Usability Testing Demystified

Bill Killam, MA CHFP
Common Testing Types

Nonfunctional Testing (Compliance Testing)
Regression Testing
Unit Testing
Integration Testing
Regression Testing
System Testing
Operational Acceptance Testing (Operational Readiness Testing)
User Acceptance Testing
Usability testing is assessing the ability of a specific group of people to perform specific tasks in a specific environment.
What are we really testing?
Perception
Count the black dots! :o)
User Consistency
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 as 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 affected.

• 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 than when people are asked to nod their head when they hear the same anchor point.
Answer Substitution

You are shown a picture of a person running for office and asked if you think they will win. There are far too many variables for you to make a good prediction, so the task is too hard for us to work out consciously.

As a result, our unconscious substitutes the hard question for an easier one – does the person look like a person who will win? Our unconscious provides an answer to that new question, but our conscious self reports it as the answer to the first question without realizing the substitution.
The Reality First Bias

• Feelings are from outside stimulus. When we imagine how we would feel, this is a pre-feeling. But we use our current feelings before assessing our pre-feelings. As a result, when asked questions, we integrate our current state into our answers, even if it's 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 the last event was positive, people tend to rate an overall experience as positive. If the last event was negative, people tend to rate an overall experience as negative. If asked about our lives when it's raining, people report lives that are worse then when the weather is nice.
Affect Heuristic

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.

Half of the participants in an experiment 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. The Superman group picked up more pencils and started helping sooner than the group that flew over in a helicopter.

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.
Misc. Unconscious Influences

• People holding cold drinks rate people as more aloof, but people are seen as more friendly when holding a warm drink

• J.Lo. and JZ's relationship was thought more likely to end when people were sitting in a wobbly chair
Priming
FINISHED FILES ARE THE RESULT OF YEARS OF SCIENTIFIC STUDY COMBINED WITH THE EXPERIENCE OF MANY YEARS
Attention
Test Your Attention
How many legs does this elephant have?
The Cognitive Reflexive Test
A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball.

How much does the ball cost?
If it takes 5 machines 5 minutes to make 5 widgets.

How long would it take 100 machines to make 100 widgets?
In a lake, there is a patch of lily pads. Every day, the patch doubles in size.

If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?
A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball.

How much does the ball cost?

Answer: The ball costs $.05. Which means the bat costs $1.05, otherwise the difference is not $1.
If it takes 5 machines 5 minutes to make 5 widgets.

How long would it take 100 machines to make 100 widgets?

Answer: 5 minutes - the rate is 1 widget every 5 minutes per machine, so the rate is now 1 every 5 minutes X 100 machines.
In a lake, there is a patch of lily pads. Every day, the patch doubles in size.

If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

Answer: 47 days. The size doubles each day, so it its completely covered in 48 days, it was 1/2 that size the day before.
Usability Testing Methods
Performance Based Testing

• This testing requires very carefully picked sample or just a very, very large sample to address the variability in populations (external validity)

• For validity, this testing requires a directly measurable outcome or dependent variable that is known to be associated with the alternatives being evaluated, which are the independent variables (internal validity)

• Requires noninterference with task performance and a realistic operating environment, which essentially means operational systems working in their intended environment without participant awareness of being tested (reliability)
Think Aloud Protocol

- Used extensively in cognitive interviews, its use has wained over the years.

- Though probably the most commonly used format because of its simplicity, but probably the most disruptive and least useful for user-based testing with interactive products (interactive mock ups, prototypes, and operational systems).

- Possible useful with non interactive artifacts (storyboards, static wireframes) as long as its limitations are understood.
Interrupted Task-based Testing

- A hybrid methodology
- Allow for semi-accurate performance data collection or the exploration of potential usability issues
- Highly dependent on the skill set of the observer for both accurate administration of testing and evaluation and generalization of the results
Some people look at this picture and see a monster.

We look at it and see low shutter speed and an incorrect F-stop. Kodak.
Questions?