Shneiderman's Three Pillars of Design

I.   Guidelines Documents
II.   User-interface Management Systems and Rapid Prototyping Tools
III.   Usability Laboratories and Iterative Testing

A really good website about Usability is:
http://www.infodesign.com.au/usability/default.html

Software Usability Testing

Defining Product Usability

User friendly – not just some words on a shrink-wrapped box.
♦ Easy to use
♦ Easy to learn (and teach)
♦ Easy to relearn
♦ Easy to unlearn
♦ Easy to avoid harm
♦ Easy to support
♦ Easy to audit
♦ Easy to share within a group
♦ Easy to integrate into existing operations

PC World, et al, -- the tables show feature comparisons, you have to read to find usability comparisons.
Usability is scientific – its **Testable**.

Defining Software Usability Testing

Why Usability Testing is important

♦ Designers and developer's intuition about a product are not always correct
♦ Designers and developer's terminology do not always match the users
♦ People differ, therefore the "average" user does not exist
♦ Usability design principles and guidelines are not sufficient
♦ Informal feedback is inadequate for product evaluation
♦ Time, money, and resources spent on usability evaluation are worthwhile
♦ Products built in pieces will usually have system-level inconsistencies
♦ Problems found late in the process are more difficult to fix
♦ Problems fixed in development will mean reduces support cost later
♦ Usability evaluations can provide competitive advantages over competitors products
♦ Yourdon – the legal question

In addition to UPA (usability professional's association), there's also ACM SIGCHI and HFES (human factors and ergonomics society).

Product Usability Goals and Objectives

Usefulness
♦ enables a user to achieve his or her goals
♦ an assessment of the user's motivation for using the product
♦ captured by performance data

Effectiveness (ease of use)
♦ defined quantitatively by either speed of performance and error rate, and is tied to some percentage of total users.
♦ e.g.. 95% of all users will be able to load the software correctly on the first attempt in less than 10 minutes.
♦ captured by performance data

Learnability

- ♦ user's ability to operate the system to some defined level of competence after some predetermined amount and period of training.
- ♦ also refer to the ability of infrequent users to relearn the system after periods of inactivity.

Attitude (likability)

- ♦ user's perceptions, feelings, and opinions of the product, usually captured through both written and oral interrogation.

Goal Setting

- ♦ What does it mean to be "easy to use"?
- ♦ Some proposed definitions:
  - • "I like it
  - • "I always do it that way"
  - • "That is the way the xxx system does it"
  - • "It is easy to implement"
- ♦ Much better:
  - • Can be learned in less than 20 minutes.
  - • User will perform 30 error-free operations per minute.
  - • The error rate will be lower than 1 per 40 operations.
  - • Tasks will be performed in 30% of the time it took before the system was used.
  - • Users will have a high satisfaction with the system as measured by a survey.
- ♦ **Explicit, specific, measurable** metrics.
  - A. Time to complete a task
  - B. Percent of task completed
  - C. Percent task complete per unit time
  - D. Ratio of success to failure
  - E. Time spent in errors
  - F. Percent or number of errors
  - G. Percent or number of competitors better than
  - H. Number of commands used
  - I. Frequency of help and documentation used

J. Time spent using help or documentation
K. Percent favorable/unfavorable user comments
L. Number of repetition of failed commands
M. Number of times interface misleads user
N. Number good/bad features recalled by users
O. Number of available commands not involved
P. Number of "regressive behaviors"
Q. Number of runs of success and of failure
R. Number users preferring your system
S. Number of times users need to work around a problem
T. Number of times a user is disrupted from a work task
U. Number of times user loses control of the system
V. Number of times user expresses frustration or satisfaction
W. etc.
- **Caution**
A. Is each attribute measurable in practice?
B. Are users specified clearly enough to find them?
C. Are there resources to measure all attributes?
D. Do all project members agree on each attribute?
E. How well do the attributes capture "usability" for the system?
♦ Tradeoffs, so have to pick relevant metrics.
♦ Measures (Nielsen, chapter 2, pp. 26-37):
- Learnability: Time to learn how to do specific tasks (at a specific proficiency)
- Efficiency: (Expert) Time to execute benchmark (typical) tasks. Throughput.
- Errors: Error rate per task. Time spent on errors. Error severity.
- Subjective satisfaction: Questionaire.
♦ Pick Levels for your system:
- Minimum acceptable level
- Desired (planned) level
- Theoretical best level
- Current level or competitor's level

Characteristic	Type of Usability Evaluation
Efficient	Time (or count clicks or page views) realistic tasks. Must use working versions of the software and plausible sample data.
Effective	Evaluate tasks for how accurately they were completed, and how often they produce errors.
Engaging	User satisfaction surveys or qualitative interviews can gauge user acceptance and attitudes towards the software.
Error Tolerant	Include task scenarios with potential problems in test use scenarios
Easy to Learn	Control how much instruction is given to test participants, or carefully recruit users with different levels of domain knowledge and experience.

In planning usability evaluations, be sure that the most important characteristics are included, and tested in a realistic way.

Cost Benefits of Usability Testing

The costs of not testing

You pay for usability eventually. It may not be a direct expense for up-front testing, but when people use a flawed product, it costs money.

A product that is difficult to use increases a vendor's support costs and decreases a user's productivity. The following lists some of these potential expenses:

Development costs	When you develop a product, the earlier you identify a problem, the less it costs to correct it. Why? because other processes involved in releasing a product are dependent on the product's design. There are users guides, online helps,

	marketing, advertising, promotional material, and training packages that must change with the product. If the product is sold to customers before the flaw is discovered, then the cost of retrofits and retraining get factored in as well. The total cost to correct a product flaw can escalate to 500 times what it would have cost at the design stage!
Documentation costs	Documentation suffers in two areas: A. more manuals are required to support a weak product B. any change to the product will affect one or more of those manuals Revisions to manuals are expensive and do not add value to the product (the vendor usually swallows the cost of revisions). Revising support materials requires unproductive time which ties up resources.
Network support costs	As more systems go on-line, a flawed product has cumulative effects on other network users. For example, when airline reservation operators use flawed transaction processing software, they are forced to make corrections and request the same information several times to perform a single transaction. The extraneous input/output requirements slows down the entire system. Computers are tied up, operators are unproductive, and most importantly, customers are unsatisfied. Finding usability flaws and reducing error rates in such systems decreases the network activity for each transaction.
Training costs	Training costs increase when a computer system is hard to use. When procedures are counter-intuitive, even simple procedures require

	extensive training. The user has to fit into the program's environment instead of the other way around. Not surprisingly, training costs are minimized when the product is easy to use.
Customer support costs	As customers find product flaws, they call the producer for solutions. The amount of support a product requires is an indication of its usability. That is why a certain word processor company has a million dollar a month customer support line. They have a good product, but it is not easy to use.
Vendor credibility	Obviously, a customer who is satisfied with a product will come back when it is time to replace it. Conversely, an unsatisfied customer goes elsewhere the next time. A system which frustrates users is probably going to be replaced, but not by the original vendor! The cost of losing a customer is often more than the original sale. Lost is future resale potential and credibility with other customers. Word of mouth is a powerful force-- usability testing ensures positive customer reactions.

Using Cognitive Psychologists and Usability Professionals

- ◆ They are trained for the task.
  - • They know how to eliminate false causes
  - • They know statistical methods
  - • They know about double blind experiments
  - • Ethics
  - • Release forms
- ◆ Of course they also caused New Coke.

Are Usability Professionals Worth the Cost?

Comparison of Usability Evaluation Methods (UEMs)

Name	Abbreviation	Description	Synopsis	Advantages	Disadvantages
Think Aloud User Testing	Performance Testing	Form of user testing where one user's interaction with a product is videotaped and analyzed to improve the product.	• 1 participant • Uses scenario or tasks • Uses Novices or Domain experts • Users given instruction to "Think Aloud" • Minimal interaction between experimenter and user • Usually videotaped or recorded	• Very close approximation to actual individual usage • Uses less users (less expensive)	• Think aloud instruction rarely works well and is unnatural for the user
Constructive Interaction		Form of user testing where two users interaction with a product is videotaped and analyzed to improve the product.	• 2 participants • One person "drives" • Uses scenario or tasks • Tasks are solved collaboratively • Users given instruction to "Think Aloud" • Uses Novices or Domain experts • Minimal interaction between experimenter and user • Usually videotaped or recorded	• Produces an increased amount of verbal feedback concerning step by step, competitive problem solving methods and opinions. • May model real world behavior when we ask for help. • From anecdotal evidence it appears to find similar usability problems to the 1 participant model.	• Approximates a collaborative problem solving process • Users more users (more expensive)

Focus Groups		Form of data gathering, usually used during the product conceptualization phase in the product cycle where potential users are asked for their opinions on a potential product.	• A group of users is asked for their opinions about a product • Scenarios may be used • Demos may be used	• Produces a large quantity of possible design choices and new features • May improve customer relations	• Is designed for a marketing tool • Is pseudo-scientific and is not a controlled environment • Does not test user's actual interaction with the software • Records what user's think they want, not necessarily what they would actually use. See Nielsen's "The user is always right, the user is not always right."
Expert Review	ER	People having sufficient experience or an advanced degree in a related discipline (HCI experts) critique a product separately or in groups to determine areas in need of improvement.	• No guidelines • No Scenarios or tasks • Uses experts	• Uses experts • May cover entire problem space	• May be difficult to standardize or categorize rationale for design changes
Expert Walkthrough	EW	HCI experts use task scenarios to guide their analysis of the interface.	• No guidelines • Uses scenarios or tasks • Uses experts	• Uses experts • May focus on known problem areas	• Inherent bias because of task selection • Does not cover entire problem space

Heuristic Evaluation	HE	HCI experts separately review an interface and categorize and justify problems based on a short set of heuristics (rules of thumb).	• Uses short guidelines • No scenarios or tasks • Uses experts	• Uses experts • Gives multiple reviewers common rules to site for justification of reviews • Reasonably fast	• The validity of Nielsen's guidelines have been questioned and alternative guidelines exist
Heuristic Walkthrough	HW	Similar to the above definition except task scenarios are used to guide the analysis.	• Uses short guidelines • Uses scenarios or tasks • Uses experts	• Uses experts • May focus on known problem areas	• Inherent bias because of task selection
Guidelines	G	Guidelines range from several hundred items to roughly a thousand. A guidelines analysis judges whether an interface conforms to these rules.	• Uses long guidelines	• May be cheaper due to non-expert evaluators	• Time intensive and tedious • Guidelines conflict and may be biased
Guidelines Walkthrough	GW	Uses the above mentioned guidelines but uses task scenarios to guide the analysis.	• Uses long guidelines • Uses scenarios or tasks	• May be cheaper due to non-expert evaluators • May focus on known problem areas	• Time intensive and tedious • Guidelines conflict and may be biased • Inherent bias because of task selection • Does not cover entire problem space
Cognitive Walkthrough	CW	A method which fully utilizes task scenarios to stress the user's cognitive process and model, which guides the analysis.	• Uses "information processing perspective" which puts the focus on the user's cognitive process and perception • Uses scenarios and tasks	• Puts the focus on the user • May focus on known problem areas • Recognition of user goals	• May be tedious • Tries to make the designer the user • Inherent bias because of task selection • Does not cover entire problem space

A user interface (UI) for a software product was evaluated prior to its release by four groups, each applying a different technique:
- ◆ heuristic evaluation – Four UI specialist
- ◆ software guidelines – Three programmers
- ◆ cognitive walkthroughs – Three programmers
- ◆ usability testing - A tester observing six subjects

Heuristic evaluation by several UI specialists found the most serious problems with the least amount of effort, although they also reported a large number of low-priority problems.

	Advantages	Disadvantages
Heuristic evaluation	Identifies many more problems Identifies more serious problems Low cost	Requires UI expertise Requires several evaluators
Usability testing	Identifies serious and recurring problems Avoids low-priority problems	Requires UI expertise High cost Misses consistency problems
Guidelines	Identifies recurring and general problems Can be used by software developers	Misses some severe problems
Cognitive Walk-through	Helps define users' goals and assumptions Can be used by software developers	Needs task definition methodology Tedious Misses general and recurring problems

You don't need a lot of users to do usability testing. Research has shown that 5 to 6 users will uncover close to 80% of the problems.

Usability: Success or Failure for Today's Products
- ◆ How could a company break the Microsoft Monopoly?
  - • Hard to imagine.
- ◆ How can a company create and maintain a market

- Corel
- Aldus
- Borland
- Adobe

♦ The software must have high quality.
  - Microsoft's web sites tout its usability and give comparison charts.

♦ InfoWorld reviews rate Usability as ~25% of software rating.

The Art of Data Massage: Reliability and Validity

There are lies, damn lies and statistics. - Mark Twain
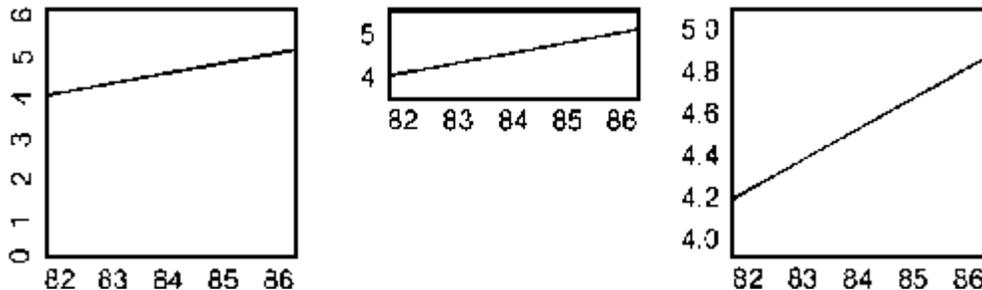
♦ Practically all statistics are based on a sample of a population. So…...
  - how was the sample chosen?
  - how big is the sample?
  - what population does it claim to represent?
  - what population does it actually represent?

♦ Water quality was determined at random locations in a stream - but somehow always get clustered around access roads

♦ Pick your favorite average

Incomes:

$9000

$9000

$9000

$12,000

$120,000

$85,000

$15,000

- Mean = $37,000
- Median = $12,000
- Mode = $9000

♦ Each is a legitimate average but can serve conflicting purposes
♦ 72% of all crow nests in a particular forest are in pine trees
  - Therefore, crows prefer to nest in pine trees.
  - But, 95% of all trees in the forest are pine!

Usability Testing Interface Styles: CUIs and GUIs

♦ The claim (self fulfilling?)
  - Worked Faster
  - Worked better
  - Higher productivity
  - Expressed lower frustration
  - Percived lower fatigue
  - Better able to self-teach and explore
  - Better able to learn more capabilities

But in view of Microsoft sponsorship and above, how valid are these claims?

Usability Testing New Product Versions: Windows 3.1 and Windows 95

http://www.acm.org/sigchi/chi96/proceedings/desbrief/Sullivan/kds_txt.htm

Usability Testing Operating Systems: Windows 95, Macintosh, and OS/2 Warp

http://pla-netx.com/linebackn/guis/

A Usability Challenge: Windows Versus Macintosh

Be sure you know and understand who is doing the test and who the users are. For example, if a Mac product is being tested by people familiar with Mac products, there really is very little learning curve and ease of learning doesn't tell you much.

When Microsoft devised the tests the suggested 20% of the Mac users. They also ran both systems with 16 meg. of memory 8 more than win needs, the min for the Mac OS. Also do't change the color setting!

The Usability Test Report Card

Pages 131-132, a good chart but too much typing for me.

Software Usability Testing—Just Do It!

Personal Firewall Comparison - Page 7/7
Review Date: November 4, 2000
Reviewer: Robert Richmond

Criteria Ratings Overall

	BlackICE	Internet Firewall	McAfee	Sygate	ZoneAlarm
User Interface:	A-	B-	A-	A	A-
Effectiveness:	B	F	A-	A+	A
Customizability:	B-	D	B-	A	B-
Overall Value:	B	D-	B	A+	A-

Comparison Summary

One great disappointment is **Internet Firewall 2000**. Its feature set is limited and the interface lacks good organization. IFW's behavior of opening ports creates an enormous security risk, one that I'm not willing to take. As for its integrated feature set, most of these options can already be found in currently available freeware. I can find no real usefulness for this product compared to the other firewalls covered in this review. I honestly would not waste the time to download the Internet Firewall 2000 demo, let alone spend $39 for the full version.

BlackICE Defender fared better than IFW2K, as it does feature several useful options. Its logging capabilities exceed the usability of any other product in this review. BID could use several improvements, however. A potential security risk is its inability to filter outgoing packet data. As intrusion programs become more sophisticated, the lack of outbound filtering could prove hazardous. Since outgoing connections are not monitored, a malicious program could transfer files and even personal information to any persons wishing to use the data for malicious purposes. Do remember that BlackICE Defender isn't free, and I believe that better freeware packages are currently available.

McAfee Personal Firewall offers decent options to protect a single system for most Internet attacks. With a few updates to the base configuration, this package could be a powerful product. The interface is easily serviceable, especially for first time firewall users. The only major flaw I noted is the lack of viable LAN support features. It is hoped that McAfee will address this issue with a future release, as they have a great foundation to build on.

Zonelabs' ZoneAlarm is a great freeware package. With simple network settings and a great interface, this adds up to a powerful Internet security package. ZA could become one of the best solutions available with a few minor improvements to the base package. The addition of more advanced configuration options would be useful, and would complete the already impressive feature set of this package. Even without this feature, ZoneAlarm proves to be a valuable product. I would have little problem in recommending this product to anyone seeking a decent personal firewall.

I would designate **Sygate Personal Firewall** as the most comprehensive personal firewall solution reviewed. SPF's robust feature set combined with an efficient interface makes it suitable for all Internet users, whether they be novices or network administrators. With a freeware personal version available, I would recommend that all Internet users take a look at this product. The current version offers excellent security, and I only see SPF improving with future releases. If you only try one product, then Sygate Personal Firewall should be it.

The results of this firewall comparison are rather interesting. In most scenarios, the freeware applications were capable of meeting and even exceeding the capabilities of the purchased packages. It is refreshing to see that freeware applications are still viable, especially in this era when most useful products are tagged with costly prices.