

User Studies and Usability Evaluations: From Research to Products

I. Scott MacKenzie

York University, Toronto, Canada

ABSTRACT

Six features of user studies are presented and contrasted with the same features in another assessment method, usability evaluation. The connection between these assessment methods and the disciplines of research, engineering, and design is analysed. The three disciplines are presented in a timeline chart showing their inter-relationship with the final goal the creation of computing products. Background discussions explore three definitions of research as well as three methodologies for conducting research: experimental, observational, and correlational. It is demonstrated that a user study is an example of experimental research and that a usability evaluation is an example of observational research. In terms of the timeline, a user study is performed early (after research but before engineering and design), whereas a usability evaluation is performed late (after engineering and design but before product release).

Keywords: Research, engineering, design, human-computer interaction, user study, usability evaluation, product development

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces – evaluation/methodology, input devices and strategies

1 BACKGROUND

In human-computer interaction (HCI) and other fields, the path from ideas to products is lengthy. Along this path, elements of research, design, and engineering work together to generate and refine ideas and to combine them with existing art into new products – products that work well and appeal to consumers (users). By “idea”, we mean new interactions or any novel twist to a user interface (UI) that will advance the state of the art by improving interactions or creating an appealing user experience.

Testing is critical. A lot of testing pertains to system components, without any relevance to HCI. However, many aspects of the system engage the user’s sensory, motor, or cognitive processes. These form the UI and fall within the purview of HCI. Testing in such cases is also critical and employs assessment methods adopted and refined in HCI.

This paper concerns research, design, and engineering and their relationship to assessment methods employed in HCI. The assessment methods of interest here are the *user study* and *usability evaluation*.¹ For the most part, we’ll assume the reader has a general understanding of what constitutes a user study or a usability evaluation. So, little is said about the mechanics of how

¹ In the author’s experience, these terms are the most common and most accepted for the assessment methods they embody. However, other terms are sometimes heard. For example, in *Designing the User Interface*, Shneiderman and Plaisant use the terms “controlled psychology-oriented experiment” for user study and “usability testing” for usability evaluation [24, chapt. 4].

to design and conduct a user study or a usability evaluation. Readers interested in such are directed to other sources, such as Shneiderman and Plaisant’s book [28] for usability evaluation or the author’s book [20] for user studies.² We are mostly interested in how a user study fits in with research, engineering, and design, and how a user study differs from a usability evaluation. These two assessment methods are worlds apart, yet their differences are often blurred and poorly understood. One goal herein is to rectify this – to identify the differences between a user study and a usability evaluation, and to show how and where each fits into the wider milieu of research, engineering, design, and products.

As it turns out, achieving this goal is trickier than it seems. And so, we delve into the nature of research itself (what it is) and research methods (how to do research). We’ll also examine how – along a lengthy timeline – research results feed into engineering and design and how the three disciplines work together in the creation of products.

Finally, note that the analyses herein echo and extend those in the author’s recent book, *Human-Computer Interaction: An Empirical Research Perspective* [20]. The relevant section is Chapter 4, “Scientific Foundations”. Let’s begin.

2 WHAT IS RESEARCH?

Research means different things to different people. “Being a researcher” or “conducting research” carries a certain elevated status in universities and corporations. Consequently, the term “research” is bantered around in a myriad of situations. Often, the word is used simply to add weight to an assertion (“Our research shows that ...”). But what is research? Surely, it is more than just a word to bolster an opinion.

Research has at least three definitions.³ First, conducting research can be an exercise as simple as *careful or diligent search*. So, carefully searching one’s garden to find weeds meets one standard for research. Or perhaps one undertakes a search on a computer to locate files modified on a certain date. That’s research. It’s not the stuff of MSc or PhD theses, but it meets one definition of research.

A second definition of research is *collecting information about a particular subject*. So, surveying voters to collect information on political opinions is conducting research. In HCI we might observe people using an interface and collect information about their interactions, such as the number of times they consulted the manual, clicked a particular button, retried an operation, or uttered an expletive. That’s research.

This third definition is more elaborate: Research is *investigation or experimentation aimed at the discovery and interpretation of facts and the revision of accepted theories or laws in light of new facts*. Most people who self-identify as

² The relevant sections from the author’s book are Chapter 5 (“Designing HCI Experiments”) and Chapter 6 (“Hypothesis Testing”).

³ <http://www.merriam-webster.com/>

researchers are likely to align with this definition. As HCI researchers, we are charged to go beyond diligent search or the collecting of information. Our work involves discovery, interpretation, revision, and perhaps experimentation – in other words, pushing the frontiers of knowledge. That's research!

It is worth adding that the research discussed here is *empirical research*. This is in contrast to theoretical research. By adding the prefix “empirical”, we invite inquiry that originates in observation and experience – the human experience.

This third definition of research is so rich it necessitates different ways to go about doing research. One goal for this paper is to establish how user studies work with HCI research. But where and how do user studies fit into the vast milieu of HCI research? As we seek to answer this question, we'll also distinguish a user study from a usability evaluation, which is an entirely different type of assessment method used in HCI.

Let's approach the question above by first delineating the different approaches or methods in conducting research.

3 RESEARCH METHODS

Research methods in the natural or social sciences (e.g., HCI) fall into three categories: observational, experimental, and correlational [27, pp. 76-83].

3.1 Observational Research

The *observational method* encompasses a collection of common techniques widely used in HCI research. These include field investigations, expert reviews, contextual inquiries, interviews, case studies, focus groups, think aloud protocols, storytelling, walkthroughs, cultural probes, and so on. The approach tends to be qualitative rather than quantitative, focusing on the *why* or *how* of interaction, as opposed to the *what*, *where*, or *when*. The goal is to understand human thought, feeling, attitude, emotion, passion, sensation, reflection, expression, sentiment, opinion, mood, outlook, manner, style, approach, strategy, and so on. These human qualities are highly relevant but difficult to directly measure.

With the observational method, behaviors are studied by observing and studying phenomena in a natural setting, as opposed to crafting constrained behaviors in an artificial laboratory. The phenomena observed and studied are real-world interactions between people and computers. Real-world phenomena are high in relevance and practical value, but lack the precision available in controlled laboratory settings. As a result, observational methods tend to achieve *relevance* while sacrificing *precision*.

A common example of HCI research following the observational method is a *usability evaluation*. A usability evaluation seeks to assess the usability of a system's UI to identify specific problems [7, p. 319]. Although there are more than 100 different usability evaluation methods [15], the general approach is to engage users in doing tasks with a particular UI. The goal is to find problems that compromise usability. Some usability evaluation methods involve expert users who observe and assess the interface.

A key goal in this paper is to distinguish a usability evaluation from a user study. Let's continue.

3.2 Experimental Research

The *experimental method* (also called the *scientific method*) acquires knowledge through controlled experiments conducted in laboratory settings. In the relevance-precision continuum, it is clear where controlled experiments lie. Since the environment is

artificial, relevance to real-world phenomena is diminished. However, the control inherent in the methodology brings precision, since extraneous factors – the diversity and chaos of the real world – are reduced or eliminated.

A controlled experiment requires at least two variables: a *manipulated variable* and a *response variable*.⁴ In HCI, the manipulated variable is typically a property of an interface or interaction technique that is presented to participants in different configurations. Manipulating the variable simply refers to systematically exposing participants to different configurations of the interface or interaction technique. A response variable is any property of human behavior that is observable, quantifiable, and therefore measurable. The most common response variable is time – the time to complete an interaction task. Other possibilities include the reciprocal of time (speed), measures of accuracy, counts of relevant events, etc.

A *user study* is an experiment with human participants. Full stop! Here, we see one of the strengths of HCI as an interdisciplinary field. The methodology for an HCI experiment with human participants – a user study – is plucked wholesale from one of HCI's constituent fields: experimental psychology. Experimental psychology is an established field with a long history of research involving humans. In a sense, HCI is the beneficiary of this more mature field. To be clear, “methodology” here refers to all the choices one makes in undertaking the user study. These pertain to the research questions, the people (participants), ethics, the apparatus (hardware and software), the interface, the tasks, task repetitions and sequencing, the procedure for briefing and preparing participants, the variables, the data collected and analyzed, and so on.

3.3 Correlational Research

The *correlational method* involves looking for relationships between variables. For example, a researcher might be interested in knowing if users' privacy settings in social networking applications are related to their personality, IQ, level of education, employment status, age, gender, income, and so on. Data are collected on each item (privacy settings, personality, etc.) and then relationships are examined. For example, it might be apparent in the data that users with certain personality traits tend to use more stringent privacy settings than users with other personality traits.

The correlational method is characterized by quantification since the magnitude of variables must be ascertained (e.g., age, income, number or level of privacy settings). For nominal-scale variables, categories are established (e.g., personality type, gender). The data may be collected through a variety of methods, such as observation, interviews, on-line surveys, questionnaires, or direct measurement. Correlational methods often accompany experimental or observational methods, for example, if a questionnaire is included in the procedure.

Correlational methods provide a balance between relevance and precision. Since the data were not collected through a controlled experiment, precision is sacrificed. However, data collected using informal techniques, such as interviews, bring relevance – a connection to real-life experiences. The conclusions available through correlational research are circumstantial, not causal. But, that's another story.

* * *

⁴ Synonymous terms are *independent variable* and *dependent variable*, respectively.

Figure 1 combines the points above by positioning each research method in a chart showing the emphasis on relevance vs. precision. The most important point here is that any attempt to describe one methodology as better or worse than another is misguided. Each methodology brings something the others cannot.

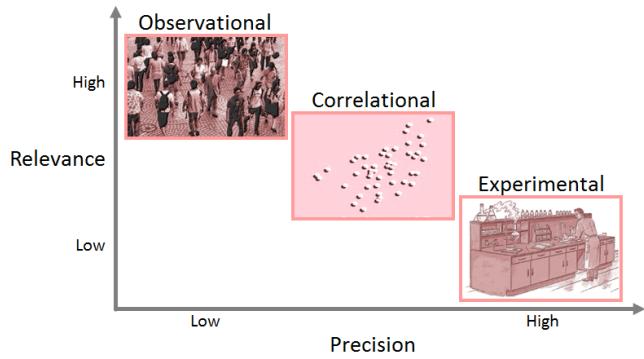


Figure 1. Research methods and their emphasis on relevance and precision.

Terms frequently used in assessing research methods are *internal validity* and *external validity*. High internal validity means the outcomes observed really exist (rather than being artifacts of uncontrolled factors). High external validity means the outcomes are broadly applicable to other people and other situations.⁵ Despite the best attempts of researchers to design their work to achieve high internal validity and high external validity, there is an unavoidable tension between these two properties. See Figure 2. In other words, efforts to strengthen one tend to compromise the other.

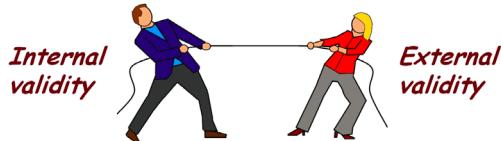


Figure 2. The unavoidable tension between internal validity and external validity. Sketch courtesy of Bartosz Bajer.

In terms of the research methods just described, observational methods tend to emphasize relevance and achieve high external validity while experimental methods tend to emphasize precision and achieve high internal validity. Consequently, we also note that a usability evaluation tends to emphasize external validity while a user study tends to emphasize internal validity.

4 A USER STUDY IS NOT A USABILITY EVALUATION

The discussion above places user study in the realm of experimental research and usability evaluation in the realm of observational research. Let's continue to distinguish these two assessment methods.

A user study requires at least one manipulated variable (aka independent variable) which in turn must have at least two levels or configurations. Users are exposed to the levels of the manipulated variable while their behaviour or performance is observed and measured. Thus, comparison is germane to a user study. In fact, a user study frequently has multiple manipulated variables, each with a different number of levels or

⁵ A related term is *ecological validity*. Ecological validity refers directly to the methodology (using materials, tasks, and situations typical of the real world), whereas external validity refers to the outcome.

configurations. This leads to a common description of a user study as, for example, a “ $3 \times 2 \times 4$ design”.

The idea of a manipulated variable is foreign to a usability evaluation. By and large, a usability evaluation is a one-of assessment – an evaluation of a particular user interface (UI). The UI might be a commercial product, but frequently it is a prototype. The goal is to find faults or weaknesses in the UI, often with reference to accepted UI design principles. A usability evaluation qualifies as research (“collecting information about a particular subject”), but the methodology is observational, not experimental.

Clearly, usability evaluation is hugely important for companies bringing products to market. But, a usability evaluation is not a user study. In a usability evaluation, there is no manipulated variable. Users do a variety of tasks while problems are noted; but the different tasks are not levels of a manipulated variable. The tasks are chosen simply to encompass the range of activities supported and for which a usability assessment is sought.

Another distinction between a user study and a usability evaluation is the level of detail in the inquiry. A user study is low-level. The goal is to assess and compare two or more interaction details to determine which works better. The details studied and compared are the minutiae of interaction – for example, whether a particular task or operation is performed better with tactile feedback vs. popup animation. In the context of a user study, “better” means preferred scores on one or more response variables.

The tasks employed in a user study are constrained. There are two goals in designing a task for a user study: (i) to avoid extraneous influences that might inject noise in the response variable, and (ii) to provide the capability to distinguish the levels of the manipulated variable.

A usability evaluation is high-level. The goal is to assess a UI or application as a whole, and to uncover potential problems that an end-user might confront. This is a higher level of inquiry, since the evaluation involves complete tasks done on a complete (perhaps prototype) UI or application.

The tasks employed in a usability evaluation are natural and unconstrained. The goal is to fully reflect the way an end-user interacts with and experiences the system.

* * *

We have identified several features that distinguish a user study from a usability evaluations. The research methods are different. Consequently so too are precision and relevance in the findings. A user study, being experimental, is more likely to attain high internal validity, whereas a usability evaluation, being observational, is more likely to attain high external validity. A user study includes manipulated variables (at least one) to expose users to different configurations of an interface, whereas a usability evaluation is a one-of assessment of a user interface seeking to uncover faults or weaknesses.

Perhaps the most important distinction between a user study and a usability evaluation lies in the *when*, as opposed to the *what*. To understand this, we must step back and consider the big picture: the path from research to products. Enter engineering and design.

5 RESEARCH VS. ENGINEERING VS. DESIGN

There are many ways to distinguish research from engineering and design. Researchers occasionally work alongside engineers and designers, but the skills and contributions each brings are different. Engineers and designers are in the business of building things. Their goal is to create products that strive to bring

together the best in form (design emphasis) and function (engineering emphasis). Designers strive to create products with high appeal to the end user. Engineers, on the other hand, are more likely to focus on a checklist of functional requirements.

The form and function roles for engineering and design are shown in Figure 3. The chart is deliberately vague, as the disciplines overlap and have boundaries that are subject to interpretation. Obviously, the goal is to create products that are high in form *and* function (top-right in chart). These are products that work well and that users enjoy.

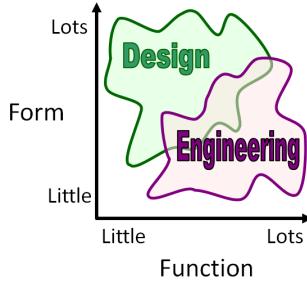


Figure 3. Form vs. function. Design emphasises form, engineering emphasises function.

One can imagine that there is certain give and take between form and function. Finding the right balance is key. However, sometimes the balance tips one way or the other. When this occurs, the result is a product or a feature that achieves one (form or function) at the expense of the other. An example is shown in Figure 4 (left). The image shows part of a notebook computer, manufactured by a well-known computer company. By most accounts, it is a typical notebook computer. The image shows part of the keyboard and the built-in pointing device, a touchpad. The touchpad design (or is it engineering?) is interesting. It seamlessly merges with the system casing. The look is elegant – smooth, shiny, metallic. But something is wrong. Because the mounting is seamless and smooth, tactile feedback at the sides of the touchpad is missing. While positioning a cursor, the user has no sense of when his or her finger reaches the edge of the touchpad, except by observing that the cursor ceases to move. This is an example of form trumping function. One user's solution is also shown in Figure 4 (right). Duct tape is added on each side of the touchpad to provide the all-important tactile feedback.



Figure 4. Form trumping function. See text for discussion.

Figure 4 reveals a failure in either design or engineering. In a sense there are no failures, as such, in research. In fact, it is perfectly common in research that a favorable or desired outcome fails to materialize. That's the nature of research – a point we'll develop further shortly. The tag “failure” is used only in so far as a poor or faulty UI feature was engineered or designed in, despite the checks and balances on the way to the final product. Really, this is a failure in assessment. One might wonder how competent or thorough the assessment was for the touchpad implementation in Figure 4, or whether any assessment with users was performed

at all. Perhaps there was a tight timeline on getting the product to market.

6 THE PATH TO PRODUCTS

Designers and engineers work in the world of products. The focus is on designing and bringing to market complete systems or products.

Research is different. Research tends to be narrowly focused. Small ideas are conceived of, prototyped, tested, then advanced or discarded. New ideas build on previous ideas and, sooner or later, good ideas are refined into the building blocks – the materials and processes – that find their way into products. But research questions are generally small in scope.

Engineers and designers also work with prototypes, but the prototype is used at a relatively late stage, as part of product development. A researcher's prototype is an early mock-up of an idea, and will not directly appear in a product.

The march forward for research is at a slower pace than engineering and design. Figure 5 shows a timeline sketch for research, engineering, and design. Products are the final goal. The raw materials for designers and engineers are materials and processes that already exist (dashed line) or emerge through research.

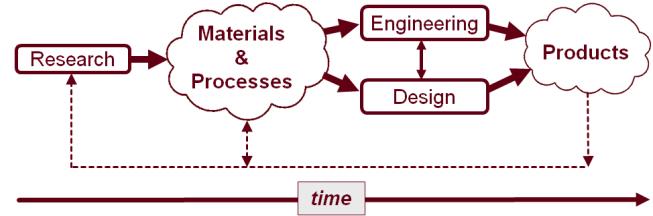


Figure 5. Timeline for research, engineering, and design.

Let's develop a case for the timeline in Figure 5 through four examples.

6.1 Computer Mouse

First, consider the computer mouse. The mouse is a hugely successful product that, in many ways, defines computing after 1984, when the Apple *Macintosh* was introduced. The *Mac* popularized the graphical user interface (GUI) and point-select interaction using a computer mouse. The mouse was first introduced a few years earlier, in 1981, when Xerox introduced the *8100 Star Information System*, or *Star* for short [16]. See Figure 6 (right). But, the mouse was developed as a research prototype in the mid-1960s by Engelbart [8]. See Figure 6 (left).

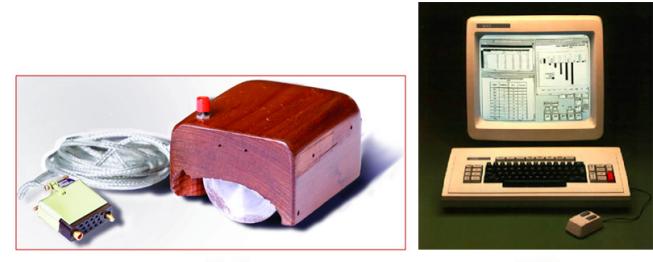


Figure 6. The computer mouse emerged from research in the 1960s, but took about 15 years to appear commercially in the Xerox *Star*.

Remarkably, it took about 15 years for the mouse to be engineered and designed into a commercial product. Among the

engineering required was replacing wheels attached to shafts of potentiometers with a rolling ball assembly [25].

The example of the mouse supports the timeline in Figure 5: The research occurred in the 1960s and the engineering and design occurred in the years leading up to 1981 when the computer mouse, as a commercial product, first appeared.

6.2 Apple iPhone

There are many scenarios with a timeline similar to the computer mouse. Consider the Apple *iPhone*, introduced in June 2007. As Selker notes, “with the *iPhone*, Apple successfully brought together decades of research” [26]. Many of the raw materials in this successful product came by way of low-level research, undertaken well before Apple’s engineers and designers set forth on their successful journey.

Among the *iPhone*’s interaction novelties are finger gestures, such as flick and two-finger pinch, and re-orienting the display when the device is rotated. To most *iPhone* users in 2007, these were new and exciting interactions. But, were these interactions really new? As interactions embedded in a commercial product, perhaps. But Apple’s engineers and designers were likely guided or inspired by research that came before them. For example, multi-touch gestures date at least to the 1980s [6, 12]. One example of two-finger input was described even earlier, in 1978, by Herot [13]. And the flick gesture? Flick gestures were used in 1963 by Sutherland in his celebrated *Sketchpad* graphics editing system [29]. What about changing the aspect ratio of the display when the device is rotated? New? Not really. Tilt, as an interaction technique for user interfaces dates to the 1990s [14, 24]. An example is *tilt me*, presented in 1998 by Harrison et al. [11].

Figure 7 is a re-vamped version of Figure 5 showing the interactions just described. They emerged from research. Dates are indicated for each as is the 2007 launch of the Apple *iPhone*. We have taken the liberty of stretching the timeline for research (research takes time!) and replacing the dashed lines with a bubble representing the materials and processes in existing products.



Figure 7. Flick, two-finger (pinch), and tilt interactions emerged from research that occurred well before the introduction of the *iPhone* in 2007.

6.3 Press-to-select Touchpad

An interaction that emerged commercially in 2008 was *press-to-select* for touchpads. Selection on earlier touchpad devices required the user to either press a separate button or perform a quick down-up tap on the touchpad surface. The Apple *Macbook Pro* and the *Blackberry Storm*, both introduced in 2008, circumvented this through a novel interaction. After moving the cursor, the user could select simply by pressing harder on the touchpad surface. All-important tactile feedback was provided by designing (or was it engineering?) the touchpad to operate like a giant button that the user presses down. See Figure 8 (right two images).



Figure 8. Touchpad press-to-select appeared in research in 1997 and then in the Apple *Macbook Pro* and *Blackberry Storm* in 2008.

Was press-to-select an idea that originated with the engineers or designers at Apple and Blackberry? Apparently not. About 10 years earlier, the same interaction appeared in research presented at the ACM SIGCHI conference [21, 22]. See Figure 8 (left). The *tactile touchpad* allowed the user to select simply by pressing harder on the touchpad surface.⁶ A relay below the touchpad provided the all-important tactile feedback when the sensed finger pressure exceeded a threshold. When the *Macbook Pro* and *Storm* were introduced, the link to this research was recognized by at least one reviewer [2, 3].

6.4 3dof Mouse

For our final example, we return to the computer mouse. The mouse is an indirect pointing device; it operates on a 2D surface to control an object on a 2D display. In 2D, there are three degrees of freedom (3dof): x and y translation and z-axis rotation. For example, an object sitting on a table can be moved left-right, forward-back, and it can be rotated. If the object is a mouse, it can be moved in three degrees of freedom, but only translation is sensed and reported to the computer. The 3rd dof (z-axis rotation) is missing.

This mouse deficiency was addressed in 1997 by research that added z-axis rotation to the sensing capability of a mouse [23]. See Figure 9 (left). Of course, the 3rd dof is not needed for common point-select tasks. For example, in moving a cursor to select objects, only the x and y translation data are needed. But, the added dof is potentially useful for other mouse interactions, such as graphics editing. With a 3dof mouse, the common rotate tool is not needed. An object can be maneuvered simultaneously in three degrees of freedom.⁷ Other possible applications are gaming and virtual reality. Adding the 3rd dof to a mouse is an interesting, sometimes useful, and relatively simple enhancement.

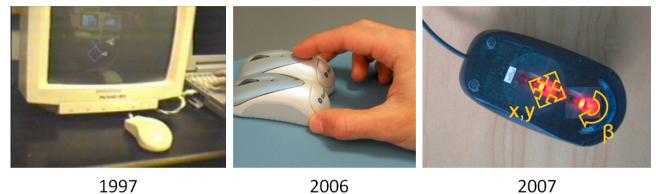


Figure 9. 3dof mouse. The initial research occurred in 1997 with refinements reported about 10 years later.

The 3dof mouse has garnered reasonable interest since the research was published in 1997. Two examples about 10 years on are also shown in Figure 9 (right two images) [1, 10]. The examples use updated technology and propose refinements to the

⁶ A YouTube video demonstrates the operation of the tactile touchpad: <http://www.youtube.com/watch?v=fxfu-Yo6yEk>.

⁷ A YouTube video demonstrates the operation of the 3dof mouse: http://www.youtube.com/watch?v=nQvowU_gzpc.

interaction.⁸ At the present time, however, there are no commercial examples of a 3dof mouse. The research – interesting as it may be – has languished. End of story (perhaps).

7 THE NATURE OF RESEARCH

The first three examples above are reminders of the adage *hindsight is 20 / 20*. It is easy to examine products – successful products – and then look back to their genesis in research. Such stories are plentiful and compelling. However, overly fixating on success stories leads to a skewed understanding of the nature of research. The reality of research is quite different.⁹

In many ways, the fourth example, the 3dof mouse, is the most interesting. The end-of-story closing appends “perhaps”. Will a 3dof mouse eventually appear as a commercial product? Perhaps. Or, perhaps not. But, the story of the 3dof mouse is typical in research – much more so than the stories in the first three examples.

The simple truth is, most research in HCI (and other fields) never arrives in products in any tangible or directly identifiable form. This claim is easily verified. Just pull a research paper at random from a 10-year-old publication, such as a conference proceedings. Peruse the paper and try to identify the product where the research appears. Most likely, there is no such product. If there is a link, it is likely indirect and goes something like this: Research in paper A is cited by and influences research in papers B, C, and D. These papers describe research that was picked up by others who extended or refined the work, as reported in papers E, F, and G. Some of this latter work comes to the attention of engineers or designers for company H, and a variation of the original research arrives in one of company H’s products. That kind of thing.

The absence of a direct link between most research and products is a simple truth, and an inconvenient truth. It is not the sort of outcome researchers put in their CV (or is admitted in a proposal for research funding!). But, the reality is simple: Most research provides knowledge, ideas, context, guidance, tools, examples, etc. The extension to previous work is usually modest – a small *nip* here, a little *tuck* there.¹⁰ And the value to future work is also modest. Put another way, research tends to be incremental, not monumental. This is exemplified in the 3dof mouse.

Are the small nips and tucks a waste of time? Not at all. Importantly, the small contributions create the context and enabling methods and processes (Figure 5) that make later and perhaps more compelling advances possible. It is arguably impossible to predict in advance whether a particular research initiative will have direct value in a future product or will contribute in more modest ways.

At this juncture, it is instructive to consider the words of John Wanamaker on his substantial investment in advertising: *Half the money I spend on advertising is wasted. The problem is, I don’t know which half.*¹¹ The wisdom here is straight: One cannot

⁸ See, as well, the 3dof input methods presented by Lee and Zhou [18] and Lee and Bang [17].

⁹ Discussions of research in this section are directed at the 3rd definition of research given in Section 2.

¹⁰ One meaning for the idiom *nip and tuck* is given at dictionary.com: *Each competitor equaling or closely contesting the speed, scoring, or efforts of the other.*

¹¹ John Wanamaker (1838-1922) was an early proponent of advertising and marketing. His ads directed consumers to the goods sold from his successful department stores in Philadelphia and elsewhere.

simply choose to do research that generates high-impact results. The entirety of the effort is needed.

A healthy program of research is open-ended, inquisitive, without deadlines, and without deliverables. There may be a timeline on the funding, and that’s fine, but constraining research to deadlines and deliverables may stifle the very creativity that is needed – akin to writer’s block. Given the right environment, a deliverable might arise, but requiring research to identify and define the deliverable in advance of conducting the research is absurd. Defining the deliverable *is* the work!

A notable example of the above point occurred in the 1940s at AT&T’s Bell Labs. A group of scientists observed some interesting phenomena when certain stimuli were applied to crystals of germanium. They decided that this was something to look into, to study further. And so they did. At that juncture, they really weren’t sure where their investigations would lead. There was no deliverable, just inquiry and investigation. This is the nature of research. Fast forward a few years and *voila!* – the transistor.¹² The transistor was a truly transformative discovery. All electronic devices, including computers, would change forever. But, nobody asked for the transistor. It emerged from open-ended inquiry, from research.

Unfortunately, there is a growing trend by governments and their funding agencies to ignore or deny the nature of research [30]. This fuels a shift in thinking, whereby universities are expected to serve industry [4]. The term of the day is *partnership*. Opportunities abound for university researchers with an industry partner. For basic research, not so much. An example is the Engage Grants program from Canada’s leading funding agency, the Natural Sciences and Engineering Research Council (NSERC).¹³ Funding is open to university researchers with an industry partner. The program description is an interesting read. There is a lot of talk about research.¹⁴ But this is butted up against the imperative of “specific short-term objectives”. Is the pursuit of specific short-term objectives research? More honestly, it’s product development. An Engage Grant proposal akin to “investigating the effects of certain stimuli on crystals of germanium” would have little chance of success.

Of course, the fruits of research in HCI do not simply crystallize from the ether. Researchers and consumers live in the same material world. Returning to Figure 5, the dashed line from “Products” back to “Materials and Processes” and “Research” should be viewed in the broadest sense possible. It is the “empirical” in empirical research, which is to say, relying on observation or experience. Of course, it’s not just researchers who observe and experience, so too do engineers and designers.

8 THE NATURE OF PRODUCTS (ENGINEERING AND DESIGN)

The results of research move forward and join the Methods and Processes bubble in Figure 5. This bubble holds the palette of possibilities – a mix of research results and state of the art features in existing products. The possibilities are fodder for other researchers and also for engineers and designers to consider and draw upon in creating products. But, the transition from research to products, via engineering and design, is complex and strained due to the vastly different environments in which the three disciplines work.

¹² The scientists were John Bardeen, Walter Brattain, and William Shockley. In 1956 they were awarded the Nobel Prize in Physics for their discovery of the transistor effect.

¹³ <http://www.nserc-crsng.gc.ca> (search for “Engage Grants”)

¹⁴ “Research” appears over 40 times in the program description.

If research is open-ended, products are the stuff of deadlines. Engineers and designers work within the corporate world, developing products that sell, and hopefully sell well. Open-ended product development is a luxury that simply does not exist in the product-driven deadline-driven world of engineering and design.

There are, of course, issues of scale. Engineers and designers will have vastly different experiences in a corporation with, say, 50 employees, compared to a corporation with 50,000 employees. The smaller company is likely to struggle most of the time, for example, getting products to customers, marketing their products, providing customer service, meeting payroll commitments, etc. New products must arrive on tight timelines. Deadlines rule.

The larger company may also struggle, but the scale of the operation provides opportunities that do not exist in a smaller company. There is room for open-ended research and perhaps even some whimsical prototyping (engineering and design) of bold new ideas. But, we digress.

Given the points above, it is worth asking whether engineering and design can be research. Different opinions are likely. In this author's view, the answer is, generally, no. It is perfectly reasonable and common to do research on the design process or the engineering process (e.g., [4, 5]), but the act of designing itself is not research. Design is design. Similarly, engineering is engineering. Engineers and designers combine existing knowledge (materials and processes) in interesting and novel ways. But, that's not research. That's... well... that's what engineers and designers do. The *iPhone* is an excellent example: Apple's engineers and designers combined decades of existing knowledge and research in a successful product (as noted in Section 6.2).

Yet, engineers and designers are often expected to "do research" or to "be researchers". Dig deep on this and it is often apparent that an institutional culture is at work. With the word "research" attached to engineering and design (and a lot of other things), doors open, awards are granted, funds flow, job titles are elevated, and so on. But, the expectation of being a researcher or doing research is, arguably, not a natural one for engineers and designers.

Commentary on this point is offered by Gaver and Boyers [9]. Their *Photostroller* is a classic example of designers doing design. The project involved designing for older people living in a care home. The designers visited the care home to learn the scope of their task, and to understand the loneliness, social interaction, withdrawal, humor, sadness, and so on, of the residents. They sketched, they created design workbooks, and they presented lo-fi prototypes during daylong sessions. Finally... the *Photostroller*, a trolley that shows a slideshow of photographs. See Figure 10 (and [9] for complete details).



Figure 10. The *Photostroller* by designer's Gaver and Boyers [9]. Photos courtesy of Bill Gaver.

Gaver and Boyers also expound on the difficulty of mixing design with research. One section in their paper is titled "How

Can Design Be Research?" An uncomfortable connection is clearly apparent [9, p. 42]:

- *Do we need to add research questions or methodological rigour to design practice for it to count as research?*
- *Do we have to change design practices to make our contributions to HCI look more like research?*
- *Is the result still design, or have we lost something in the process?*
- *These questions have been vexing the HCI design community – and us – for some time. The problem is that novel products alone do not seem sufficient to count as research.*

There is something amiss when designers feel compelled to make their work "count as research". The message here is simple. Designers do design – and they should be allowed to apply their skills without dressing it up as research, to appease an institutional culture that expects something that is not there.

Gaver and Boyers also note that theory was of limited value for their efforts, and that

- *It was by looking at specific examples of practice that we found guidance for our work.* [9, p. 40]

This last point speaks directly to Figure 5 herein. The resources that engineers and designers draw upon are the materials and processes that exist in current products or arrive through research.

9 TIMELINE FOR USER STUDIES AND USABILITY EVALUATIONS

Sections 5, 6, 7, and 8 provide a broad context for an important feature that distinguishes a user study and a usability evaluation – *when*. In short, a user study is done early in the timeline, a usability evaluation late. This is illustrated in Figure 11, which augments Figure 5 to show when each of these assessment methods takes place. A user study is a concluding step in a research initiative. The goal in a user study is to assess a research idea. The idea assessed is typically a small-scale modification to interaction, such as press-to-select for touchpads [21]. The modification is compared with alternative possibilities.¹⁵ A usability evaluation follows engineering and design and is closer to an end product. The assessment is of a complete user interface.

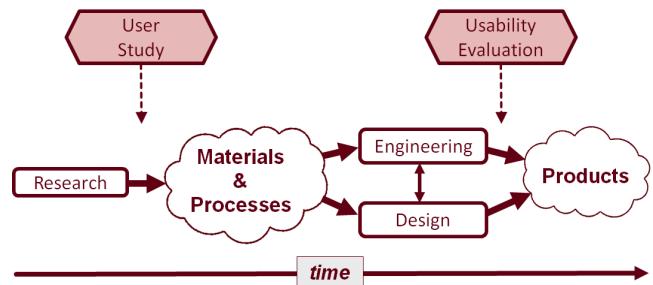


Figure 11. Research-engineering-design timeline. A user study follows research, whereas a usability evaluation follows engineering and design.

From Figure 11, it is clear why a user study is associated with research: It is the final step. The figure suggests that a usability

¹⁵ The interaction of interest and the alternatives are the levels of the manipulated variable. For the cited paper [21], the levels were (i) press-to-select with tactile feedback, (ii) finger lift followed by a down-up tap, and (iii) pressing a separate physical button. The primary response variable was Fitts' throughput [19]. Throughput was highest for the press-to-select condition.

evaluation is far removed from research, since it is further along the research-engineering-design timeline. Of course a usability evaluation is research, since the objective is “collecting information about a particular subject” (see Section 2).

10 CONCLUSION

We have contrasted two testing or assessment methods commonly used in HCI, a user study and a usability evaluation. At least six distinguishing features were presented. These are brought together in Figure 12.

Feature	User Study	Usability Evaluation
Research method	Experimental	Observational
Manipulated variable(s)?	Yes	No
Precision-relevance emphasis	Precision	Relevance
Validity emphasis	Internal	External
Level of inquiry	Low	High
Place in timeline	Early	Late

Figure 12. Contrasting the features of a user study and a usability evaluation.

The features in Figure 12 cover a broad range of topics relating to research, engineering, design and the creation of computing products. Important properties of research, design, and engineering were presented to frame the discussions on user studies and usability evaluations. In particular, a timeline chart was presented showing research as the first step. Research results provide materials and processes. Engineers and designers combine research results with existing practice in creating new computing products. Assessment is critical, both for research (user studies) and for engineering and design (usability evaluation).

ACKNOWLEDGEMENT

Many thanks to Marilyn Tremaine for her unwavering support and encouragement.

REFERENCES

- [1] R. Almeida and P. Cubaud, "Supporting 3D window manipulation with a yawing mouse," in *Proceedings of the 4th Nordic Conference on Human-Computer Interaction - NordiCHI 2006*, New York: ACM, 2006, pp. 477-480.
- [2] K. Arthur. (2008, Oct 9). *The BlackBerry Storm's three-state screen*. Available: <http://www.touchusability.com/blog/2008/10/9/the-blackberry-storms-three-state-touchscreen.html> (accessed 7 Mar 2015).
- [3] K. Arthur. (2008, Oct 14). *Three-state touch: Now Apple has it too!* Available: <http://www.touchusability.com/blog/2008/10/14/three-state-touch-now-apple-has-it-too.html> (accessed 7 Mar 2015).
- [4] M. Blythe, "Research through design fiction: Narrative in real and imaginary abstracts," in *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems - CHI 2014*, New York: ACM, 2014, pp. 703-712.
- [5] T. Brown, *Change by design: How design thinking transforms organizations and inspires innovation*: New York: HarperCollins, 2009.
- [6] W. Buxton, R. Hill, and P. Rowley, "Issues and techniques in touch-sensitive tablet input," in *Proceedings of SIGGRAPH '85*, New York: ACM, 1985, pp. 215-224.
- [7] A. Dix, J. Finlay, G. Abowd, and R. Beale, *Human-computer interaction*, 3rd ed.: London: Prentice Hall, 2004.
- [8] W. K. English, D. C. Engelbart, and M. L. Berman, "Display selection techniques for text manipulation," *IEEE Transactions on Human Factors in Electronics*, vol. HFE-8, pp. 5-15, 1967.
- [9] B. Gaver and J. Boyers. (2012, July + August) Annotated portfolios. *interactions*, 40-49.
- [10] J. Hannagan, "TwistMouse for simultaneous translation and rotation," B. Comm. Dissertation, University of Otago, Dunedin, New Zealand, 2007.
- [11] B. Harrison, K. P. Fishkin, A. Gujar, C. Mochon, and R. Want, "Squeeze me, hold me, tilt me! An exploration of manipulative user interfaces," in *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems - CHI '98*, New York: ACM, 1998, pp. 17-24.
- [12] A. G. Hauptmann, "Speech and gestures for graphic image manipulation," in *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems - CHI '89*, New York, New York: ACM, 1989, pp. 241-246.
- [13] C. F. Herot and G. Weinzapfel, "One-point touch input of vector information for computer displays," in *Proceedings of SIGGRAPH 1978*, New York: ACM, 1978, pp. 210-216.
- [14] K. Hinckley, J. Pierce, M. Sinclair, and E. Horvitz, "Sensing techniques for mobile interaction," in *Proceedings of the ACM Symposium on User Interface Software and Technology - UIST 2000*, New York: ACM, 2000, pp. 91-100.
- [15] M. Y. Ivory and M. A. Hearst, "The state of the art in automating usability evaluation of user interfaces," *ACM Computing Surveys*, vol. 33, pp. 470-516, 2001.
- [16] J. Johnson, T. L. Roberts, W. Verplank, D. C. Smith, C. Irby, M. Beard, and K. Mackey, "The Xerox Star: A retrospective," *IEEE Computer*, vol. 22, pp. 11-29, 1989, Sept.
- [17] B. Lee and H. Bang, "A mouse with two optical sensors that eliminates coordinate disturbance during skilled strokes," *Human-Computer Interaction*, vol. 30, pp. 122-155, 2015.
- [18] K.-M. Lee and D. Zhou, "A real-time optical sensor for simultaneous measurement of three-DOF motions," *IEEE/ASME Transactions of Mechatronics*, vol. 9, pp. 499-507, 2004.
- [19] I. S. MacKenzie, "Fitts' throughput and the remarkable case of touch-based target selection," in *Proceedings of the 16th International Conference on Human-Computer Interaction - HCII 2015*, Berlin: Springer, 2015. p. (in press).
- [20] I. S. MacKenzie, *Human-computer interaction: An empirical research perspective*: Waltham, MA: Morgan Kaufmann, 2013.
- [21] I. S. MacKenzie and A. Oniszczak, "A comparison of three selection techniques for touchpads," in *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems - CHI '98*, Los Angeles, CA, New York: ACM, 1998, pp. 336-343.
- [22] I. S. MacKenzie and A. Oniszczak, "The tactile touchpad," in *Extended Abstracts of the ACM SIGCHI Conference on Human Factors in Computing Systems - CHI '97*, New York: ACM, 1997, pp. 309-310.
- [23] I. S. MacKenzie, R. W. Soukoreff, and C. Pal, "A two-ball mouse affords three degrees of freedom," in *Extended Abstracts of the ACM SIGCHI Conference on Human Factors in Computing Systems - CHI '97*, New York: ACM, 1997, pp. 303-304.
- [24] J. Rekimoto, "Tilting operations for small screen interfaces," in *Proceedings of the ACM Symposium on User Interface Software and Technology - UIST '96*, New York: ACM, 1996, pp. 167-168.
- [25] R. E. Rider, "Position indicator for a display system," U.S. Patent 3,835,464 Patent 3,835,464, 1974.
- [26] T. Selker. (2008, December) Touching the future. *Communications of the ACM*, 14-16.
- [27] D. Sheskin, *Handbook of parametric and nonparametric statistical procedures*, 5th ed.: Boca Raton: CRC Press, 2011.
- [28] B. Schneiderman and C. Plaisant, *Designing the user interface: Strategies for effective human-computer interaction*, 4th ed.: New York: Pearson, 2005.
- [29] I. E. Sutherland, "Sketchpad: A man-machine graphical communication system," in *Proceedings of the AFIPS Spring Joint Computer Conference*, New York: ACM, 1963, pp. 329-346.
- [30] C. Turner, *The war on science: Muzzled scientists and willful blindness in Stephen Harper's Canada*: Vancouver: Greystone Books, 2013.