.

3.4   Real Time – Broad: Life Studies

Exploratory Ethnographies
The real time–broad cell covers an area of work that is basically understanding people without intervention, or with as little intervention as possible. Two anthropologists were on the research team. They were tasked with exploratory work. Their research areas were broad and not necessarily tied to technology; they could consider areas that might benefit from the introduction of technology. This set of work included research

in different geographical locations to understand emerging markets, the digital divide, the relationship between baby-boomers and their parents, dawn to dusk lives of small businesses, and other topics [6, 2005]. Each of the projects was uniquely designed, for example some were single day shadowing of participants, others were longitudinal over the course of a year.

The challenge with this type of work was to allow sufficient freedom in the research to truly enable the discovery about peoples' lives. The second challenge was how to share the insights from this work with the engineering and product marketing team. One strength of the work was in creating team member empathy for people and situations. This led to devising creative ways to communicate the findings, including photo-story narrations at the espresso coffee stand [4], posters in the buildings, and engagement through the creation of events related to the populations studied.

Not every observation leads to feature improvement, but it does provide the rich perspective of peoples' lives and their contexts that enable team members to realize how our products or potential products might fit into those lives.

Customer Engagements

Getting product teams involved in site visits is an activity that has been promoted for many years. We invested time in programs that weren't research but which were designed to drive empathy with customers. When team members are empathetic with their users they are more receptive to recommendations from user research. We created programs entitled, 'Know-a-knowledge-worker' or 'Get-to-know-an-IT-Pro'. Senior team members and executives were assigned a participant and provided with sufficient guidance to be able to conduct a site visit, and then spent time with a targeted customer to understand what they did in their day-to-day life at work, traveling to interact with them in their work context. The participants were not recruited based on their use of a particular technology, but based on what they did at work. We kept the requirements on reporting back from the visits to a minimum, as at the end of the day the benefit was to have more than 100 people on the team who had experienced what their customers would be doing. It was clear that the visits made an impression as reference to the visits would come up in discussions during development.

4 Summary

Although I have mapped the research that took place for Windows Vista to the User Research Framework, it is important to realize that the quadrants didn't act in isolation. It was the rich integrated insights gained from working in all these ways that provided us with a holistic view of our customers. The framework also provided a way of describing the size of investment in each quadrant. Teams get anxious when they can't clearly see a connection between research and specific feature impact. Even with this framework the majority of resources are invested in narrow product work – that is the most obvious opportunity to impact product, however we know from our experience that paying attention to the other quadrants has valuable impact on the experience in less obvious ways. Many of the programs and tools established during the Windows Vista development cycle have continued to be used and enhanced by the

Windows 7 team, by other user research teams at Microsoft, and even to assist in the marketing of Windows Vista.

References

1. Benedek, J., Miner, T.: Measuring Desirability: New methods for evaluating desirability in a usability lab setting. In: Usability professional association, 2002 Conference Proceedings (2002),
   `http://www.microsoft.com/usability/UEPostings/`
   `DesirabilityToolkit.doc`
2. Enterprise Learning Framework,
   `http://www.microsoft.com/technet/desktopdeployment/bdd/elf/`
   `welcome.aspx.com/presspass/features/2005/apr05/`
   `04-04Ethnographer.mspx`
3. Nowicki, J.: Learning Windows Vista in the Work Place. Microsoft white paper (2006),
   `http://download.microsoft.com/download/d/9/b/d9b9587e-427c-`
   `439f-b90c-69a1e643de4c/windows_vista_workforce.doc`
4. Steele, N., Lovejoy, T.: Engaging our audiences through Photostory. Visual Anthropology Review 20(1) (2004)
5. Windows Customer Feedback, `http://wfp.microsoft.com/Welcome.aspx`
6. Microsoft Presspass,
   Large-Scale Research Project Aims to Make Windows Vista Useful, Fun for All (January 2007), `http://www.microsoft.com/presspass/features/2007/jan07/`
   `01-29LivingWithVista.mspx;`
   Making technology conform to people's lives (April 2005); Interview with anthropologist Tracey Lovejoy, `http://www.microsoft.com/presspass/features/2005/`
   `apr05/04-04Ethnographer.mspx;`
   Developers Tap Real-Life Families to Find Out What Consumers Really Want from Windows Vista (January 2007),
   `http://www.microsoft.com/presspass/features/2007/jan07/`
   `01-29VistaDevelopers.mspx`