LIBERATING Usability Testing

Phil Carter | Auckland University of Technology | phil.carter@aut.ac.nz

To find out if something is useful, you use it and find out. There is movement between your use of the thing and your thinking about it. Remarkably, this everyday *inquiry within experience* has not been highlighted as the essence of usability testing. When it is, usability testing can be liberated from specialists and restrictive procedures [3]. It is humanized so it can be owned by anyone and used in a wide range of ways in response to the emergent needs of a software-development endeavor. In addition, when usability expands to include usefulness and what works well, it becomes much more generative. This article seeks to build awareness and precision in our practice of usability testing.

Abstractions

Many methods can be used to investigate a person's use of an artifact. If we say that the person's actual experience of using an artifact at the time of using it is the object of study, then most methods have abstractions or distances away from this object of study.

In a survey or questionnaire, the respondent is involved in a reflective process; that is, they are thinking about an event (their use of the artifact) that occurred in the past in a different situation. So there are abstractions in the dimensions of time and place. There is an abstraction in the primary focus of the study in that it is the questionnaire, not the artifact. And there is an abstraction in the cognitive framework in that the respondent is not relating to their own framework, but to the questionnaire designer's. Table 1 lists these abstractions and also outlines the abstractions in other methods which are typically used to evaluate usability.

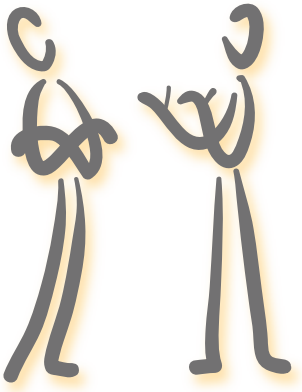
In a similar way to questionnaires, the typical application of interviewing does not deal directly with a person's actual thinking, feeling, and actions when using an artifact. Two other common methods used in usability evaluations—heuristic evaluations and cognitive walkthroughs—are also distanced from the core focus of the inquiry but are less likely to spill into reification because they do not pretend otherwise.

We might expect that the anthropological approach, which has the studying of behavior within context as a distinguishing characteristic, would highlight the mechanism of inquiry within experience. However, people using this approach have their research sensibilities attuned to the complex and fascinating socio-environmental contexts of use [5], not to the precision of a person's actual experience of using an artifact.

Usability testing, however, aims a direct inquiry into the experience of using an artifact while the artifact is being used. From the user's point of view, they simply relate to their own thinking and experience at the time of using the thing. As consumers with generations of shopping practice, we might also see this inquiry as an everyday type of thing. It is perhaps this marvellous sim-

	Questionnaire	Interviewing	Heuristic evaluation	Cognitive walkthrough	Usability testing
Time	No	No	Yes	Yes	Yes
Place	No	No	No	No	No
Artifact (as primary focus)	No	No	No	Yes	Yes
User	Yes	Yes	No	No	Yes
User's cognitive framework	No	Yes	No	No	Yes

Table 1: Inquiring into the actual use of an artifact: Are these methods typically done to the conditions of use?



plicity that has obscured the power of the mechanism of directness, the heart of usability testing.

Expression Within Experience

There are differences between a person's own testing of a recently purchased product and a usability test. In a usability test there is an expression of the experience of using an artifact, a description of that, and then a communication to the developers of the artifact. However, these activities can also be built on everyday competencies, and so effective practice can be achieved by nonspecialists.

Talk aloud or *think aloud* has been usability's main technique for a user to articulate their experience—particularly their thinking—to the others involved in the testing. Nielsen et al [7] provide an informative genealogy of this technique from psychology. However, this thing we are talking about, self-reflective consciousness, is at the core of what it is to be a human being. Every functional one of us is competent. The ease with which most people can immediately *talk aloud* on their first usability test is evidence of this shared competence.

Our experience over many years in a usability lab is that *talk aloud* is very effective in exposing many of the factors at play within a user. However, at times further clarification is needed. For example, the test user may be frowning, but there is no verbal expression relating to that. Is it because of the color contrast on the screen, incomprehension of a button icon, or indecision about navigation? Kuniavsky suggests the moderator could say, "You frowned when that dialog box came up. Is there anything about it that caused you to do that?" [6, p. 119]. What is the effect of a question such as this? It introduces a cognitive process that most users will seek to satisfy. If they are primarily engaged in feeling or acting, they will switch attention. Even if they are thinking, it will most likely require a shift to finding reasons, a process that can dominate attention. Table 2 outlines the possible effects of some other types of interventions that can be given in response to a test user's frowning but saying nothing.

gives a few for the moderator in usability testing: Don't force opinions; restate answers; follow up with examples; use artifacts to keep people focused on the present and to trigger ideas; be aware of your own expectations; never say the participant is wrong; listen carefully to the questions that are asked of you; keep questions simple in language and in intent; probe

Intervention	User response
Why are you frowning?	Requests a cognitive process that may interfere with the user's experience if it is affective or kinaesthetic. Searching for reasons and rationale is a constructive cognitive activity.
What are you thinking?	The user's experience may still be affective/kinaesthetic. Requesting a cognitive process may interfere with this.
What you are experiencing?	Use of the word "experiencing" removes the threat of not being in line with affective or cognitive focus of the test user; however, the question still requires a cognitive process that may be premature.
What is it? What's happening?	More informal and more neutral than the previous two statements, and so likely to give the user more room to continue to use the system and give feedback when ready.
You're frowning...	Offers information from an external source that may assist the user to increase their self-awareness. However, this may also be presumptuous, and the user may argue against it.
You're wrinkling your brow.	More observational and less processed than the previous statement. However, the user may feel exposed.
Something's up, eh?	If said with emotive tone and force that's similar to what the user is displaying, this statement may assist the user in staying with their experience and feeling invited to express it in their own time.
Nothing...	Something may eventuate. If not, and the moment is lost and the user becomes involved with another aspect of the system, then a pause may be used for reflection, or the user may be invited to back up.

Table 2. Possible effects on user by different interventions

While the preceding analysis can wake us up to the precision of language, it is not sufficient as a guideline for practice. There are many other factors that have impact, such as the moderator's tone, attitude, comfort level, and friendliness. Likewise, there is the personality, attitude and capabilities of the user. And so there is also the relationship between these two people. Guidelines can quickly proliferate. Kuniavsky [6]

expectations; ask why a lot; investigate mistakes; probe nonverbal cues; respect the evaluator's ideas; and focus on personal experience. While these are all good guidelines that may be integrated into a moderator's functioning with prolonged and repeated practice, they are too numerous for nonspecialists to own and adopt them.

Being Alongside

Table 2 indicates that the intervention "Something's up, eh?" results in a desirable outcome. Such an intervention is not generated from a focus on what to say or adherence to a complex list of guidelines but emerges from a type of relationship between the user and the moderator where the moderator is alongside the user. *Being alongside* is one way to describe the attitude a moderator can adopt that will evoke effective interventions such as this one. By being alongside, the moderator's own experience will inform her understanding of the user's experience. They need not, therefore, ask so many questions, but in combination with simple observations, they will be able to offer a comment that is gift-like—it doesn't demand anything in response.

For example, the user may be smiling, clearly enjoying themselves, and the moderator alongside the user, also having a good time, says, "This is fun, eh?" It's an easy opening for the user to keep connected with their experience and to elaborate on it in their own time. There is also a lot of space for the moderator to keep connecting with the emergent functioning of the user. The subsequent interaction can have the quality of a conversation between friends.

Such friendly conversations are not dominated by safety concerns; rather, they evoke an inquiry that is informed by all the dimensions human beings generate. The intellect can be informed by the affect and the imagination, as well as what the eyes see and the brain thinks. Friendliness is not something that requires an emotional precursor to get going or emotional fusion to be sustained. Friendliness is uplifting for the moderator and creates a sense of ease and enjoyment that stimulates intrigue and curiosity. Users benefit enormously from naive inquiry because naivete destroys distrust and getting it right. There is less anxiety to seek understanding before experiencing and less pressure to fix anything up; all habits are grounded in culture and are accentuated by our technical understanding. It appears that the beneficial qualities of inquisitiveness and openness tend not to emerge when someone is actively pursuing them, but unfold from the generation of a friendly, cooperative working relationship.

Many guidelines on interviewing strongly emphasize that the moderator be neutral. Seeking neutrality will interfere with the moderator's ability to be alongside and be informed about the user. Worse, neutrality is impossible to sustain for any length of time, and its pretence will destroy genuineness and threaten the user. If, in contrast, a moderator can inculcate in themselves a flexible movement between being open and naive and having responses, then the user is more likely to feel secure, knowing they are dealing with a real human being. The user will also be more likely to bring forward tentatively ask questions because they anticipate they will be treated in a genuine way.

The relationship establishes a strong motivation for a truly ethical approach that can have an ongoing attention to the well-being of all. There is no need to give the advice that a user often hears at the beginning: *It is not about you*. It is about them. It *is* about their encounter with the artifact. When they experience feeling dumb or vulnerable, do we then say it is not about them? This invalidation is not useful. Instead, a cooperative interest in the work can be established. If right from the beginning, all participants know that the common purpose is to uncover and explore the usability of this particular artifact; then the attraction of discovery becomes stronger, and the sense of companionship in the journey develops. Under such conditions, the emotional need for safety ceases to dominate and true safety is established.

Much can be done in the pretesting warm-up phase to establish a cooperative working relationship.

Typically, the warm-up phase is seen as being for the benefit of the test user, but it is also a great opportunity for the moderator to generate respect and gratitude for the test user. The test user has often come along out of a sense of generosity and wanting to contribute to the work of making a better artifact. When one person experiences gratitude, it is enormously generative within a relationship. In addition, emphasize and clarify the value of the artifact during the warm-up phase so that the whole enterprise can take on greater respect and dignity.

Descriptions for the Developers

With the common everyday habit of *inquiry within experience* firmly established as the functioning heart of usability testing, there is also lots of space for fresh ideas in response to the different challenges of software-development projects.

One area is how to create precise and pertinent descriptions of usability that are immediately sensible to developers and do not consume precious time in post-testing analysis. With a cooperative working relationship established, the logger or scribe can become more involved and feed tentative descriptions back to the user for clarification and further refinement. The describing task need not be a logger grinding out logs on their own. The descriptions get their own airtime and "usability test." Users can be engaged in a co-inquiry that can, at any time, return to and be grounded in the actual use of the artifact. The moderator and scribe can pay particular attention to creating descriptions that are precise and detailed enough for the readers (system analysts, designers, developers, programmers, project managers) to understand. This cooperative in-vivo process of creating clear descriptions can assist thinking about usability; ambiguity is uncovered and so can be dealt with during the testing, removing post-testing guesswork.

In terms of physical arrangement, participants can be flexibly arranged according to each situation to facilitate this cooperative work. Often a triangle setup is useful because everyone can see everyone else. In this setup, the moderator sits beside but slightly behind the user, and the scribe sits across on an angle. The scribe can have a second monitor to make following the user's onscreen actions easy.

A clear and explicit group commitment to the importance of identifying the usability pluses and minuses encourages most test users to be authentic and to value and express both their joys and difficulties. Under such conditions, system developers can be productively introduced into the room. This can be very curative if developers have experienced—or even perceive—usability as something being done to them. How much more productive to have the intelligence and the design that options developers are weighing up, and the deep and extended thinking they are throwing into the work to be part of the mix [4].

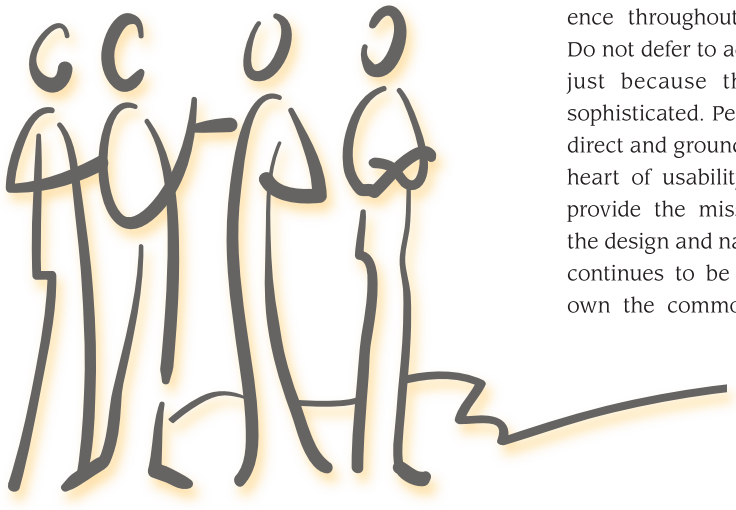
Usability as Usefulness

Great advantage occurs when the actual human value of an artifact becomes the core criteria on which usability inquiries are grounded. For example, if learning outcomes become integrated into the usability assessment criteria for an educational system, then clear links can be made between the elements of the system that support or hinder learning. Likewise, if the different interface features and interactions of an e-mail application are related to people's ability to communicate and work, then deep understanding results about the interaction between the human and the tool. "Usability is meaningless unless it is usability for something worthwhile." [11, p. 58]. Without the focus on worthwhileness, what will keep usability from being compromised by compliance mentalities and converted into checklists? The building of scenarios and personas and much of the good work grouped under user-centered design also help us to put the artifact into its proper living context. Once again, the heart of these approaches is simple and humane: Get to know someone and appreciate the different aspects of their life.

Usability findings gain a much greater generative capacity when the focus is what works as well as what doesn't. Extensions and enhancements to what works well can be precisely identified. Software

designers and developers get the praise and appreciation they deserve and get the encouragement and motivation to continue. One should be alert to the limitations that can occur when a mindset or methodology begins with problematic findings. However, escape can occur through expansion; we ask, what is the human value of the artifact we are involved in designing and developing? When we see meaning and value in our work, we invest more imagination and careful thinking. We will not be so easily governed by compliance or committee, but we will be freed up no matter what the external structures.

Freed up, we will not preempt our judgements or get narrowed in our thinking. History gives us many examples that might help us become humble in the knowledge that we don't know what we don't know. Test showings of pilot episodes of *All in the Family* and the *Mary Tyler Moore Show* scored low in terms of likability, and both shows were nearly dumped before they got going [2]. New was shocking, and not altogether welcome. In another fascinating example, Brian Eno illustrates that the relationship with an artifact and its restrictions can be highly desired and even mandatory to creativity [1]; we can put aside even the treasured maxim of *designing tools that are transparent to the task* so that we can attend to the actuality.



Wrap-Up

Sit alongside someone as they use something, and find out how it's going: That's usability testing. It's that simple. When we appreciate the everyday, common-sense essence of usability testing, we free ourselves from the specific mandates of different methods [9]. There is room for us to attend to what is actually happening. We avoid the narrowness of thinking we are dealing with a process of extraction—we are in a human process. We bring the human qualities of inquisitiveness and friendliness because they are most effective. Human beings with life experience become the experts [8]. No group need get precious or struggle to own usability [10]. It can be taken up and integrated into any framework.

Do not underestimate the engine of usability testing just because it is simple. Own it, and overly controlling and fear-motivated forces will not so easily get purchase and colonize usability for their restrictive purposes. See the simple nobility of this inquiry. Do one thing at a time and notice that even one little piece of cooperative work ripples influence throughout an organization. Do not defer to academic structures just because they appear more sophisticated. Perhaps it will be the direct and grounded inquiries at the heart of usability testing that will provide the missing link between the design and natural sciences that continues to be elusive. When we own the common human core of the usability inquiry and see it as based on usefulness, then usability may offer even more than has been imagined [11]. We will surely need everyone's involvement and creativity if we are to move forward in these times.

REFERENCES **1.** Eno, B. (1999). Revenge of the Intuitive. *Wired, 7*(1). **2.** Gladwell, M. (2005). *Blink: The Power of Thinking without Thinking.* Victoria, Australia: Allen Lane. **3.** Holzinger, A. (2005). Usability engineering methods for software developers. *Communications of the ACM, 48* (1), 71-74. **4.** Hornbaek, K. & Frokjaer, E. (2005). Comparing usability problems and redesign proposals as input to practical systems development. In *Proceedings of CHI 2005.* Portland, Oregan, USA (April 2-7), 391-400. **5.** Kantner, L., Sova, D. H., & Rosenbaum, S. (2003). Alternative methods for field usability research. In *Proceedings of SIGDOC'03.* San Francisco, California, USA (12-15 October), 68-72. **6.** Kuniasky, M. (2003). *Observing the User Experience: A Practitioner's Guide to User Research.* San Francisco: Morgan Kaufmann. **7.** Nielsen, J., Clemmensen, T., & Yssing, C. (2002). Getting access to what goes on in people's heads? - Reflections on the think-aloud technique. In *Proceedings of NordiCHI 10/02.* Arthu, Denmark (October), 101-110. **8.** Norman, D. A. (2005). Whose profession is this? Everybody's, nobody's. *Interactions*, May/June, 51. **9.** Seffah, A. & Metzker, E. (2004). The obstacles and myths of usability and software engineering. *Communications of the ACM, 47* (12), 71-76. **10.** Siegal, D. & Dray, S. (2005). Avoiding the next schism: ethnography and usability. *Interactions*, March/April, 58-61. **11.** Spool, J. & Schaffer, E. M. (2005). In *Proceedings of CHI 2005.* Portland, Oregon, USA (April 2-7), 1174-1175.

ABOUT THE AUTHOR

Phil Carter began working in computing by operating a HP 3000 in the early 1980s. He then lived in Taiwan for seven years before returning to New Zealand to study computer and cognitive science. His PhD focused on Expert Systems and Psychodrama. He is also a certified Psychodramatist. He has setup and been director of a usability research centre and is currently teaching usability to computing and business students at the Auckland University of Technology.