An Introduction to Usability Testing

Bill Killam, MA CHFP
Adjunct Professor
University of Maryland
hkillam@umd.edu
(bkillam@user-centereddesign.com)

Caveat

Caveat

The title of this tutorial is accurate…

…as long as we replace the terms “usability” and “testing”

Background

Origin of the Species

- “Usability testing” is the common name for multiple forms both user and non-user based system evaluation focused on specific aspects of a product design – the user experience
- Done for many, many years prior, but popularized in the media by Jakob Neilson in the 1990’s

What does “usability” mean?

- ISO 9241
  - “Extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.”
- ISO 9126
  - “A set of attributes that bear on the effort needed for use, and on the individual assessment of such use, by a stated or implied set of users.”

Simplified

- Usability is the ability of a specific group of people to perform specific tasks in a specific environment.
- Usability testing would, therefore, be *testing* the ability of a specific group of people to perform specific tasks in a specific environment.
- But the ability for users to perform tasks alone is not all we care about.

Elements of the User Experience

- Accessibility
  - A precursor to usability: if users cannot gain access to the product, all other elements of the experience are moot points
- Functional Suitability
  - Does the product contain the functionality required by the user? This is the product's utility, but related to usability in terms of desirability to try to use it.
- Functional Discoverability
  - Can the user “discover” the functions of a product?
- Ease-of-learning
  - Can the user figure out how to exercise the functionality provided once it has been discovered?
- Ease-of-use
  - Can the user exercise the functionality accurately and efficiently once its learned?
- Ease-of-recall
  - Can the knowledge of operation be easily maintained over time?
- Safety
  - Can the user operate the system in relative safety, and recover from errors?
- Subjective Preference
  - Do user’s like using it?

Conclusion

- We are not just interested in “usability”, we are interested in several, interrelated aspects of a product, including multiple types of use related elements (ease of learning, ease of use, and possibly ease of recall), depending on the product.

What does it mean to “test” something?

- Definition

noun: test; plural noun: tests

1. a procedure intended to establish the quality, performance, or reliability of something, especially before it is taken into widespread use.
2. the means by which the presence, quality, or genuineness of anything is determined
3. an act of using something to find out whether it is working correctly or how effective it is

- To conduct a true test, the test itself has to be valid

Validity

- “Validity is the degree to which the results of a research study provide trustworthy information about the truth or falsity of the hypothesis.”*
- Construct validity is the degree to which a test measures what it purports to measure
  - e.g., Is someone's opinion about a product’s ease of use usability an accurate measurement of a product actual ease of use?
- Content validity is the extent to which a measure represents all facets of a construct
  - e.g., Is time on task a complete measure efficiency?
- Criterion validity is the extent to which a measure is consistent at with other measures taken (concurrent validity) or if measured in the future (predictive validity)
  - e.g., Are observed or measured levels of difficulty consistent with reported levels of difficulty (concurrent validity)?
  - e.g., Would you perform a task with errors today, but perform differently on the same task if attempted later?

*Cherulnik, P.D. 2001. Methods for Behavioral Research: A Systematic Approach

Validity (continued)

- Internal validity refers to the situation where the "experimental treatments make a difference in this specific experimental instance."** In other words, how you set up and run a study determines if you have internal validity.
- Threat to internal validity include
  - Recruitment/selection bias
  - Any interference with the participants, including the mere presence of a moderator and/or observer
  - Differences in how task are administered
  - Lack of objective measures

**Cambell, D.T. & Stanley, J.C. (1963) Experimental and Quasi-experimental Designs for Research

Validity (concluded)

- External validity asks the question of “generalizability” – can the result from the experiment correctly predict the behavior of the larger audience they represent.
- Threats to external validity include
  - Small sample size
  - Lack of a representative sample
  - Differences in environment (lab versus real life)

“Test” Types

- Within the user-centered design process, two types of “tests” are described
  - Formative
  - Summative
- “Formative” usability tests are intended to be diagnostic. The UPA says this type of testing is used to “form” the design. Formative testing is recommended throughout the design process. By nature, they will be qualitative in nature.
- “Summative” usability tests are intended to “establish a baseline” or to “compare the product against usability requirements.” Summative testing is intended to be done on a completed design. They can be quantitative in nature, but this is not common.

The Law of Large Numbers

- The law of large numbers states that very large sample will be representative of the population from which it is drawn.
- It works because differences in population members cancel each other.

Example: A Cancer Study

- A study of the incidence of kidney cancer in the 3,141 counties of the United States reveals the following fact:
- The counties in which the incidence of kidney cancer is highest are mostly rural, sparsely populated, and located in traditionally Republican states in the Midwest, the South, and the West.

Example: A Cancer Study (concluded)

- The study also revealed the following fact:
- The counties in which the incidence of kidney cancer is lowest tend to be mostly rural, sparsely populated, and located in traditionally Republican states in the Midwest, the South, and the West.

The Law of Small Numbers

- Imagine a large urn filled with marbles. Half the marbles are red, half are white. The experimental procedures is to draw marbles, record the colors, throw the balls back into the urn, and then do it all again and repeat.
- Person one draws balls seven marbles at a time. They will find that the outcome "two red, two white" occurs 6 times as often as the extreme outcome or "four red" or "four white."
- Person two draws balls marbles four at a time. They will find that the outcome extreme outcome "four red" or "four white" occurs 8 times more often than the person who drew marbles 7 at a time
- The actual probabilities of the extreme values in these experiments are 1.56% for marbles drawn 7 at a time and 12.5% percent for marbles drawn 4 at a time.

The Law of Small Numbers (concluded)

- Suppose you have run an experiment on 20 subjects, and have obtained a statistically significant result which confirms your theory ($p < .05$, two-tailed).
- You now have a chance to run an additional group of 10 subjects.
- The probability that the second group will show statistically significant results is about 50/50.

Typical Participant Sizes

- The behavioral research rule of thumb when working with human subjects is that you need a minimum of 25-30 people before you'll see data begin to regress to the mean assuming a medium effect size
- You could do statistical analysis on the results of any usability testing, but this analysis is valid only if your test is known to be both valid and reliable
- You can generalize to the larger population only if you had a representative sample in the study, but your confidence interval is critical to include

Typical Participant Sizes (concluded)

- Usability is typically done with very few people per round
  - Neilson says you only need 5 people (but not for the right reason)
  - Krug says you only need 2 or 3 people (also not for the right reason)
  - The IUSR and the related ISO standard says 3 per user group, profile, or persona but other Federal documents (e.g., EHR Testing Guidelines) suggest 10-15 participants
- A single day of testing can test with, at most, 8-9 people

Use of Confidence Intervals

- When working with samples, a confidence interval provide a way to represent the uncertainty in test results. And there is always a level of uncertainty.
- Since each sample and each test is different, the confidence level tells the reader the likelihood that another sample will provide the same results. (In other words, if you ran the test again, what value are you likely to get the next time?)
- Typical confidence intervals in research include the 90% or 95% confidence interval, though behavioral research often uses a 80% confidence interval.

Use of Confidence Intervals (concluded)

- Confidence intervals when testing with, say, 8 people range from 37% (0 out of 8 or 8 out of 8) to between 50%-70% (all other values)
  - For example, if 6 out of 8 people successfully completed a task in your test, you can only predict that somewhere between 20% and 97% of all people would complete the task (assuming all conditions for validity and reliability have been met)
  - Alternately, if you want to confidently state, based on your testing, that 90% of all users will be able to successfully complete a task, you need to test 430 people and 400 of them have to successfully complete the task
- None of this even matters if all conditions of validity have not been met

A Valid Summative (Quantitative) Research Example

Example 1

- You own a company that sells product on the web. You have always required people to register to purchase from the site. It has been suggested that sales would increase if you allowed people to purchase as a “guest.” (Your hypothesis.)
- You operationalize your dependent variable as an objective measure with good contract validity (sales or number of abandoned shopping carts)
- You create a new design for your site that allows people to purchase without registering. You make NO OTHER CHANGES.
- You set up two servers – one hosts the current design and one hosts the new design (Your independent variables.)
- For a one month (or more) period, you run both servers. Every other person who comes to the site is routed to the other server. If you have large traffic, you will have near equivalent groups in both populations. (Your sample population.)
- If there is a difference in sales in favor of the new site design, you can conclude that adding the ability to check out as a guest is the likely cause of higher sales. Knowing the value of N (your population sizes), statistics can provide a confidence interval for this finding.

Why this works

- Large samples of populations are possible, making it likely to generate representative samples
- Ordered assignment of participants to each server works due to the random nature of each visitor.
- To population performing tasks at the same time avoid confounding variables
- The participants are unaware of their participation so observer effects, projected responding, etc. are avoided

Limitations

- You can never prove a hypothesis. You can only see if you have sufficient data to reject the alternative to your hypothesis (the antithesis or a null hypothesis). You still have a level of uncertainly, though the probability of this in this case is likely to be very low.
- Some participant may have returned and become part of the other participant group. so there is likely some noise in the data.
- You have to be willing to experiment on your users (what if you are wrong in your hypothesis and sales went down?)
- You have to have a completed design to use or a fully operational simulation.

Other Applications

- You could compared a design system to benchmarks (e.g., 90% people should be able to complete the purchase process within 4 minutes), but you need question the validity of the requirement.
- It would be possible, but more difficult, to try to measure subjective aspects of the user experience (e.g., satisfaction, efficiency, functional suitability, functional discoverability), and be careful how you collect this data (internal validity).Subjective data is inherently unstable.

Example 2

- You have 3 potential designs for break lights - standard (your control), redundant, and redundant centered (CHMSL)
- You operationalize your dependent variables as an objective measure with good construct validity (reduction in rear end collisions, reduction in costs of repeat for rear end collisions)
- You obtain 3 sets of equivalent drivers operating in a real world environment (e.g., cab drivers in NYC)
- After the end of the experiment trail, you measure differences in the number and cost of rear end collisions
- Statistical evaluations will tell you if there is a statistical significant differences in your dependent variables

Limitations

- You can never prove a hypothesis. You can only see if you have sufficient data to reject the alternative to your hypothesis (the antithesis or a null hypothesis). You still have a level of uncertainly, though the probability of this in this case is likely to be very low.
- This has good internal validity but poor external validity (generalizability) since you don't know how other drivers in other environment will compare. (Results from the original study was an approx. 50% reduction. Results for the general public the first year of incorporation was 8.5%, CI 6.1-10.9.)
- You have to be willing to experiment on your users (what if you are wrong in your hypothesis and accidents went up?)
- You have to have a completed design to use or a fully operational simulation.

Example 3

- You have a design and ask 9 people to evaluate it in a lab (a typical scenario)
- Performance is not on an interval scale and you have no comparison to make
- You have little or no internal validity – your presence alone as well as the altered environment introduces lots of confounding valuables.
- You have no external validity – your audience is too small to be representative
- Count anything you want, you get no useful quantitative data from this type of an evaluation

Psychology Primer

What are we really testing?

Consciousness

- “Your consciousness is like a tiny stowaway on a transatlantic steamship taking credit for the journey without taking into account the massive engineering underfoot.” - David Eagelman, *Incognito: The Secret Lives of the Brain*

2 System Theory

- The car versus elephant analogy. - Daniel Gilbert, *Stumbling on Happiness*
- Kannaman and Taversky
  - System 1: Automated and Unconscious Processes
  - System 2: Conscious Processes

System 1: Unconscious/Automated Processes

- Fast processing of information
- Multi threaded processing
- Used to construct our understanding (perception) of the world

- >90% (perhaps >99%) of our daily functioning is precessed at the system 1 level

System 2: Conscious Awareness (Attention)

- Slow processing of information
- Highly limited in processing capacity and focus
- Generally just monitors system 1 (though we assume its doing more).

- Able to influence system 1 processing, but not “in control”
- Nearly always unaware of the effects of system 1 processing on decision making.

How fast is fast?

- Nolan Ryan's fast ball was clocked at 101 mph. At that speed, the ball will cross the 60' 6" distance to home plate in 0.04 seconds. That means, in order to hit a fast ball, the signal must reach the eye, be sensed by the occipital node of the brain, be processed, a signal sent to the motor cortex, signals sent to the arms and legs, and the body move in that much time.
- But conscious processing of input data does not even begin for .05 seconds after input is received.

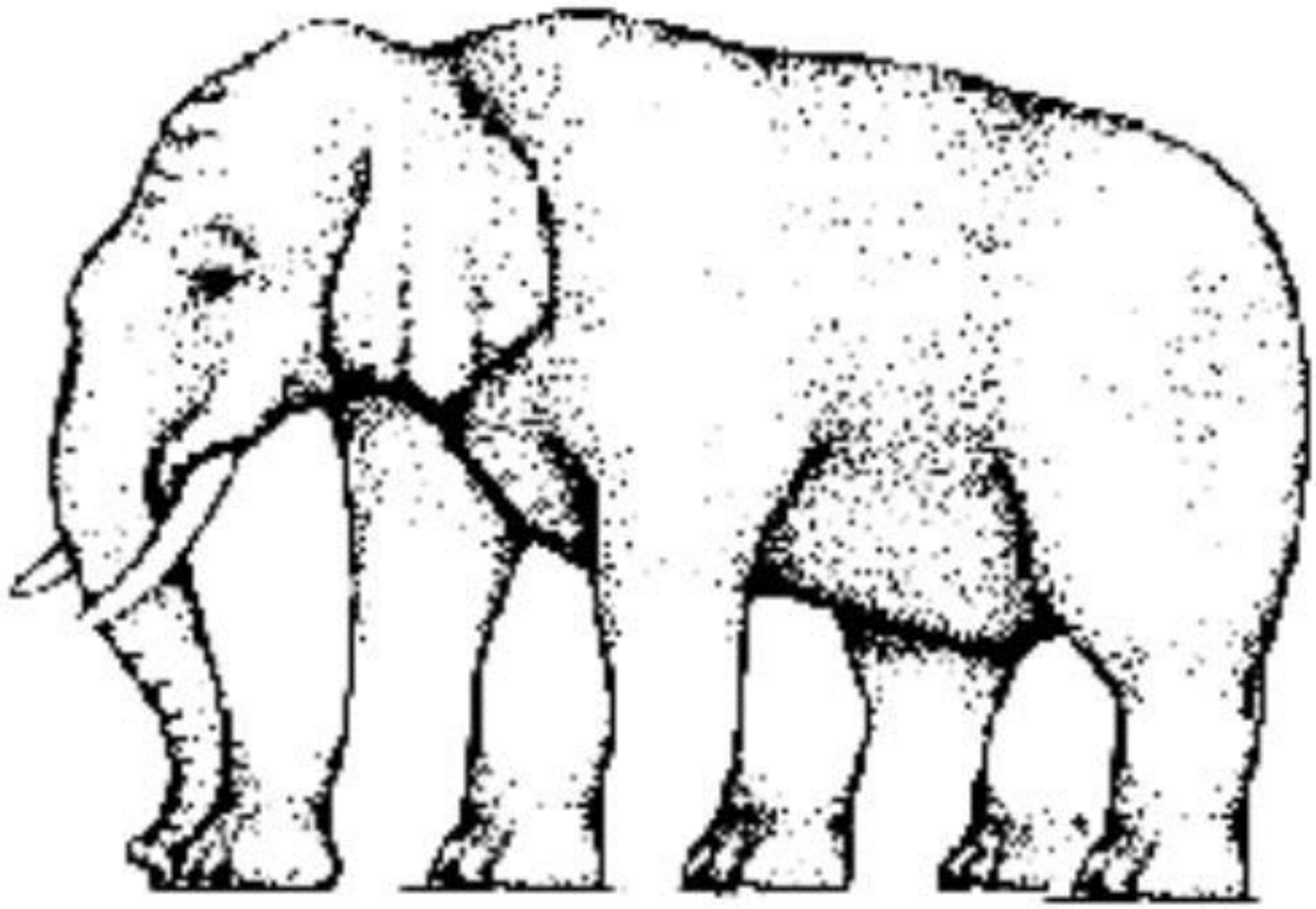
Attentional Blindness

https://www.youtube.com/watch?v=Ahg6qcgoay4

Change Blindness

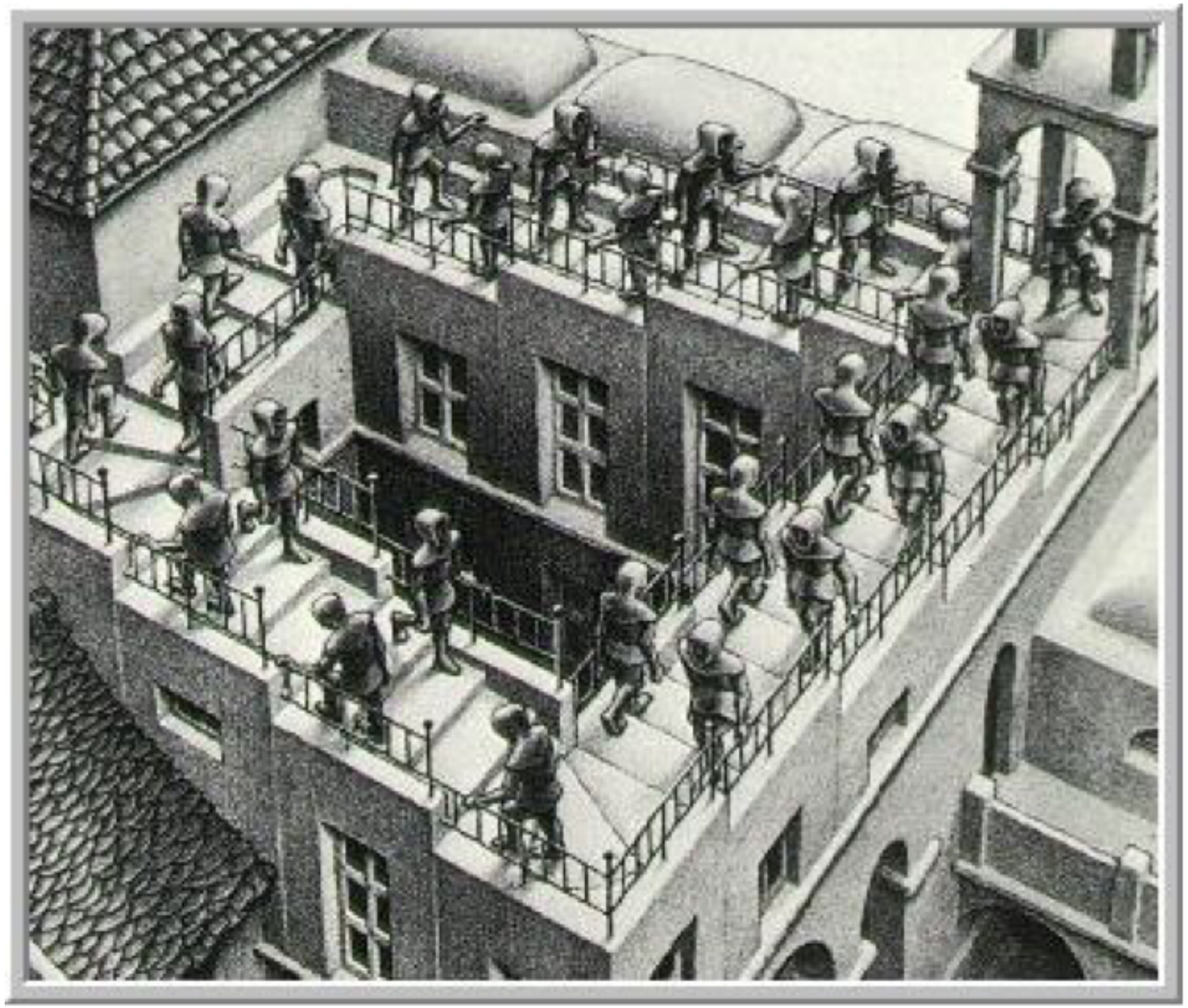
- Example 1
- Example 2

Limits of Attentional Focus



How many legs does this elephant have?

Limits of Attentional Focus (concluded)



Developing Expertise

- We (usually) need conscious awareness to learn an activity, but as we become proficient, even expert, the thinking and decision making moves from level 2 (conscious) to level 1 (unconscious)
- Consider driving a car. When you first leaned to drive a car, it required all of your attention. You could not (should not) listen to the radio, engage in a conversation, etc. But as you became more skilled, you moved the activity from conscious (level 2) thinking and decision making to non conscious (level 1) thinking and decision making
- However, as a result, you are also no longer conscious of what you're doing while driving

Goal of Interaction Design

- The goal of interaction design is to allow product interaction (*how* we do what we are doing) to occur (ideally) as all non conscious (level 1) thinking, thus allowing our limited, single threaded conscious attention to focus on the goal (*what* we are doing). Ideally, this would be to the point we don't even notice the device we used to get the job done.

- The less often we have to redirect our attention from our task to attend to how we accomplish the task, the more *transparent* the product design, the easier it is to use, the less errors we make, the faster we work, the happier we are.

Example

USER-CENTERED DESIGN

Loudoun County Public Schools / Overview

www.lcps.org/loudoun

Public Surpl...lus Auctions | INFM605-Fall 2013 | Entrepreneurs | 401(k) | We are hirin...ility People | Toyota Financial Services | 2016-12?ym...mini=2016

DISTRICT | Select A School | Translate this page from... | SIGN IN | REGISTER | Search This Site...

"A Climate for Success"

LOUDOUN COUNTY PUBLIC SCHOOLS

SCHOOL MENU | SCHOOL ATTENDANCE BOUNDARIES | CONTACT US! | CALENDAR OF EVENTS

PARENTS | STUDENTS | STAFF

Home | Our District | Academics | Administration | Employment | Information | Purchasing | School Board | Services

National African-American History Month Activities

National African-American History Month is always observed the entire month of February.

- National African-American History Month Activities
- Douglass History Presentation February 9th at Balch Library
- Briar Woods Presents 'Guys and Dolls'
- Stone Hill Middle School Leo Club Induction
- Monroe Technology Center Open House
- Trailside Middle School Mascot Revealed

view full ARTICLE | view all HEADLINES | view NEWS ARCHIVES

WHAT DO YOU WANT TO DO?

UPCOMING EVENTS

Today

5:30 PM - 7:00 PM **Health, Heath, Safety, Wellness & Transportation Committee**

6:30 PM - 8:30 PM **Community Public Meeting on Middleburg Charter School**

Tomorrow

5:30 PM - 7:00 PM **Adhoc Committee on AOS/MATA facility**

5:30 PM - 7:30 PM **Legislative & Policy Committee Meeting**

6:30 PM - 10:30 PM **School Board Work Session & Adoption - FY15 Operating Budget**

View Calendar

ANNOUNCEMENTS

Featured Video

Raiders Girls Varsity Volleyball Ranked 18th in Nation

Watch

YourVoice

Powered by Kiz Insight

Loudoun County Public Schools listens to your views on critical issues. The results of the Community Budget Priorities Survey were presented during the School Board meeting on January 14th and are viewable here. Thank you for your participation.

QUICK LINKS

- Apply Online Now
- Bus Routes
- Budget (Operating) Information
- BoardDocs - School Board Information
- Closing and Delay Information
- Connect-Ed Phone Messaging System Opt-Out Form
- CTE Non-discrimination policy
- District Profile
- Energy & Environment
- Fine Arts
- FOIA Requests
- FY 14 Salary Scale
- High School Athletics
- Kindergarten
- LCPS Comment and Complaint Form

USER-CENTERED DESIGN

About / Parent Information tab

www.lcps.org/site/Default.aspx?PageID=21

DISTRICT | Select A School | Translate this page from... | SIGN IN | REGISTER | Search This Site...

"A Climate for Success"
LOUDOUN COUNTY
PUBLIC SCHOOLS

SCHOOL MENU | SCHOOL ATTENDANCE BOUNDARIES | CONTACT US! | CALENDAR OF EVENTS

PARENTS | STUDENTS | STAFF

Home | Our District | Academics | Administration | Employment | Information | Purchasing | School Board | Services

About

- Overview
- 2013-2014 School Facts
- Directory
- Directions
- District Profile
- School Hours
- Closing and Delay Information
- Parent Liaison
- 2014 Graduation Dates
- 2013-2014 Student Calendar
- 2013-14 Calendario Escolar Estudiantil
- 2013-2014 Teacher & Staff Calendar
- 2013-2014 Key Dates Calendar
- 2014 - 15 Student Calendar

Home > Our District > About

Parent Information

(Click on link to view information)

- Academic Program of Studies (high school and middle school)
- Assistive Technology
- Bullying Prevention and Awareness
- Bus Schedules
- Calendar
- Career and Transition Services
- CAMPUS
- CAMPUS (Spanish)
- Child Find Center
- Closing and Delays
- Connect-Ed Opt Out Form
- Diagnostic and Prevention Services
- District Information Profile
- Driver Education (Behind-the-Wheel) Information
- Elementary School Report Card Manuals
- Eligiblity for Special Education and Section 504
- Epinephrine Administration Policy
- FOIA Request
- FAQ's on Virginia High School League (VHSL) Transfer Rule
- Kindergarten Information
- Home Instruction
- LCPS Complaint Form
- Loudoun Education Foundation (LEF)
- Loudoun Education Alliance of Parents (LEAP)
- Loudoun School - Business Partnerships
- Medication at School
- Minority Student Achievement Advisory Committee
- Online Fee Payments
- Online Mathematics Resources
- Parent Emergency Guide (English)
- Parent Emergency Guide (Spanish)
- Parent Information Management System (PIMS)
- Parent Resource Center
- Positive Behavioral Interventions and Supports (PBIS)
- Promotion and Retention Guide
- Reconsideration of Instructional Materials Request
- School Counseling Services
- School Board Policy 6-21 Appeal Form (New Posting)
- Speaking to the School Board
- Student Registration
- Special Education Advisory Committee
- Special Permission
- Sports Physical Registration
- Summer School Registration
- Student Health Services
- Substance Abuse Prevention Program
- Tranfer to Under-Enrolled Schools Policy

Last Modified on December 3, 2013

Go to "http://www.lcps.org/site/Default.aspx?PageID=21"

USER-CENTERED DESIGN

LCPS – Payment Portal

webinter.lcps.org/paymentportal/

"A Climate For Success"
LOUDOUN COUNTY
Public Schools

PAYMENT PORTAL

Payment Portal Programs

- Advanced Placement Test Fees
- PSAT - Academy of Science Applicants Only (LCPS Students)
- PSAT - Academy of Science Applicants Only (Non-LCPS Students)
- Student Athletic Fees
- Student Behind The Wheel
- Student Parking
- Student Parking - Monroe Tech Center
- Transcript Fee-FORMER STUDENTS ONLY
- Summer in the Arts
- Summer School Payment
- Teacher License

Refund/Return Policy | Privacy Policy | Feedback

Go to "http://webinter.lcps.org/paymentportal/public/HandlePaymentWizard.aspx?cl=1&programName=PSAT"

USER-CENTERED DESIGN

LCPS – Payment Portal

webinter.lcps.org/paymentportal/public/HandlePaymentWizard.aspx?cl=1&programName=ap_fees

"A Climate For Success"
LOUDOUN COUNTY
Public Schools

PAYMENT PORTAL

Loudoun County Public Schools - Payment Portal

Step 1: Select the items you would like to pay for

Available Items

To add an item to the shopping cart: click on the item then click on the 'Add selected item' button or drag the item to the shopping cart.

Item	Description	Cost
AOS AB Calculus - AP w/ Statistics (Academy of Science Students Only)	Advanced Placement Course	$83.41
AOS BC Calculus - AP w/Statistics (Academy of Science Students Only)	Advanced Placement Course	$83.41
AOS Biology - AP (Academy of Science Students Only)	Advanced Placement Course	$83.41
AOS Chemistry - AP (Academy of Science Students Only)	Advanced Placement Course	$83.41
AOS Environmental Science - AP (Academy of Science Students Only)	Advanced Placement Course	$83.41
AOS Physics - AP (Academy of Science Students Only)	Advanced Placement Course	$83.41
Biology - AP	Advanced Placement Course	$83.41
Calculus AB - AP	Advanced Placement Course	$83.41
Calculus BC - AP	Advanced Placement Course	$83.41
Chemistry - AP	Advanced Placement Course	$83.41
Comparative US Government - AP	Advanced Placement Course	$83.41
Computer Science A - AP	Advanced Placement Course	$83.41
Economics - Macro - AP	Advanced Placement Course	$83.41
Economics - Micro - AP	Advanced Placement Course	$83.41
Environmental Science - AP	Advanced Placement Course	$83.41
European History - AP	Advanced Placement Course	$83.41
French - AP	Advanced Placement Course	$83.41
German - AP	Advanced Placement Course	$83.41
Human Geography - AP	Advanced Placement Course	$83.41
Language & Composition - AP	Advanced Placement Course	$83.41
Latin - AP	Advanced Placement Course	$83.41
Literature & Composition - AP	Advanced Placement Course	$83.41
Music Theory - AP	Advanced Placement Course	$83.41
Physics C - AP	Advanced Placement Course	$83.41
Psychology - AP	Advanced Placement Course	$83.41
Spanish - AP	Advanced Placement Course	$83.41
Statistics - AP	Advanced Placement Course	$83.41
Studio Art - AP	Advanced Placement Course	$83.41
U.S. Government & Politics - AP	Advanced Placement Course	$83.41
U.S. History - AP	Advanced Placement Course	$83.41
World History - AP	Advanced Placement Course	$83.41

A $81.00 Advance Placement Test fee (AP) will be administered for each AP test the student chooses to take. Fees are to be paid in advance before the test is taken. Payment of the fee does not guarantee a testing slot as all requirements must be met and applications accepted by your school in advance. In addition, payment of the testing fee does not guarantee a passing grade. Payment of the AP fee will be required for all AP tests taken. Although the Loudoun County School Board strongly believes students benefit from the Advance Placement program, this program is voluntary. On line payment of the AP testing fees is $83.41 per test. Parents and students who wish to receive a discount of $2.41 may elect to pay at the school site with cash, check, or money order. All refunds will be issued after the student fills out a refund request form and has it approved by the high schools director of guidance. Refunds will be submitted via check.

USER-CENTERED DESIGN

LCPS – Payment Portal

webinter.lcps.org/paymentportal/public/HandlePaymentWizard.aspx?cl=1&programName=ap_fees

"A Climate For Success"
LOUDOUN COUNTY
Public Schools

PAYMENT PORTAL

Loudoun County Public Schools - Payment Portal

Step 1: Select the items you would like to pay for

Available Items

To add an item to the shopping cart: click on the item then click on the 'Add selected item' button or drag the item to the shopping cart.

Item	Description	Cost
AOS AB Calculus - AP w/ Statistics (Academy of Science Students Only)	Advanced Placement Course	$83.41
AOS BC Calculus - AP w/Statistics (Academy of Science Students Only)	Advanced Placement Course	$83.41
AOS Biology - AP (Academy of Science Students Only)	Advanced Placement Course	$83.41
AOS Chemistry - AP (Academy of Science Students Only)	Advanced Placement Course	$83.41
AOS Environmental Science - AP (Academy of Science Students Only)	Advanced Placement Course	$83.41
AOS Physics - AP (Academy of Science Students Only)	Advanced Placement Course	$83.41
Biology - AP	Advanced Placement Course	$83.41
Calculus AB - AP	Advanced Placement Course	$83.41
Calculus BC - AP	Advanced Placement Course	$83.41
Chemistry - AP	Advanced Placement Course	$83.41
Comparative US Government - AP	Advanced Placement Course	$83.41
Computer Science A - AP	Advanced Placement Course	$83.41
Economics - Macro - AP	Advanced Placement Course	$83.41
Economics - Micro - AP	Advanced Placement Course	$83.41
Environmental Science - AP	Advanced Placement Course	$83.41
European History - AP	Advanced Placement Course	$83.41
French - AP	Advanced Placement Course	$83.41
German - AP	Advanced Placement Course	$83.41
Human Geography - AP	Advanced Placement Course	$83.41
Language & Composition - AP	Advanced Placement Course	$83.41
Latin - AP	Advanced Placement Course	$83.41
Literature & Composition - AP	Advanced Placement Course	$83.41
Music Theory - AP	Advanced Placement Course	$83.41
Physics C - AP	Advanced Placement Course	$83.41
Psychology - AP	Advanced Placement Course	$83.41
Spanish - AP	Advanced Placement Course	$83.41
Statistics - AP	Advanced Placement Course	$83.41
Studio Art - AP	Advanced Placement Course	$83.41
U.S. Government & Politics - AP	Advanced Placement Course	$83.41
U.S. History - AP	Advanced Placement Course	$83.41
World History - AP	Advanced Placement Course	$83.41

A $81.00 Advance Placement Test fee (AP) will be administered for each AP test the student chooses to take. Fees are to be paid in advance before the test is taken. Payment of the fee does not guarantee a testing slot as all requirements must be met and applications accepted by your school in advance. In addition, payment of the testing fee does not guarantee a passing grade. Payment of the AP fee will be required for all AP tests taken. Although the Loudoun County School Board strongly believes students benefit from the Advance Placement program, this program is voluntary. On line payment of the AP testing fees is $83.41 per test. Parents and students who wish to receive a discount of $2.41 may elect to pay at the school site with cash, check, or money order. All refunds will be issued after the student fills out a refund request form and has it approved by the high schools director of guidance. Refunds will be submitted via check.

LCPS – Payment Portal

webinter.lcps.org/paymentportal/public/HandlePaymentWizard.aspx?cl=1&programName=ap_fees

AOS AB Calculus - AP w/ Statistics (Academy of Science Students Only)	Advanced Placement Course	$83.41
AOS BC Calculus - AP w/Statistics (Academy of Science Students Only)	Advanced Placement Course	$83.41
AOS Biology - AP (Academy of Science Students Only)	Advanced Placement Course	$83.41
AOS Chemistry - AP (Academy of Science Students Only)	Advanced Placement Course	$83.41
AOS Environmental Science - AP (Academy of Science Students Only)	Advanced Placement Course	$83.41
AOS Physics - AP (Academy of Science Students Only)	Advanced Placement Course	$83.41
Biology - AP	Advanced Placement Course	$83.41
Calculus AB - AP	Advanced Placement Course	$83.41
Calculus BC - AP	Advanced Placement Course	$83.41
Chemistry - AP	Advanced Placement Course	$83.41
Comparative US Government - AP	Advanced Placement Course	$83.41
Computer Science A - AP	Advanced Placement Course	$83.41
Economics - Macro - AP	Advanced Placement Course	$83.41
Economics - Micro - AP	Advanced Placement Course	$83.41
Environmental Science - AP	Advanced Placement Course	$83.41
European History - AP	Advanced Placement Course	$83.41
French - AP	Advanced Placement Course	$83.41
German - AP	Advanced Placement Course	$83.41
Human Geography - AP	Advanced Placement Course	$83.41
Language & Composition - AP	Advanced Placement Course	$83.41
Latin - AP	Advanced Placement Course	$83.41
Literature & Composition - AP	Advanced Placement Course	$83.41
Music Theory - AP	Advanced Placement Course	$83.41
Physics C - AP	Advanced Placement Course	$83.41
Psychology - AP	Advanced Placement Course	$83.41
Spanish - AP	Advanced Placement Course	$83.41
Statistics - AP	Advanced Placement Course	$83.41
Studio Art - AP	Advanced Placement Course	$83.41
U.S. Government & Politics - AP	Advanced Placement Course	$83.41
U.S. History - AP	Advanced Placement Course	$83.41
World History - AP	Advanced Placement Course	$83.41

A $81.00 Advance Placement Test fee (AP) will be administered for each AP test the student chooses to take. Fees are to be paid in advance before the test is taken. Payment of the fee does not guarantee a testing slot as all requirements must be met and applications accepted by your school in advance. In addition, payment of the testing fee does not guarantee a passing grade. Payment of the AP fee will be required for all AP tests taken. Although the Loudoun County School Board strongly believes students benefit from the Advance Placement program, this program is voluntary. On line payment of the AP testing fees is $83.41 per test. Parents and students who wish to receive a discount of $2.41 may elect to pay at the school site with cash, check, or money order. All refunds will be issued after the student fills out a refund request form and has it approved by the high schools director of guidance. Refunds will be submitted via check.

Add selected item(s) to shopping cart

Shopping Cart

To remove an item from the shopping cart: click on the item then click on the 'Remove selected item' button or drag the item out of the shopping cart.

Item	Description
The shopping cart is empty	
	Total:

Next

Refund/Return Policy | Privacy Policy | Feedback

Evaluation Goals

System 2 Testing

- Functional discoverability
- Ease of Learning (Intuitiveness)
  - Obvious visual affordances
  - Conforming ot industry standards and best practices or a clear and useful violation
  - Consistency of percepts
- Minimal cognitive demand

System 1 Testing

- Ease of use
  - Good conceptual model
  - Good use of visual design principles
  - Good positional feedback
- Ease of recall (hard to test)
- Safety

Testing Applications

- If the task is to perform a specific function (buy a product, get direction, etc.), then operating the system and performing the task are highly integrated
- Tasks and sub-tasks are easy to define (locate a specific product, add items to a shopping cart, update a shopping cart, locate contact info, find directions between 2 specific locations, etc.)
- Task and subtask success is easily measured (success or failure)
- This is the simplest type of product to test.

Testing: Information Systems

- If the task is information-based, it has one of several purposes
  - Gain knowledge
  - Be persuaded
  - Assist in decision making
  - Get trained Part of the task is simple (locate the information)
- But true success (the effect of the info on the user) is not easy to measure

Testing Methods

Non-User Based Testing

Compliance Testing

- The Spelling and Grammar checker of usability testing
- Possible (within limits) to be performed by anyone
- Can remove the low level usability issues that often mask more significant usability issues

Compliance Testing (concluded)

- Style Guide-based Testing
  - Checklists
  - Interpretation Issues
  - Scope Limitations
- Available Standards
  - Commercial GUI & Web Standards and Style Guides
  - Domain Specific GUI & Web Standards and Style Guides
  - Internal Standards and Style Guides
- Interface Specification Testing*

*Special Case of QC Testing that assumes a usable design to start with

Expert Reviews

Heuristic Evaluation Guidelines

- One or more usability experts review a product, application, etc.
- Subjective but based on sound usability and design principles
- Highly dependent on the qualifications of the reviewer(s)

Expert Reviews

Nikon Ad - circa 1990

Heuristic Evaluation Guidelines

- Hix and Harrison Design Guidelines (27)
- Bruce Tognazzini's First Principles of Interaction Design (15)
- Nielson's Heuristics (10)
- Ben Shneiderman's 8 Golden Rules
- Don Norman's 6 Principles of Usability
- Bill Killam's 4 Main Heuristics

1rd Heuristic

Design for the intended users (not yourself). Use their language. Use concepts they understand. Design for their capabilities and limitations.


CD-RW

X	O	X
	X	
O		

The Game of 15s

- Let's play the game of "15." The pieces of the game are the numbers 1, 2, 3, 4, 5, 6, 7, 8, and 9. Each player takes a digit in turn. Once a digit is taken, the other player cannot use it. The first player to get three digits that sum to 15 wins.

- Here's a sample game: Player A takes 8. Player B takes 2. Then A takes 4, and B takes 3. A takes 5. What digit should B take?

The Game of 15s (continued)

4 3 8
9 5 1
2 7 6
4 5 6
8 5 2
4 9 2
3 5 7
8 1 6

- Player A takes 8. Player B takes 2. Then A takes 4, and B takes 3. A takes 5. What digit should B take?

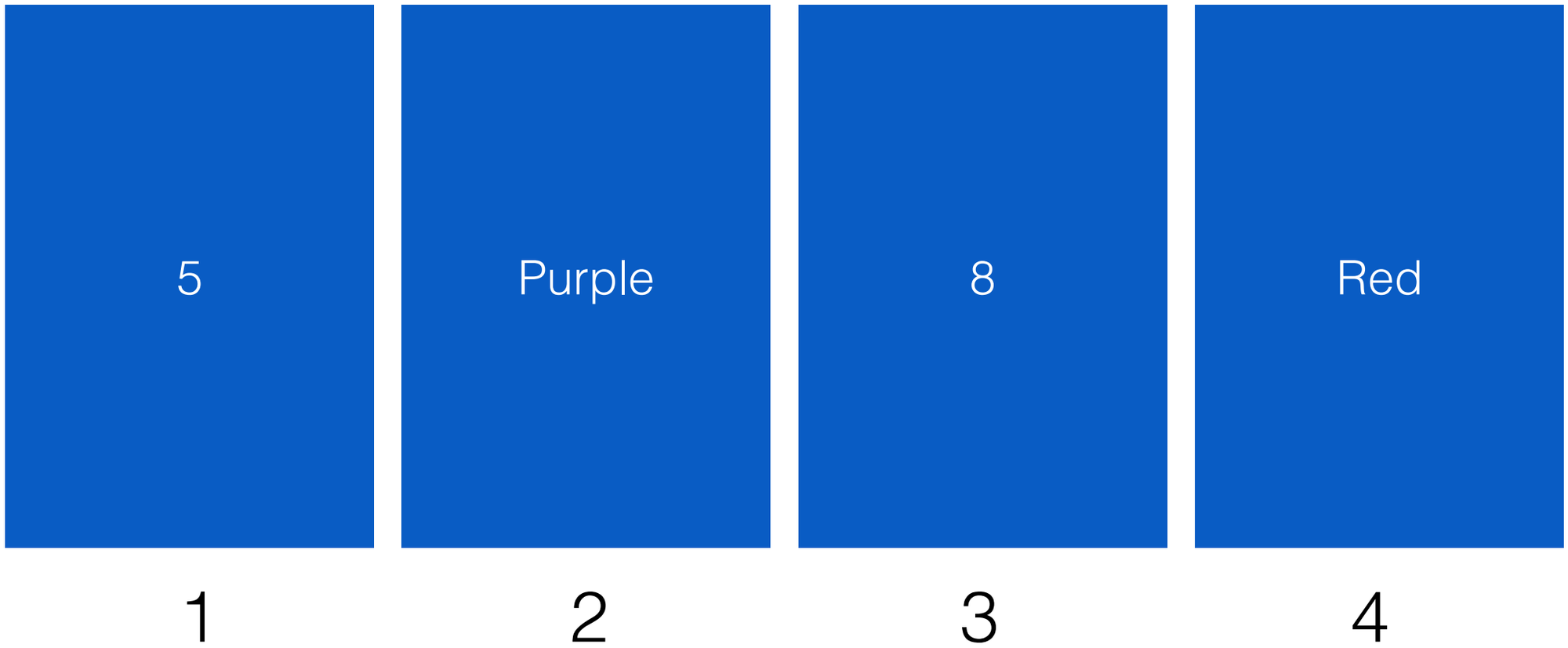
The Game of 15s (concluded)

~~4~~	3	~~8~~
9	~~5~~	1
2	7	6

Framing Problem to Match our Abilities

- The rule to test: If a card has an even number on its face, its has a primary color on its opposite face.
- How many cards in the next slide do you need to look at to confirm this rule is being followed?

Framing Problem to Match our Abilities



Framing Problem to Match our Abilities

- The rule to test: You cannot drink alcohol if you are under 18.
- How many cards in the next slide do you need to look at to confirm this rule is being followed?

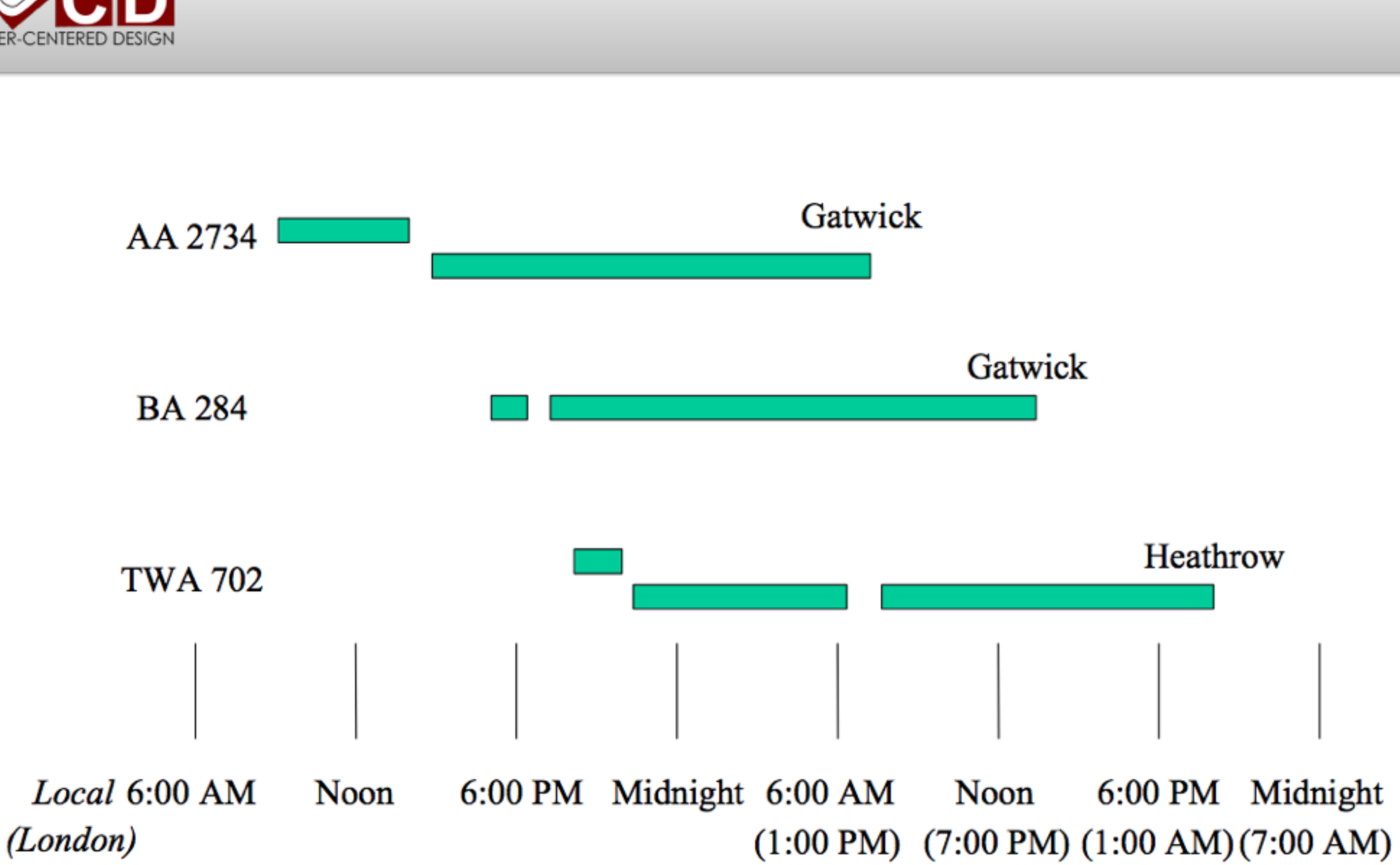
Framing Problem to Match our Abilities

The Results from Research

- The correct answer is that you need only check 2 cards - the second and thirds cards in both problems (8 and purple or Tequila and 16).
- Less then 1/4 of people solve this problem when framed in colors and numbers, but virtually everyone gets the problem correct when framed as ages and drinks.

1131	SAN	0820+1	LGW	AA	2734	FCYBM	D10	1
	AA	2734	CHG PLANE AT DFW					
X12	1805	SAN	1425+1	LGW	BA	284	FJMSB	D10	1
	2100	SAN	2030+1	LHR	TW	702	FCYBQ	*	2
	TW	702	EQUIPMENT 767 LAX L-10					

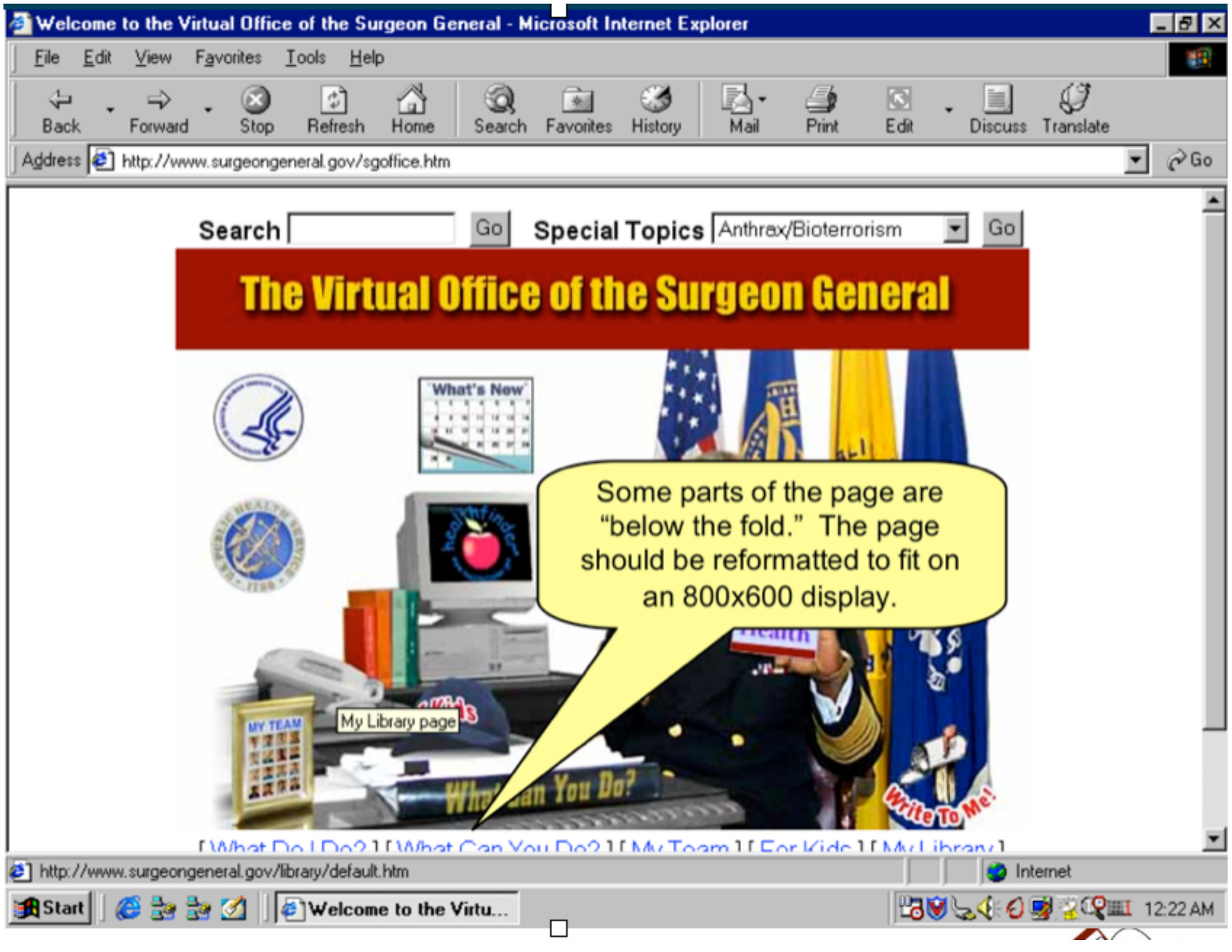
AA 2734
Gatwick
BA 284
Gatwick


TWA 702
Heathrow
Local 6:00 AM
(London)
Noon
6:00 PM
Midnight
6:00 AM
(1:00 PM)
Noon
(7:00 PM)
6:00 PM
(1:00 AM)
Midnight
(7:00 AM)

2nd Heuristic

Make it clear where and how to interact. Functional “visibility” occurs when there are obvious places to and procedures for interaction. Learning comes from adequate feedback.

J.K.Rowling
official site


Welcome to the Virtual Office of the Surgeon General - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Forward Stop Refresh Home Search Favorites History Mail Print Edit Discuss Translate
Address http://www.surgeongeneral.gov/sgoffice.htm
Go
Search Go
Special Topics Anthrax/Bioterrorism Go
The Virtual Office of the Surgeon General
What's New
Some parts of the page are "below the fold." The page should be reformatted to fit on an 800x600 display.
My Library page
MY TEAM
Write To Me!
[What Do I Do?] [What Can You Do?] [My Team] [For Kids] [My Library]
http://www.surgeongeneral.gov/library/default.htm
Internet
Start
Welcome to the Virtu...
12:22 AM

3rd Heuristic

Support the user's development of a good, complete, and unambiguous cognitive model of the product.


Recycle Bin Icon


FREE 2-Day Shipping, No Minimum Purchase
Gift Cards
Your Account | Help
GO
Cart
Your Lists
Amazon Daily BLOG
7 posts since yesterday
Posts for Bill
New iPod shuffles Announced
The new iPod shuffles, available in black or silver, store up to
Arrived
onger battery

ORHAM'S

CONTRACTORS DEPOT

Item Number

Go!

Advanced Search

Search for goods

David Burt / *dgburt*

	Criteria		Value	
Search:	Item Description	Like	cement	Go!

My Jobs
Sub Jobs
Select Goods
Orders
Quotes
Departments
Reports
Change Password
Logout

Search Results

001336 PW-45 HEATER REPLACEMENT WICK

Retail:$ 10.79
Specification: PW-45 HEATER REPLACEMENT WICK
PW-45 HEATER REPLACEMENT WICK
Substitute Items Related Items

015935 PW-12 HEATER REPLACEMENT WICK

Retail:$ 9.19
Specification: PW-12 HEATER REPLACEMENT WICK
PW-12 HEATER REPLACEMENT WICK
Substitute Items Related Items

015852 PW-68 HEATER REPLACEMENT WICK

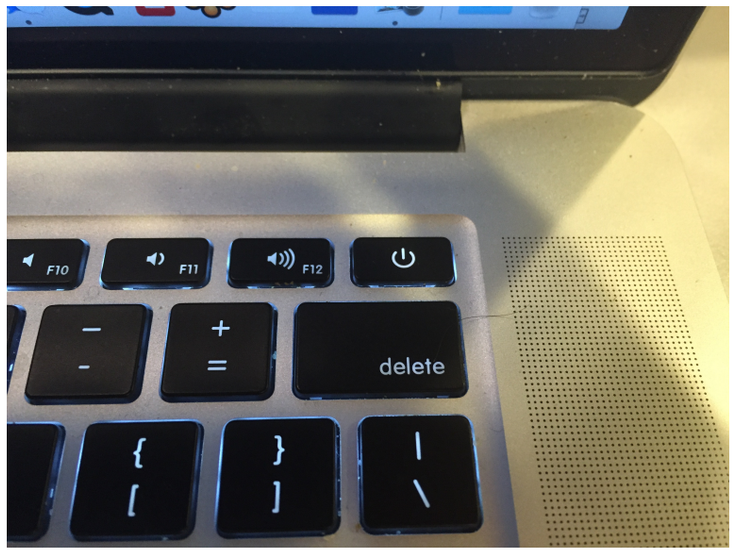
Retail:$ 13.59
Specification: PW-68 HEATER REPLACEMENT WICK
PW-68 HEATER REPLACEMENT WICK
Substitute Items Related Items

4th Heuristic

Expect, eliminate, limit the impact of, or compensate for errors

Error Type 1: Slips

- Slips are common users issues
- Hand/eye coordination or basic control of our psychomotor systems
- Exacerbated by distraction, speed, attention overload
- Unavoidable by design but need to be anticipated and addressed by the designer
- “To err is human. To forgive: Design”



Error Type 2: Lapse

- Lapses are induced by inconsistencies or lack of good ease of recall
- Example: NIST's need to "chop" a form

Error Type 3: Mistake

- Mistakes are generated by a lack of understanding or a lack of sufficient or correct information
- Lack of sufficient or correct information is the responsibility of the designer in the presentation layer of an interface
- Vernally show us as errors of omission, commission, systematic errors.
- Lack of understanding is the responsibility of the designer in interaction and in conceptual model of an interface
- Mistake are often undetectable by the end user

USER-CENTERED DESIGN

bkillam@user-centereddesign.com's Shopping Bag. (If you are not bkillam@user-centereddesign.com, click here.)

Unisex Hooded Work Jacket EDIT

Color: True Navy

Item Total
$98.50

REGULAR # 458834CT3

Size: L	QTY 1	$90.00	REMOVE

Logo 1: View Logo $8.50 REMOVE

0289831
Location: Chest Right Side

Unisex Hooded Work Jacket EDIT

Color: Black

Item Total
$95.95

REGULAR # 458834CT3

Size: L	QTY 1	$90.00	REMOVE

Logo 1: View Logo $5.95 REMOVE

0289831
Location: Chest Right Side

USER-CENTERED DESIGN

Manage a Bank Account

Back to Make a Payment

Note: Select the agreement check box to confirm that you are the owner of the account entered

***required field**

Account Type*: Checking (•) Savings ( )

Bank Account Name:

Routing #*:

Account #*:

Confirm Account #*:

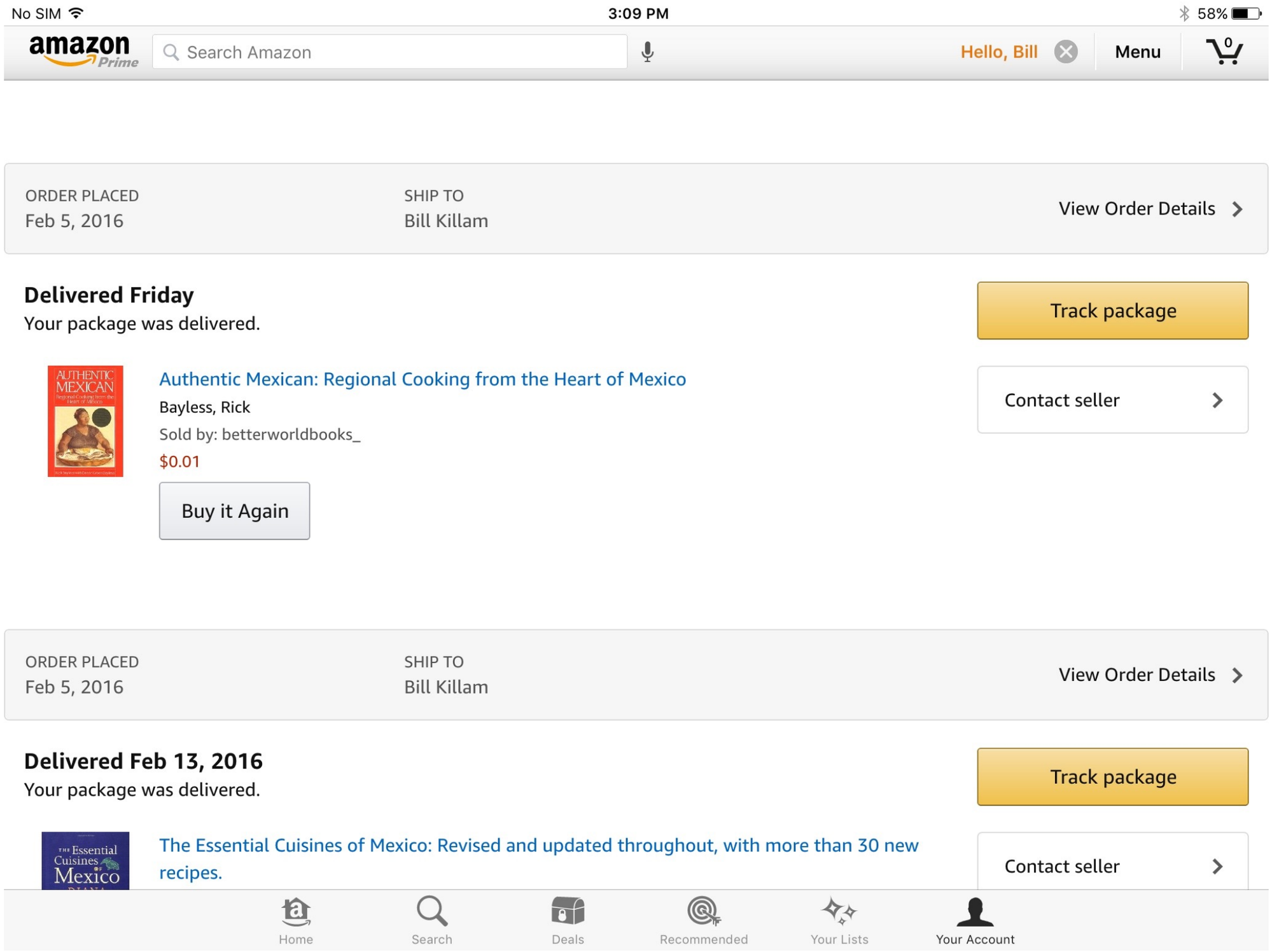
Name as it Appears on Check*: (25 alphabet characters)

|Routing #| |Account #|

☐ I acknowledge I am an owner of the account referenced above and I am authorized to make payments from that account.

Submit Delete Back

To edit or delete a payment account, click on the Account Name (ONLY One account can be added)


No SIM
3:09 PM
58%
amazon Prime
Search Amazon
Hello, Bill
Menu
0
ORDER PLACED
Feb 5, 2016
SHIP TO
Bill Killam
View Order Details
Delivered Friday
Your package was delivered.
Track package
Authentic Mexican: Regional Cooking from the Heart of Mexico
Bayless, Rick
Sold by: betterworldbooks_
$0.01
Buy it Again
Contact seller
ORDER PLACED
Feb 5, 2016
SHIP TO
Bill Killam
View Order Details
Delivered Feb 13, 2016
Your package was delivered.
Track package
The Essential Cuisines of Mexico: Revised and updated throughout, with more than 30 new recipes.
Contact seller
Home
Search
Deals
Recommended
Your Lists
Your Account

User Based Testing

Psychological Issues in Testing

Known and well documented biases

- The availability biases
  - Ease of Recall
  - Retrievability
  - Presumed Associations
- Representative Biases
  - Insensitivity to Base Rates
  - Insensitivity to Sample Size
  - Misconceptions of Chance
  - Regression to the Mean
  - The Conjunction Fallacy/ Plausibility Bias
- The Confirmation Bias
  - Outcome Bias
  - Hindsight Bias
  - Correlation equal Causation Bias
- Anchor & Adjustment Bias
- The Heuristic Effect
- Attribute Substitution
- Reality First

Anchoring Effect

- The effect of anchoring, known is the anchoring index, is approximately 50%. In other words, the anchoring effect accounts for 50% of the error between the point that would be selected if no anchoring had been provided and the point that is provided with the anchor. This is independent of domain knowledge, which makes it all the more insidious.

- When real estate agents were tested and asked to determine the actual selling price of the home after being told the asking price, but provided with both a low and high anchor point, the effect of anchoring was measured at 41%. These professionals claimed that they were not effected by the asking price. When the same study was conducted with college students with no background in home sales, the effect was measured to be 48%. However, this group were aware they were effects.

- People who are instructed to shake their head when they hear an anchor point tend to move farther away from the anchor point in negotiations then when people are asked to nod there head when they hear the same anchor point.

Other Unconscious Influence Examples

- 2013 Harvard study showed that people were more assertive after using a larger device (an iPad) then after using a smaller device (an iPhone).
- J.Lo. and JZ's relationship was thought more likely to end when people were sitting on a wobbly chair
- People holding cold drinks rate people as more aloof, but people are seen as more friendly when holding a warm drink

Priming

FINISHED FILES ARE THE RE-
SULT OF YEARS OF SCIENTIF-
IC STUDY COMBINED WITH THE
EXPERIENCE OF MANY YEARS

The Reality First Bias

- When asked questions, we integrate our current state into our answers, even if its unrelated.
- If you ask a person who just worked out on a treadmill if they would be more hungry or more thirsty when lost in the woods, they are more likely to report thirsty (91%) versus people who did not just work out on a treadmill (60%).
- If asked about our lives when it's raining, people report lives that are worse then when the weather is nice.
- If the last event was positive, people tend to rate the overall experience as positive. If the last event was negative, people tend to rate the overall experience as negative.

Affect Heuristic

- Researchers had participants enter into a room for a study. In some cases, Monopoly money was present in the room or a money related screensaver was on the computer monitor. At some point during the presumed task, the researcher knocked over a jar of pencils. All participants assisted in picking up the pencils; however, those participants that “saw” money in the room picked up, on average, half as many pencils.
- In a study, participants were asked to create 4 word sentences from 5 word sets. For some of the participants, foreword sets included words related to aging (bald, Florida, wrinkled, etc.). After the 1st task, participants were asked to walk down the hall to a 2nd room. Participants who had seen the age-related words in the 1st task walked more slowly the 2nd room.
- In a study, half of the participants experienced a flyover simulation of a city as though from an helicopter. Half of the participants experienced a flyover simulation of the city as though they were flying like Superman. At some point during the presumed task, the researcher knocked over a jar of pencils. The “Superman group” picked up more pencils and started helping sooner than the group that flew over in a helicopter.

Answer Substitution

- “You are shown a picture of a person running for office and asked if you think they will win. There are far to many variables for you to make a good prediction, so the task is too hard for system 2 to work out.
- System 1 substitutes the hard question for an easier one – does the person look like a person who will win?
- System 1 provides an answer to that new question, but System 2 reports it as the answer to the first question without realizing the substitution.

Confabulation

- If System 2 does not have access to the information of system 1, it will use logic to answer the question even if its incorrect.
- In a split brain study, people were shown a picture of a chicken's leg and a picture of a car covered in snow and then asked to point to a related picture in a set. People pointed to either a picture of a chicken or a picture of a snow shovel. If the image was shown to the left hemisphere, they could describe the reason why they pointed to this picture. If the picture was shown to the right hemisphere, they pointed to the picture but could not explain why.
- When participants were shown the picture of a chicken's leg to the left hemisphere and a picture of the car in snow to the right hemisphere at the same time, they would point to the same 2 pictures. When asked why they pointed to the picture of the shovel, participants reported that chickens produce a lot of chicken poop, so you need a shovel to clean it up.

Projected Responding

- Respondents believe they understand the goal of the project and attempt to provide the information they think is being asked for. Can be induced by the experimenter (the experimenter expectancy effect).
- Almost unavoidable.
- Encouraged by subtle differences in responses (correctly or incorrectly) perceived. Why its extremely hard to test your own designs. And why you NEVER take notes in the participants's presence.

Observer Effect/Social Facilitation

- Any of a general class of changes to a user's behavior as the results of being observed (or thinking they are being observed)
- The effect of observers is powerful and unconscious. And they don't even need to be real observers. In research on stealing and lying, children are less likely to cheat and lie there is a mirror in the room. Melissa Bateson ran a field experiment with her own (psychology) department. Coffee was paid for on a faith basis. She alternated images above the donation box - even weeks had a poster with flowers on it, odd weeks had a poster with eyes on it. On odd weeks, contributions were 3x what was received on even weeks.
- The Heisenberg Principle of usability testing.
- Most well know is the Hawthorne Effect.
- This effect causes a divergence in performance - the good do better, the poor do worse.

Anxiety


CavemanCircus.com
Which of the following is the largest?
A: A Peanut
B: An Elephant
C: The Moon
D: A Kettle

Confirmation Bias

- The tendency for the observer to see information that fits their expectation, whether conscious or unconscious.
- Almost unavoidable, but certainly unavoidable if you are responsible for the design.

Think Aloud Testing

Think Aloud Protocol

- Participants are asked to verbalize their thought process and expectations while working on a product
- Training is required
- Prompting is required to maintain the process
- Most widely used (which is not a good thing)

Think Aloud Protocol (concluded)

- Disruptive to performance
- Issues of the ability for users to be introspective
- Issues of distraction (split attention)
- Issues of verbal overshadowing
- Issues of increased anxiety
- Focuses the participant on the very aspects of the design we don’t want them to be consciously aware of
- Suitable for concept presentation and cognitive walkthroughs on non-operational products (e.g., story boards, static screen flows, Wizard of Oz walkthroughs)

Interrupted Task-based Testing

Interrupted Task-Based Testing

- An approach that allows for exploration of issues without being overly disruptive when issues are not present
- Can be used for exploratory testing on an existing design
- Can be used for exploring possible design alternatives
- Should (must) follow the ethical guidelines for the treatment of human subjects (including informed consent), confidentiality
- Should not be hampered by trying to support statistical analysis

Other Formats

Other Data/Formats

- Eye Tracking, Physiological Measures, Blink Rates, etc.
  - Objective measures that seem more realistic
  - Lack perceptual component (e.g., with eye tracking what we look directly at is not all we see, we can look directly at something and not see it, and what we perceive is not always what is in front of us)
- Remote Usability Testing
  - Has logistical advantages
  - Generates a false assumption that its more valid
  - Doable as a think aloud, but otherwise results in a hybrid (part interrupted task based and part think aloud)
  - Much of the observational data is missing
- Co-Discovery
  - 2 peoples working on a problem together
  - A highly useful hybrid approach (natural task performance and think aloud)

Interrupted Task Based Testing Procedures

Test Set-up

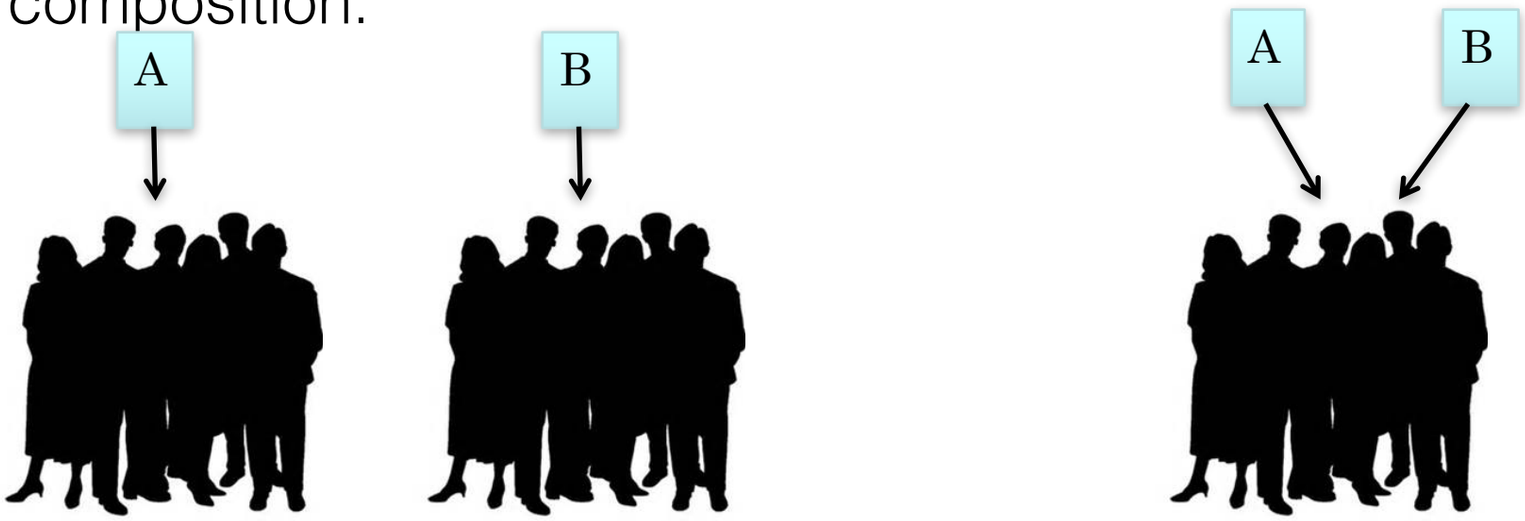
- What's the hypothesis?
  - A hypothesis is required for research
  - A hypothesis not "required" for usability testing but its present anyway - essentially that design one design is better than another
- Define Your Variables
  - Independent Variables - what you, the "experimenter" chooses to subject each group to (the designs)
  - Dependent variables - what effect you think will be "dependent" on their use of the product (speed, frustration, joy, anger, hesitation, confusion, task success, cognitive workload, etc.)
  - Constants - variables you need to control to keep the evaluation fair. Examples might be level of experience, background, domain knowledge, computer skill, etc.
  - Random Variables - those thins tour don't try to control since yo think they don't matter to the evaluation
- Confounding Variables
  - If, after the evaluation, you discover something that might the the cause of differences observed in the evaluation this is a confounding variable
  - Confounding variables typically occur when there is a failure to set a constant (e.g., you account for experience with he product but not a similar product type)

Participant Issues

- User-types
  - User versus user surrogates
  - All profiles or specific user profiles/personas?
  - Other audiences
- How many?
  - Relationship to statistical significance
  - “Discount Usability” – who’s rule?
  - No less then 3 from any group
- Recruiting
  - Friends and Family or Snowball Recruiting
  - Recruiting Firms
  - Over recruiting
- Participant stipends
- Scheduling

Within versus Between Subject Designs

- A between subject design uses a different group to evaluate each product. This type of design can be extended to as many products as needed.
- A within subject design has everyone evaluate the products so its only practical when testing two designs. But, to be as fair as possible, the order of the evaluation has to be “counterbalanced” - 1/2 of the participants try out design A first, the other half try out design B first.
- The choice of design is based on time commitment & number of designs/ products. Within subject design needs fewer people and they can compare the design themselves, but they can spend less time with each product. Between subject design allow for more time with a design and avoids learning effects, but requires two (or more) group that should be identical in composition.



Defining Task Scenarios

- “Scenarios are contrived for testing, may not be representative of real world usage patterns, and are NOT always required
- Short, unambiguous tasks to explore areas of concern, redesign, or of interest
- Wording is critical
  - In the user’s own terms
  - Does not contain “seeds” to the correct solution
- Enough to form a complete test but able to stay within the time limit
  - Flexibility is key
  - Variations ARE allowed

Preparing Test Materials

- Consent form
  - Confidentiality
  - Anonymity
  - Time, intent, expectation from them
  - Voluntary and the right to stop
- Video release form
- Receipt and confidentiality agreement
- Facilitator's Guide
  - Introductory comments
  - Participant task descriptions
  - Questionnaires, SUS, Cooper-Harper, etc.

Piloting the Study

- Getting subjects
  - Convenience sampled
  - “Hallway” testing
- Collect data
  - Check task wording
  - Check timing

Facilitating

- Rogerian principles apply
  - Unconditional Positive Regard
  - Empathy
  - Congruence
- Rogerian techniques are used
  - Open ended questions
  - Minimal encouragers
  - Reflections
  - Summarization
- Objectiveness - never plan to test your own design

Collecting Data

- The data is NOT in the interface, the data is in the user!
- Data is observational, not transcribable
  - Behavior, reactions, hesitations (movement and voice), body language, “tells”
  - Collecting participant comments may be misleading (e.g, confabulation), but may help indicate when issues are present
- Collecting subjective data
  - Pre-test
  - Post-task
  - Post-test

Writing Tasks

Need to Understand the Product

- You can't test everything
- Decide if you are testing System 1, System 2, or both
- Perform an expert review first
- Determine what's new, what's changed or what's critical that needs to be tested
- Determine what tasks allow you to assess potential issues while performing these tasks
- Have enough tasks to change out those you "finish"

Tasks Descriptions

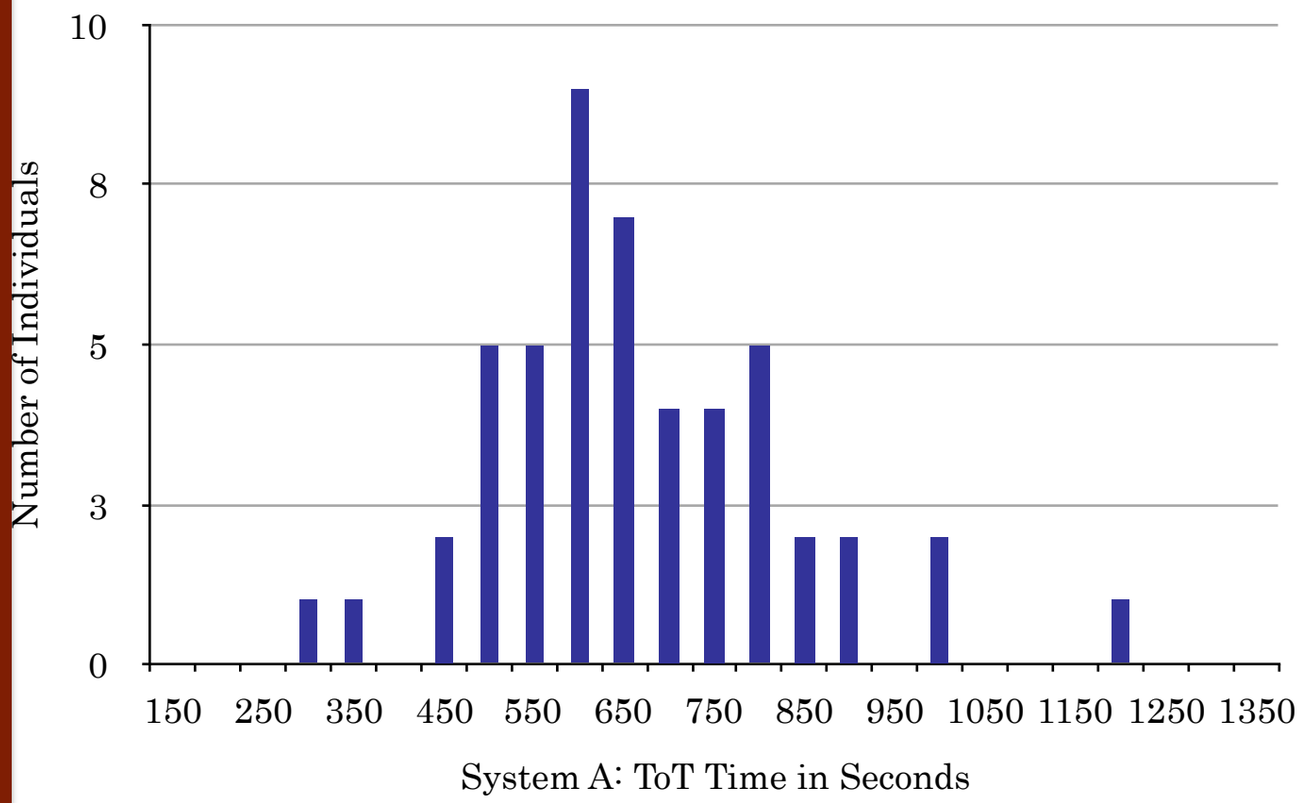
- Write short, unambiguous, tasks that expose the user to the specific function or displays to be evaluated
  - Don't Use: "You are with a friend studying for final and getting tired. You decide its time for a cup of coffee but you don't know where the nearest coffer shop is located. Using this app, try to find the nearest coffee shop to your current location."
  - Use: "Find the nearest coffee shop"
- Don't seed the answer
  - Don't Use "Figure out how to register for the site" if there is a link called "Register"
- tell them what you ant them to do, not how to do it.
  - Don't use: "Using the advanced search feature…"

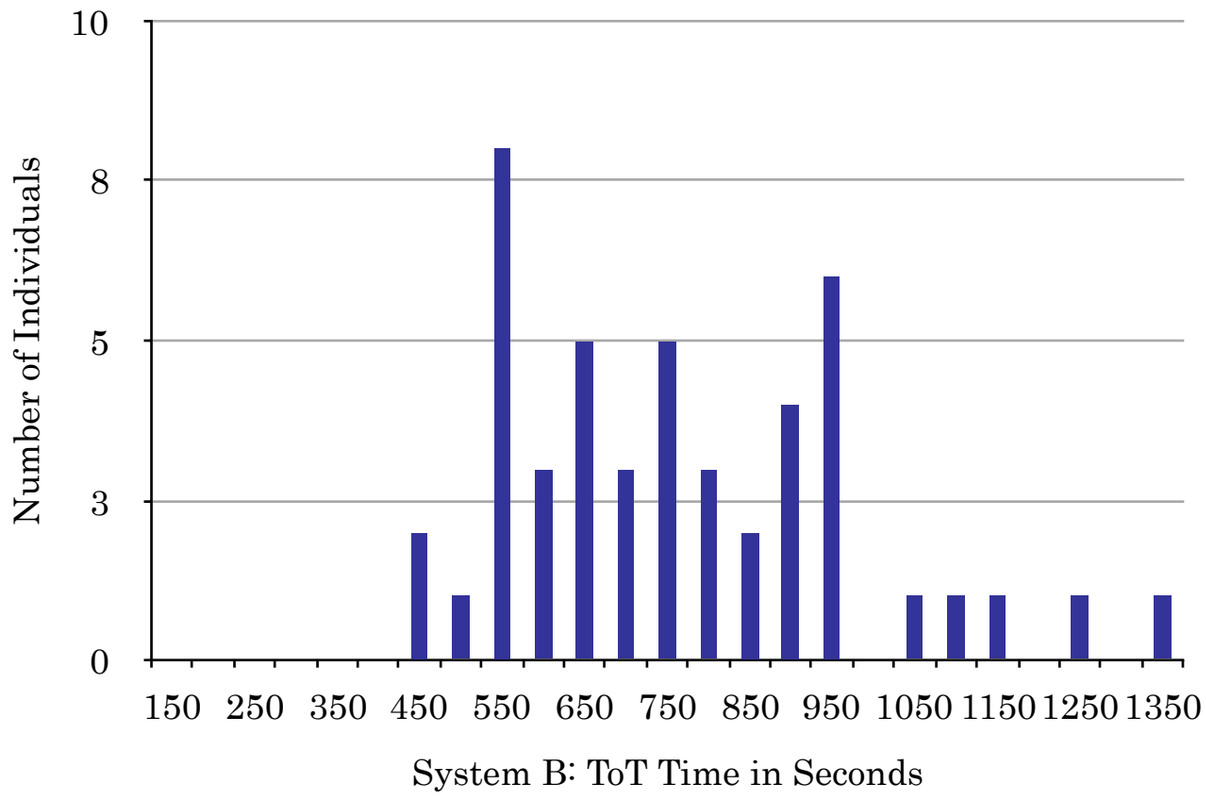
Reporting Results

Efficiency Data

- Efficiency
  - Can be operationalized in number of ways
  - Time on task being the most common, but keystones and screens have been proposed
  - These can be measured objectively, but they do not have construct validity
- External time may be important to management and some types of engineering (particularly process flow)
- External time is important when an external time limit is imposed on task performance (process flow, missile intercept, etc.)
- These measures do not correlate with effectiveness except in extreme cases. Mental effort is a better measure of efficiency.

Sample ToT Data – Controlled Experiment*





*Source: UCD, Inc. – Voting System Usability Compliance Test Development Report for NIST

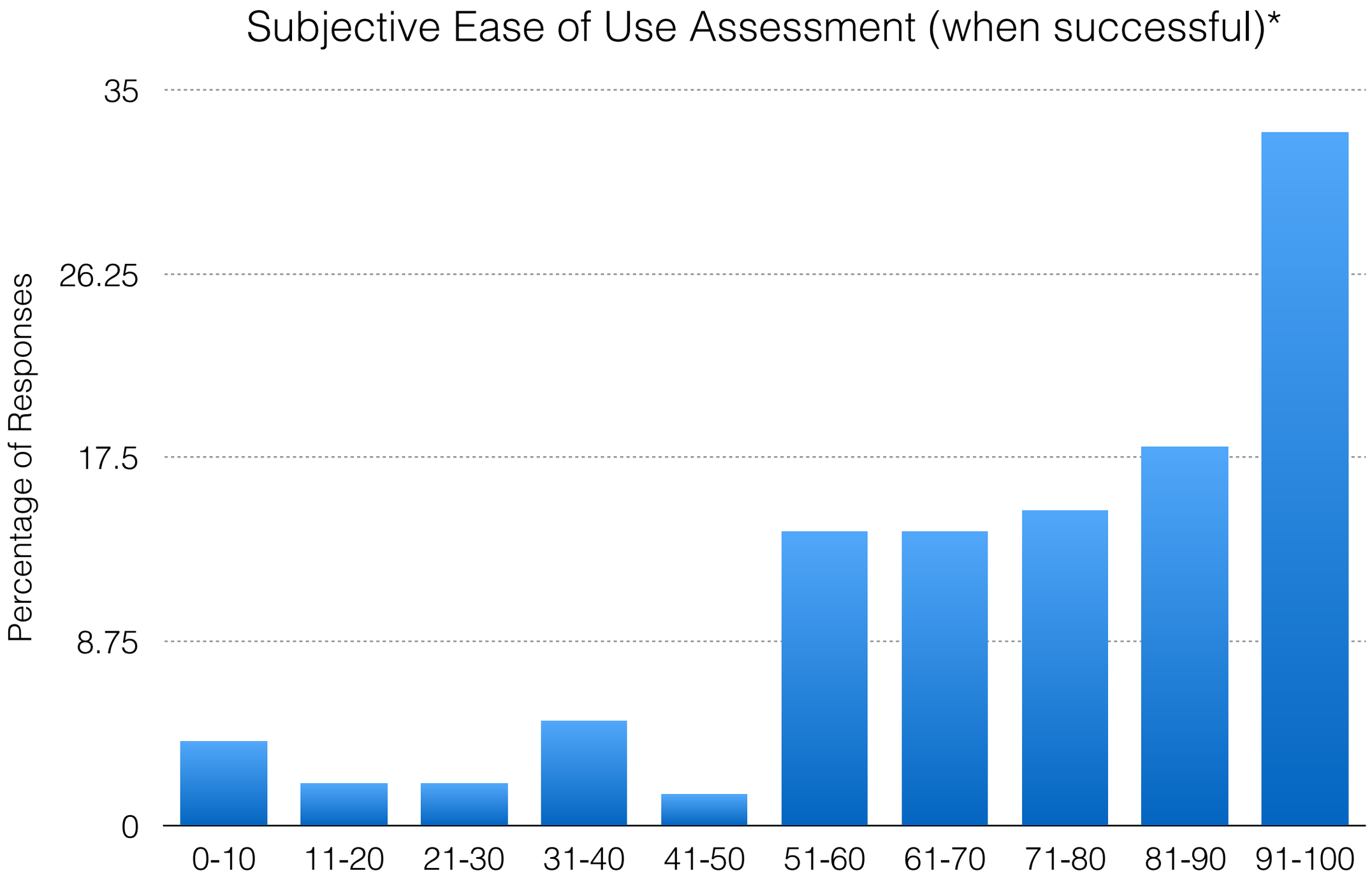
Satisfaction Data

- Satisfaction data can be operationalized in a number of ways, but is always opinion data (and non-parametric)
  - Standardized survey instrument (e.g. SUS, SUMI, QUIS)
  - Simple Likert item and Likert scale assessments
- Satisfaction data suffers from numerous issues that threaten their validity
  - Halo effect
  - Leniency bias
  - Strictness bias
  - Projected responding
  - Usability Issues–a lack of agreed understanding between the question(er)and the respondent)
- Satisfaction data does not correlate with performance

Satisfaction & Perceived Usability

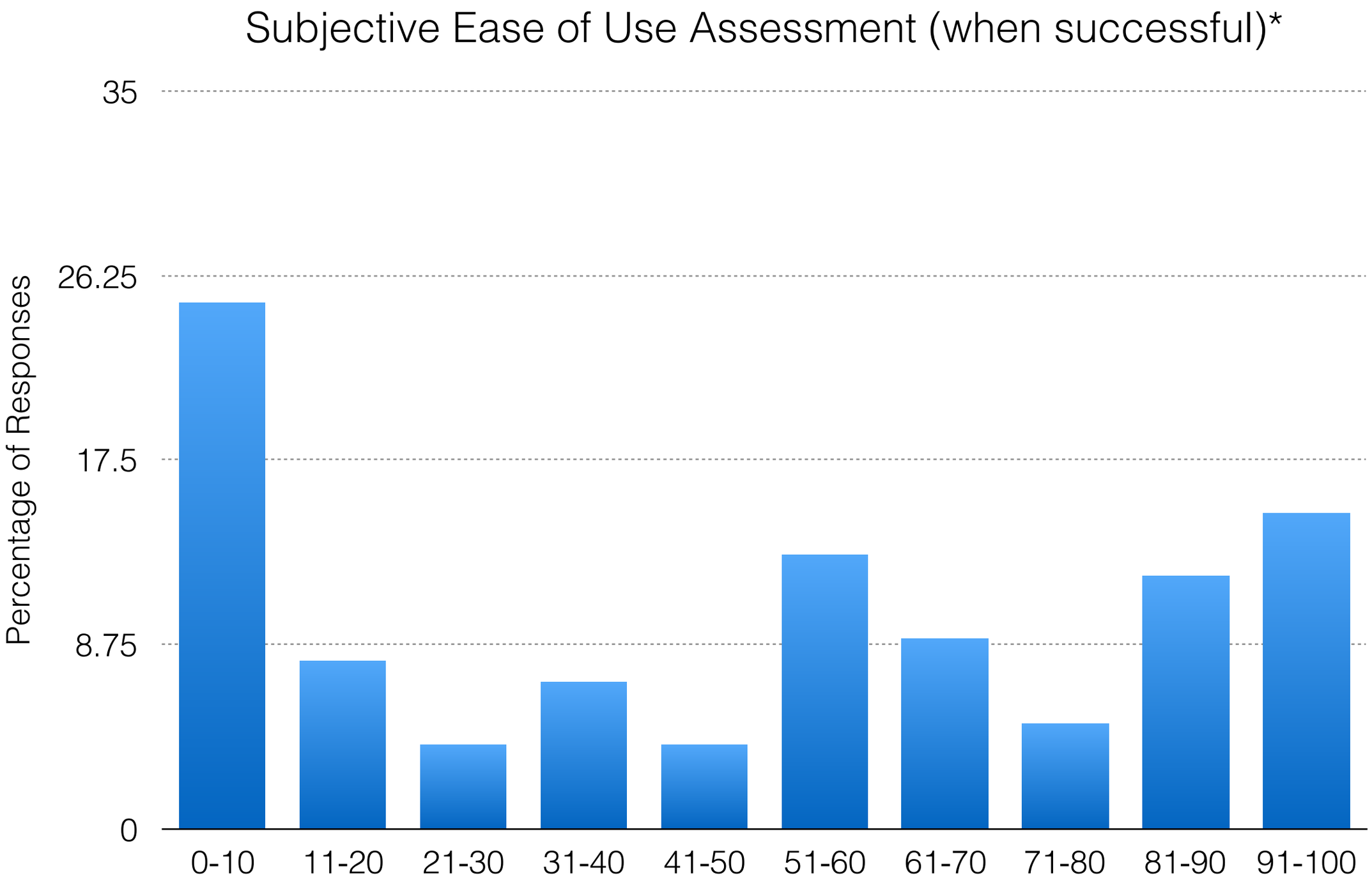
- Satisfaction and perceived usability are, however, fairly stable and often independent of performance or even efficiency data
- A simple, one question evaluation can be used to assess satisfaction and perceived usability or a tool like the SUS could be used for assessment.
- If users have prior experience with the existing design, their subjective assessment can be made against the current design.
- But opinion data is non parametric. You cannot average it or compare it directly. A Wilcoxcon Signed Rank Sum Analysis, Mann Whitney U test, or other non parametric evaluation is needed to test for statistical significance.

Post Test Analysis of Approx. 3000 Sessions*



*Source: Jeff Sauro, Measuring Usability

Post Test Analysis of Approx. 3000 Sessions*



*Source: Jeff Sauro, Measuring Usability

Satisfaction

- Approximately 1600 people participated in evaluations of voting systems, comparing paper boots to electronic systems (DREs). All participants attempted to cast identical votes.
- Statistically significant performance differences were detected between the two systems with neither system performing at 100% accuracy. (Values indicated as many as 1/3 of all participants had one or more errors on one of the systems.)
- When asked how they performed, only 1 reported they are not sure. The rest reported they had no issues and had cast all votes perfectly.
- When asked if they liked the system and would use it again, 100% reporter they would.

Opinion Data

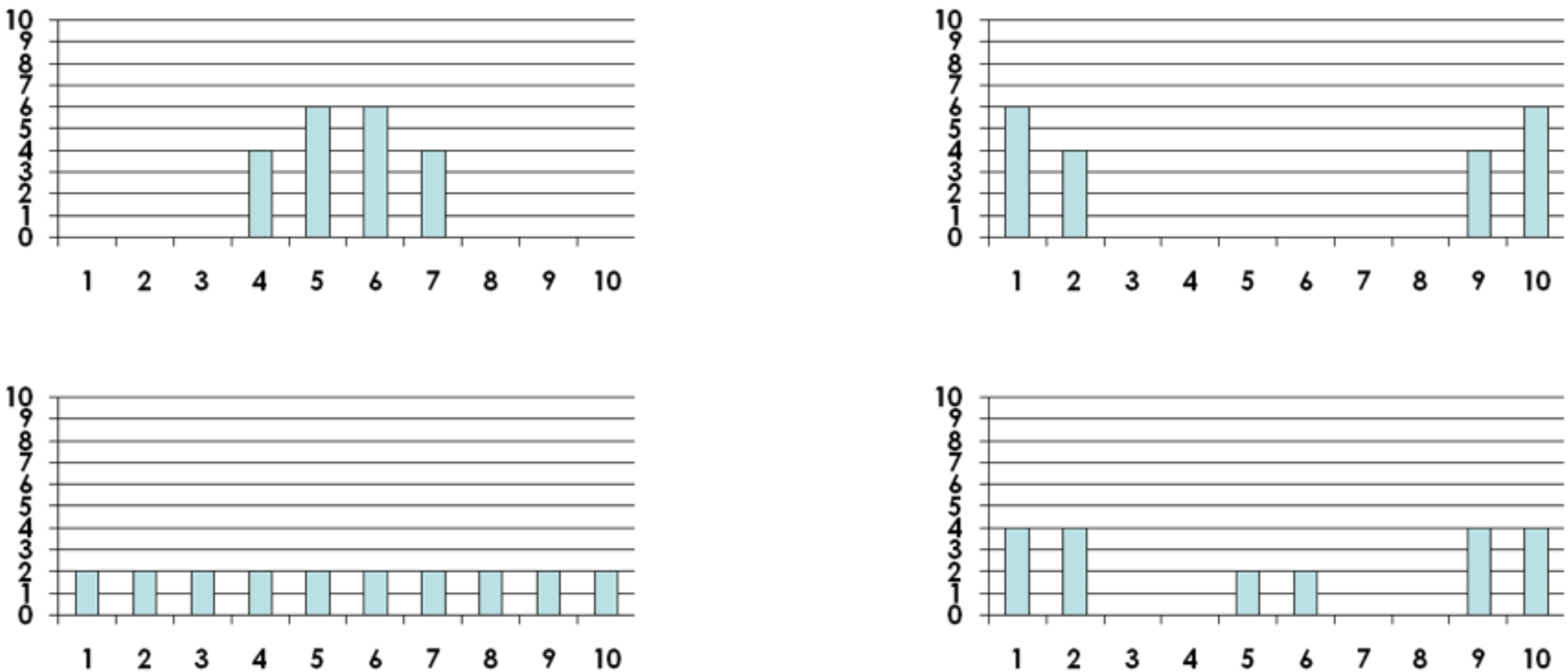
- People often collect data from users on satisfaction or their assessment of the product usability (perceived usability).
- Opinion data is non parametric. It does not follow a normal, or bell shaped curve so arithmetic averages, standard deviation, etc. are not applicable.
- Even if a scale is provided (e.g., rank the product on a scale of 1 to 10), the results cannot be treated as interval or ratio data that follows the rules of addition, subtraction, multiplication, or division (e.g., one person's score of 8 cannot be assumed to be 40% better then another person's score of 6).
- Opinion data should be analyzed as a binomial (yes/no, prefer/don't prefer, like/don't like, etc.). Then a binomial formula (similar to that used for pass/fail), can be used to determine the confidence interval of the data
- Alternately, a Likert question can be used (e.g., Statement: This design is better then the old design: Response option: Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree).Then a Wilconxin Rank Sum Test can be used to determine if there is a preference. The Wilcoxin Rank Sum Test will provide a statistical probability value.

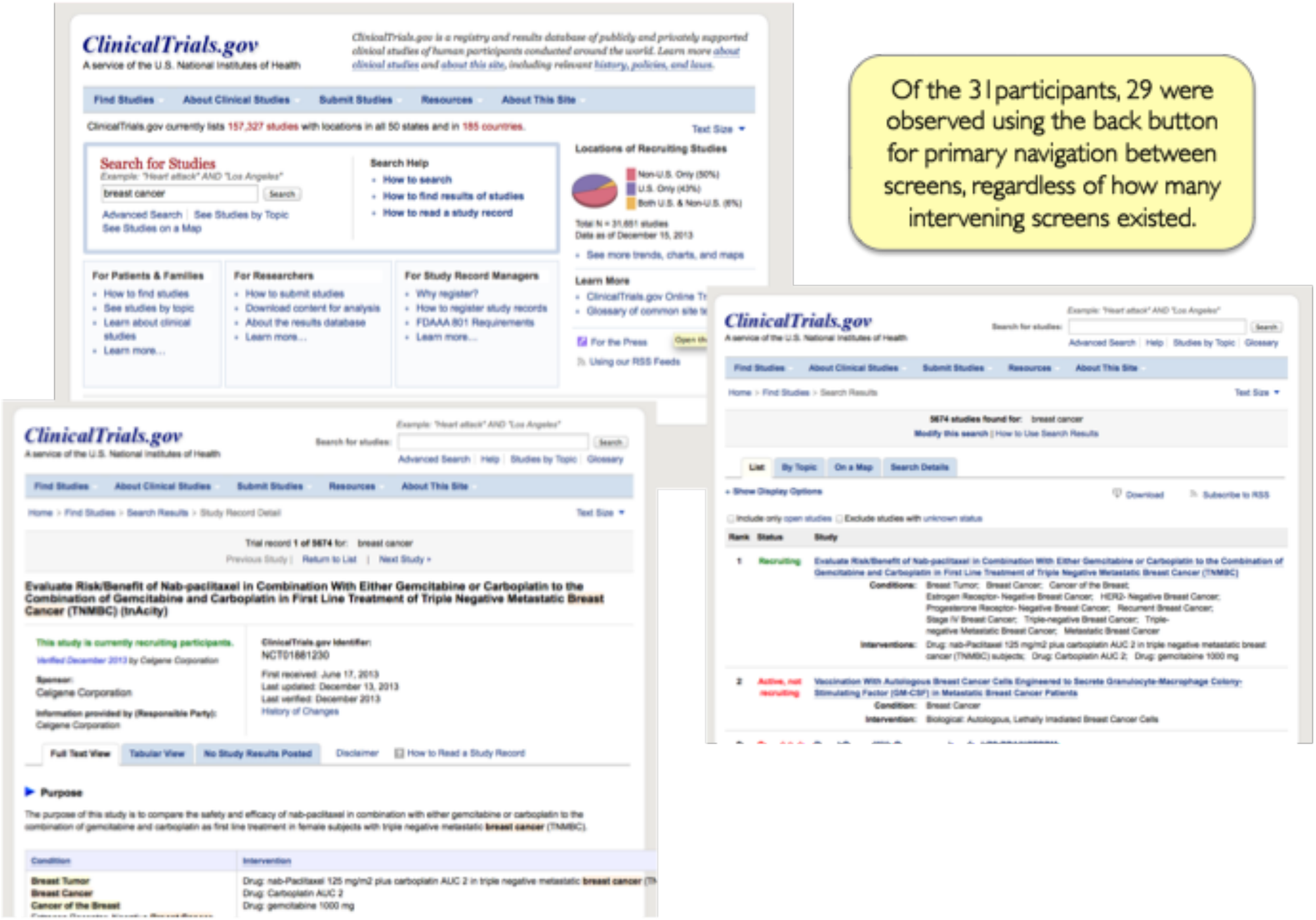
So What is Reported?

- User behaviors, with explanations of the likely cause
- Violation of industry standards and best practices
- Predictions based on prior user-based testing
- Also report, with care
  - User Comments
  - Data form standardized surveys (e.g., SUS) or other scales

Histograms

- Histograms can can tell a lot about the data
- But the data often shows other patterns such as bimodal distributions. In these cases, the average and standard deviation are not adequate…

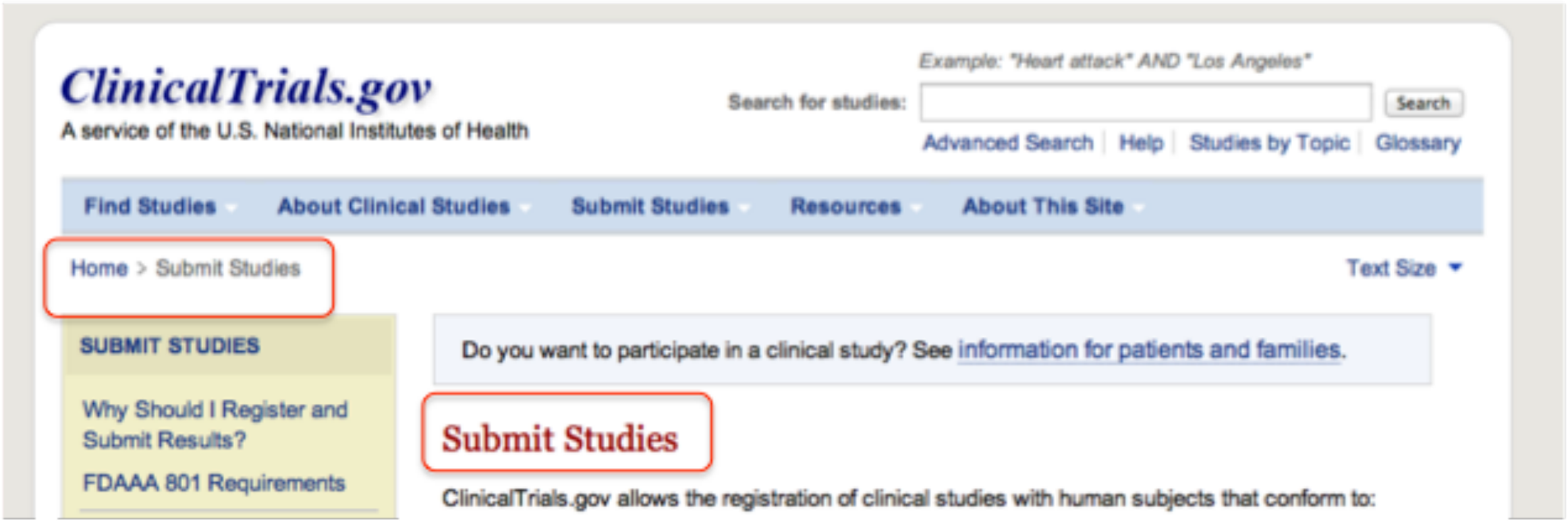




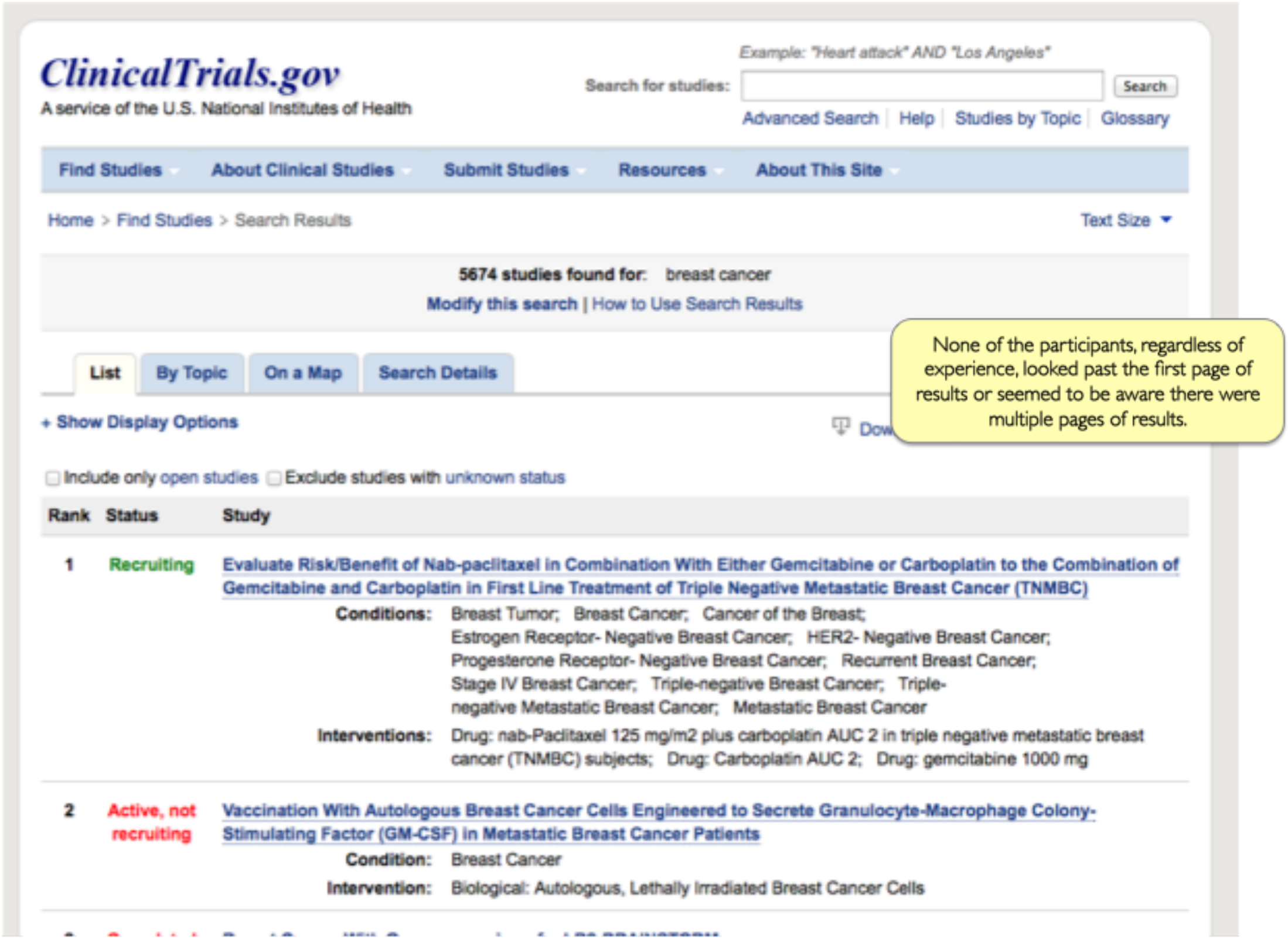
Of the 31 participants, 29 were observed using the back button for primary navigation between screens, regardless of how many intervening screens existed.

Only one of the 31 participants (an experienced user) used the breadcrumbs.

The breadcrumbs are not visually associated with the page title. In some cases, there are intervening visual elements that may preclude the user from noticing the necessary association.



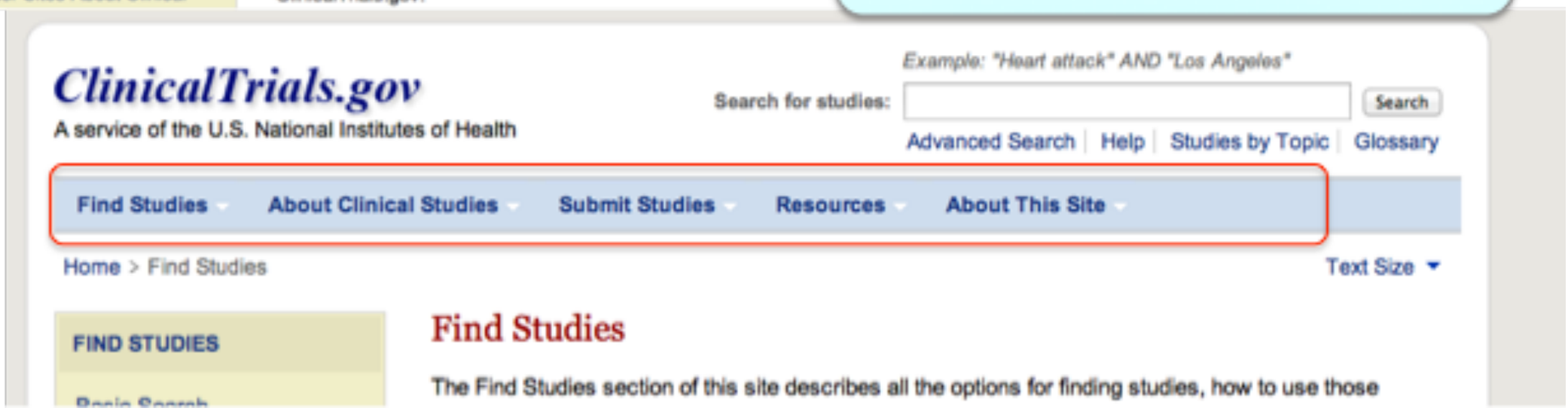
Only Researchers and Journal Editors understood the purpose of this section and did not think it was particularly important information for the site (believing this information is conveyed by other means).



None of the participants, regardless of experience, looked past the first page of results or seemed to be aware there were multiple pages of results.

The menu tabs do not provide positional feedback.

Consider more traditional tabs with a background color change or other visual indication when a tab is selected.




Home > Find Studies > Search Results
Text Size
2093 studies found for: smoking
Modify this search | How to Use Search Results
Only three participants noticed the number of results displayed on the page, which changes format (the use of bold) and terminology ("Studies" versus "Trial") on the down-level page.
Find Studies
About Clinical Studies
Resources
About This Site
Home > Find Studies > Search Results > Study Record Detail
Text Size
Trial record 1 of 2093 for: smoking
Previous Study | Return to List | Next Study ▸

ClinicalTrials.gov

A service of the U.S. National Institutes of Health

Search for studies: Example: "Heart attack" AND "Los Angeles"

Find Studies | About Clinical Studies | Submit Studies | Resources

Home > Find Studies > Search Results

5674 studies found for: breast canc

Modify this search | How to Use Search R

List | By Topic | On a Map | Search Details

+ Show Display Options

Download | Subscribe to RSS

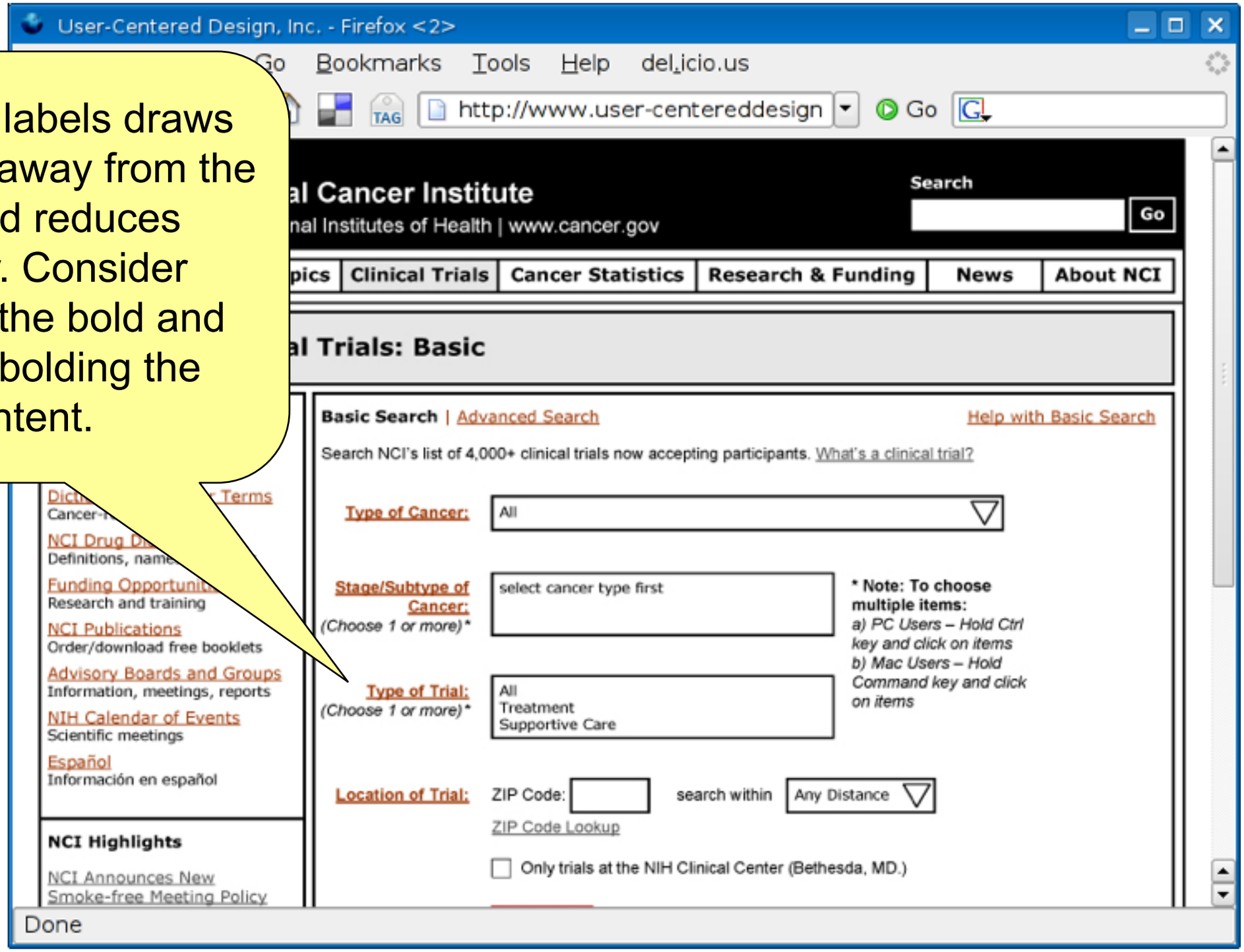
Include only open studies | Exclude studies with unknown status

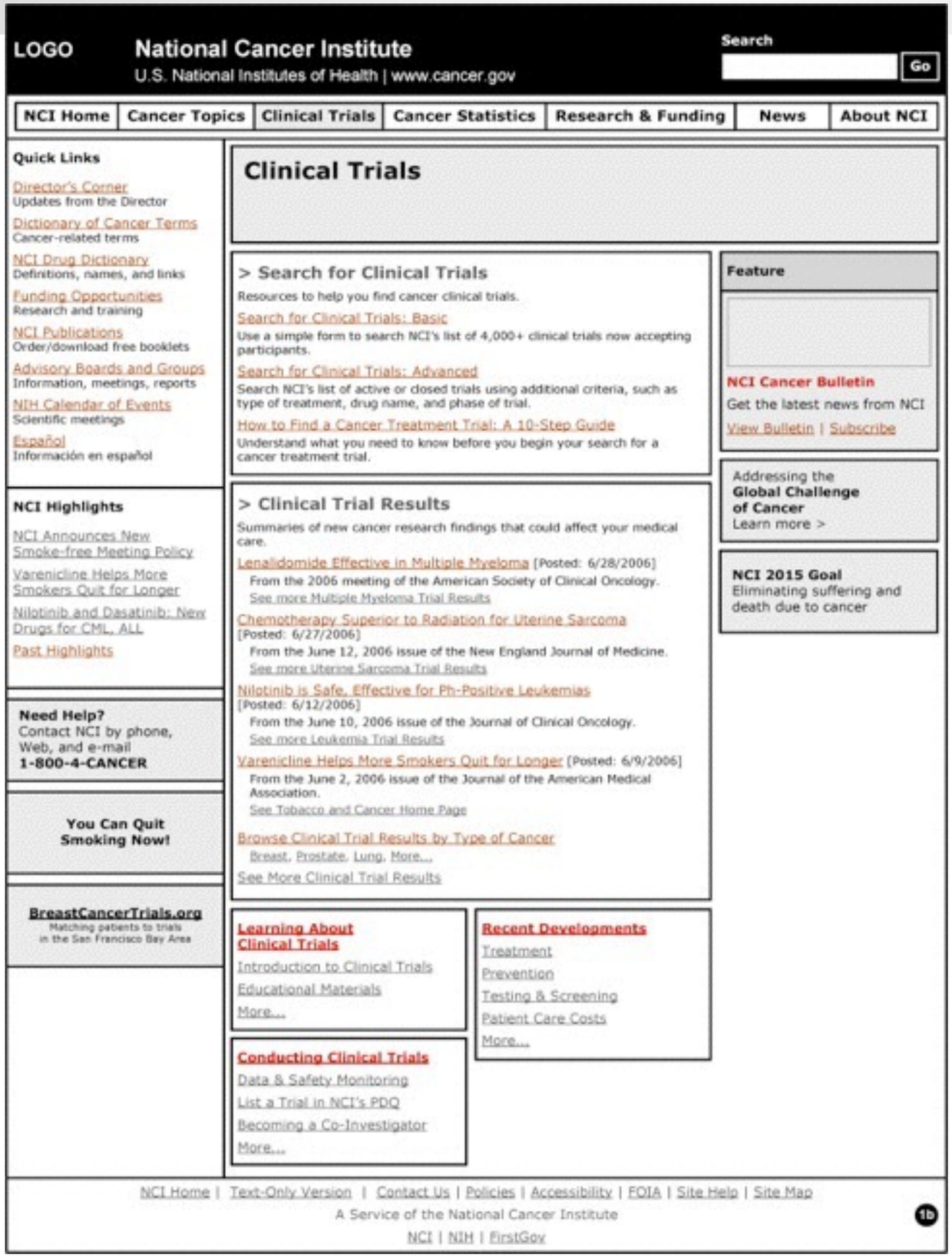
Rank	Status	Study
1	Recruiting	**Evaluate Risk/Benefit of Nab-paclitaxel in Combination With Either Gemcitabine or Carboplatin to the Combination of Gemcitabine and Carboplatin in First Line Treatment of Triple Negative Metastatic Breast Cancer (TNMBC)** **Conditions:** Breast Tumor; Breast Cancer; Cancer of the Breast; Estrogen Receptor- Negative Breast Cancer; HER2- Negative Breast Cancer; Progesterone Receptor- Negative Breast Cancer; Recurrent Breast Cancer; Stage IV Breast Cancer; Triple-negative Breast Cancer; Triple-negative Metastatic Breast Cancer; Metastatic Breast Cancer **Interventions:** Drug: nab-Paclitaxel 125 mg/m2 plus carboplatin AUC 2 in triple negative metastatic breast cancer (TNMBC) subjects; Drug: Carboplatin AUC 2; Drug: gemcitabine 1000 mg
2	Active, not recruiting	**Vaccination With Autologous Breast Cancer Cells Engineered to Secrete Granulocyte-Macrophage Colony-Stimulating Factor (GM-CSF) in Metastatic Breast Cancer Patients** **Condition:** Breast Cancer **Intervention:** Biological: Autologous, Lethally Irradiated Breast Cancer Cells

None of the participants, regardless of experience, looked past the first page of results or seemed to be aware there were multiple pages of results. When prompted, they looked around the page, scrolled, and eventually found the pagination controls at the bottom of the page.

All navigation should be in grouped together.

Bold form labels draws users eyes away from the form and reduces usability. Consider removing the bold and possibly bolding the content.





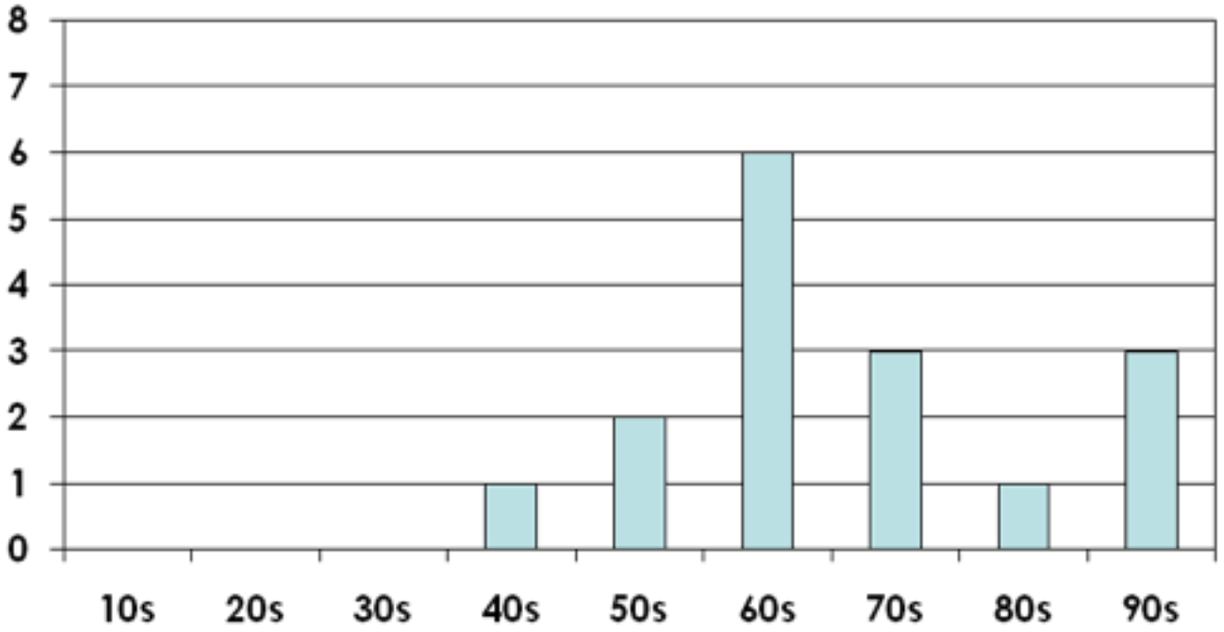
There are **50** hyper links on the home page (not including primary nav.) representing four levels within the clinical trial section and direct links to other parts of NCI

Participants (without prior exposure) failed to recognized the five primary disciplines as navigational elements. The most common expectation (if noticed at all) was that the links would provide definitions of the terms.

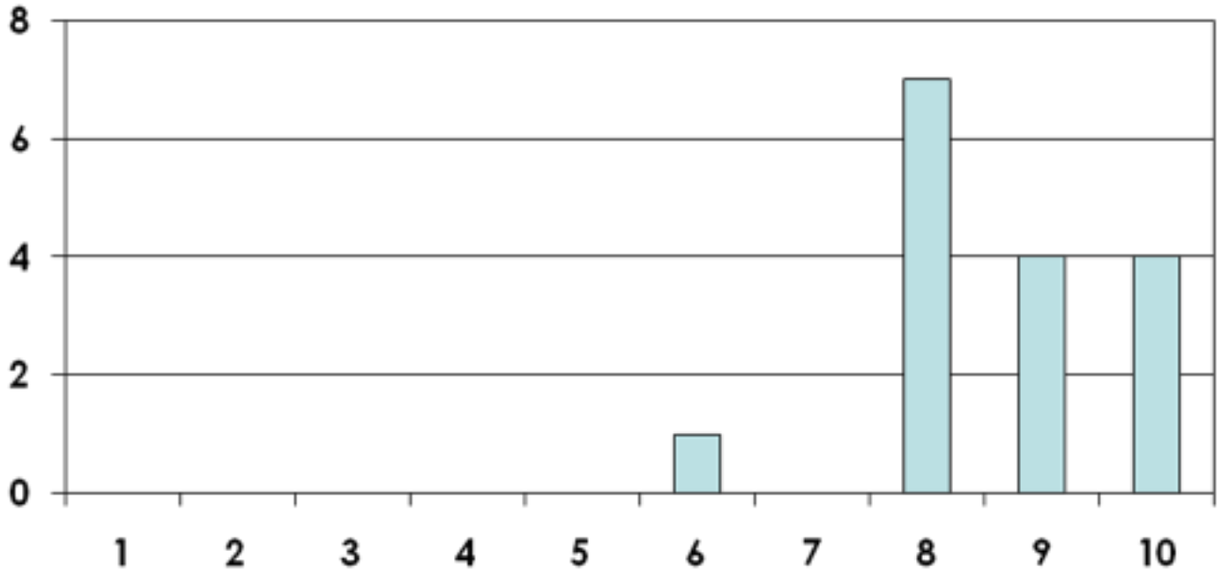
Participants had difficulty understanding what content was searched.
Many thought **all content** in Clinical Trials would be searched, not just ongoing trials

A few participants wanted to use the global NCI search to search Clinical Trials (consider labeling this "Search NCI" or "NCI Search"

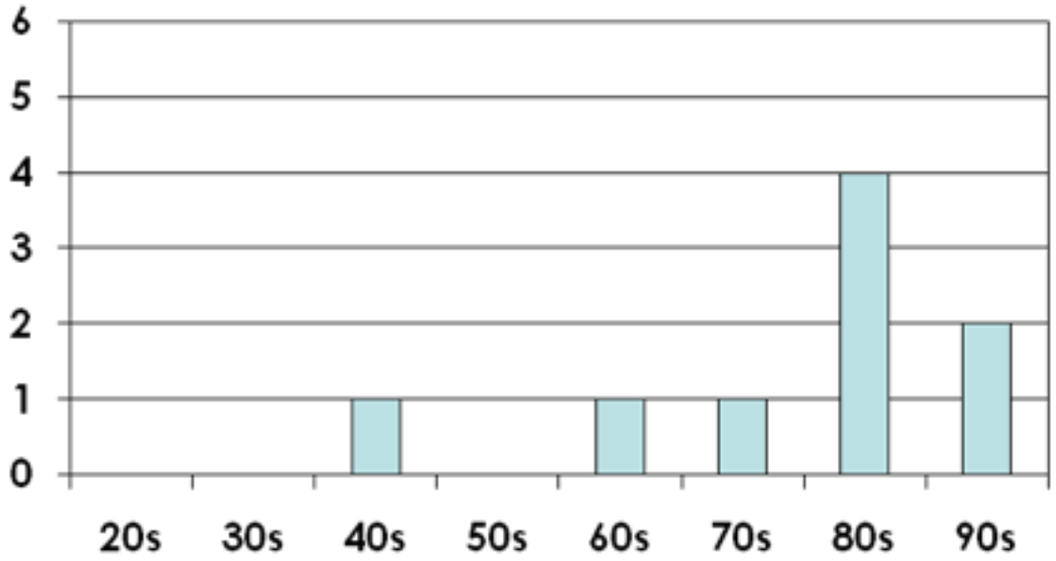
Some participants responded to the term "Find" even when the search form was on the page.



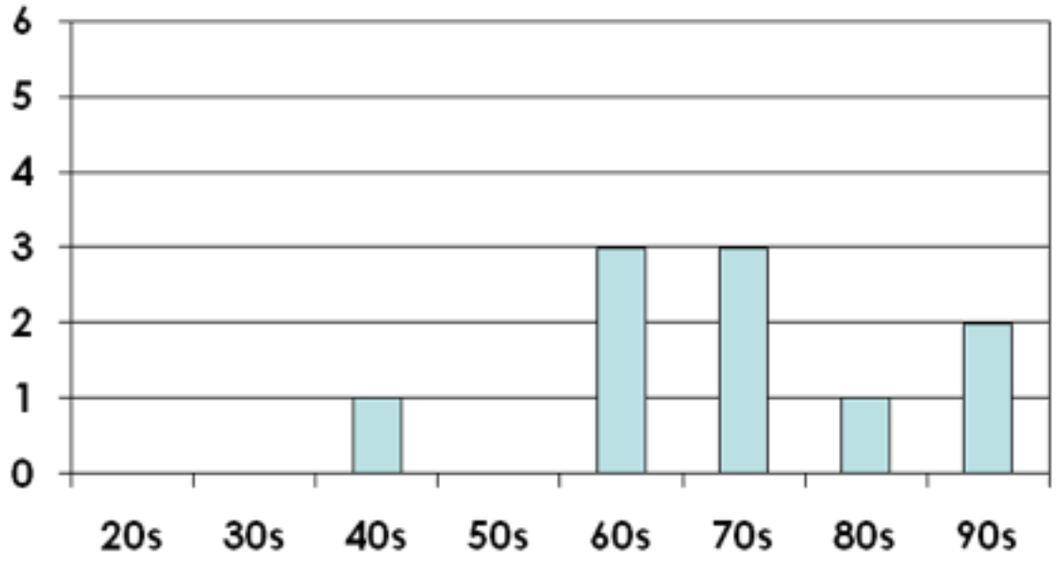
SUS



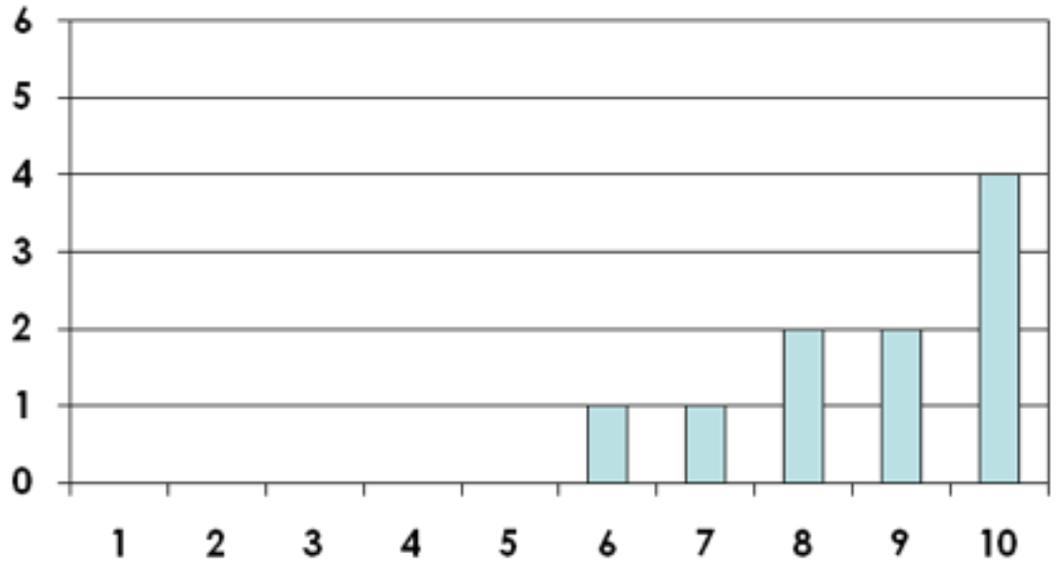
Cooper Harper



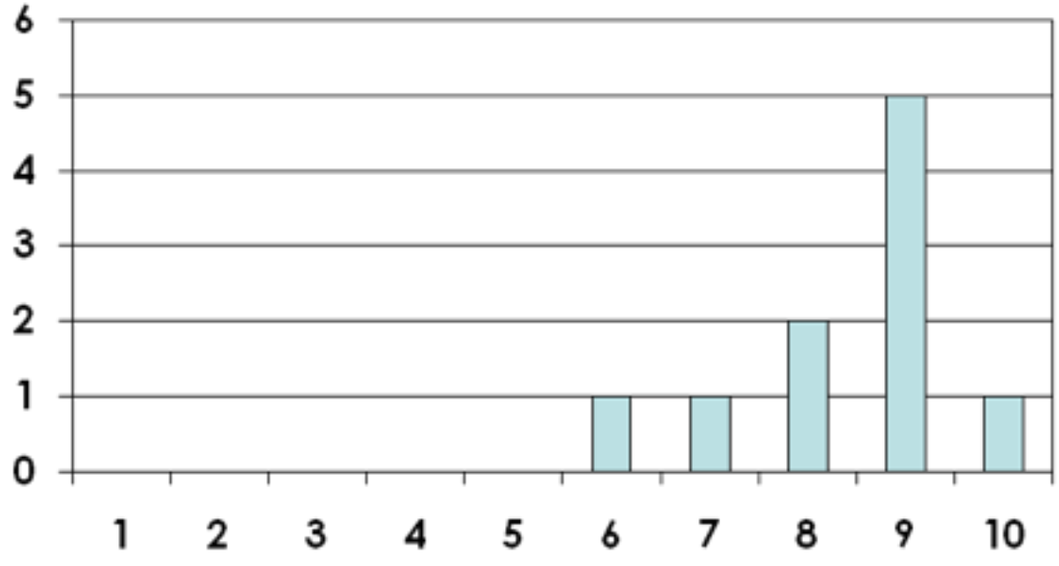
Memphis – SUS Data



DC – SUS Data



Memphis – MCH Data



DC – MCH Data

Conclusion

Origin of the Species

- Any “testing” is better than nothing, but don’t mistake “a 6 pack and friends” evaluation for testing
- Evaluations with human subject is highly valuable, it can be deeply insightful, the basic skills can be taught, but it is serious business and should not be conducted casually
- The more you know about experimental design the better your evaluations will be, but the more you know about users, designers, and usability evaluations the better the data will be

Origin of the Species

- Usability evaluations are best done early and often as part of a user-centered design process (part of what makes is user-centered)
- The intent of the evaluation should be to not just to know what happened, but to determine why it happened and to figure out what, if anything, can be done to make it better
- Unless you have the right conditions and a large sample set available, the is little distinction between a true expert review and small sample user-based testing, but experts often need users to “see” the data

Usability, Organizations, and Process

Origin of the Species

- The 1970s, when Hardware is King
  - 1950s – its an art
  - 1960s – there are degrees
  - 1970s – they're in management
- The 1980s, when Software is King
  - 1960s – its an art
  - 1970s – there are degrees
  - 1980s – they're in management
- 1990s, when "Interaction" should be King
  - 1970s – its an art
  - 1980s – there are degrees
  - 1990s – they should be in management

Processes

- System Development Models
  - Waterfall
  - Spiral
  - V-Model
- Software Development Models
  - Dynamic System Development Process (DSDP)
  - Joint Application Development Process (JAD) (circa 1970)
  - Structured Systems Analysis and Design Methodology (SSADM) (circa 1980)
  - Information Requirement Analysis/Soft System (circa 1980)
  - Object Oriented Programming (origins in 1960, but a common methodology in the 1990s)
  - Rapid Application Development (circa 1991)
  - Agile
    - Extreme Programming (circa 1990)
    - SCRUM

Processes (Concluded)

- “Interface Design Models
  - Star (Hartson & Hix, 1989)
  - LUCID (Cognetics, 2008)
  - ISO 13407/ISO 9241
  - Human Centered Design (IDEO)
  - User-Centered Design (the common term)
- Characteristics of a User-Centered Design Process
  - Design is a separate activity, distinct from development
  - Design should occur, completely, before development begins
  - Feedback is needed at many steps in the design process to…
    - Confirm the direction of design
    - Evaluate alternatives

Origin of the Species