Introduction to UX Methods & Metrics

Jeff Sauro, PhD
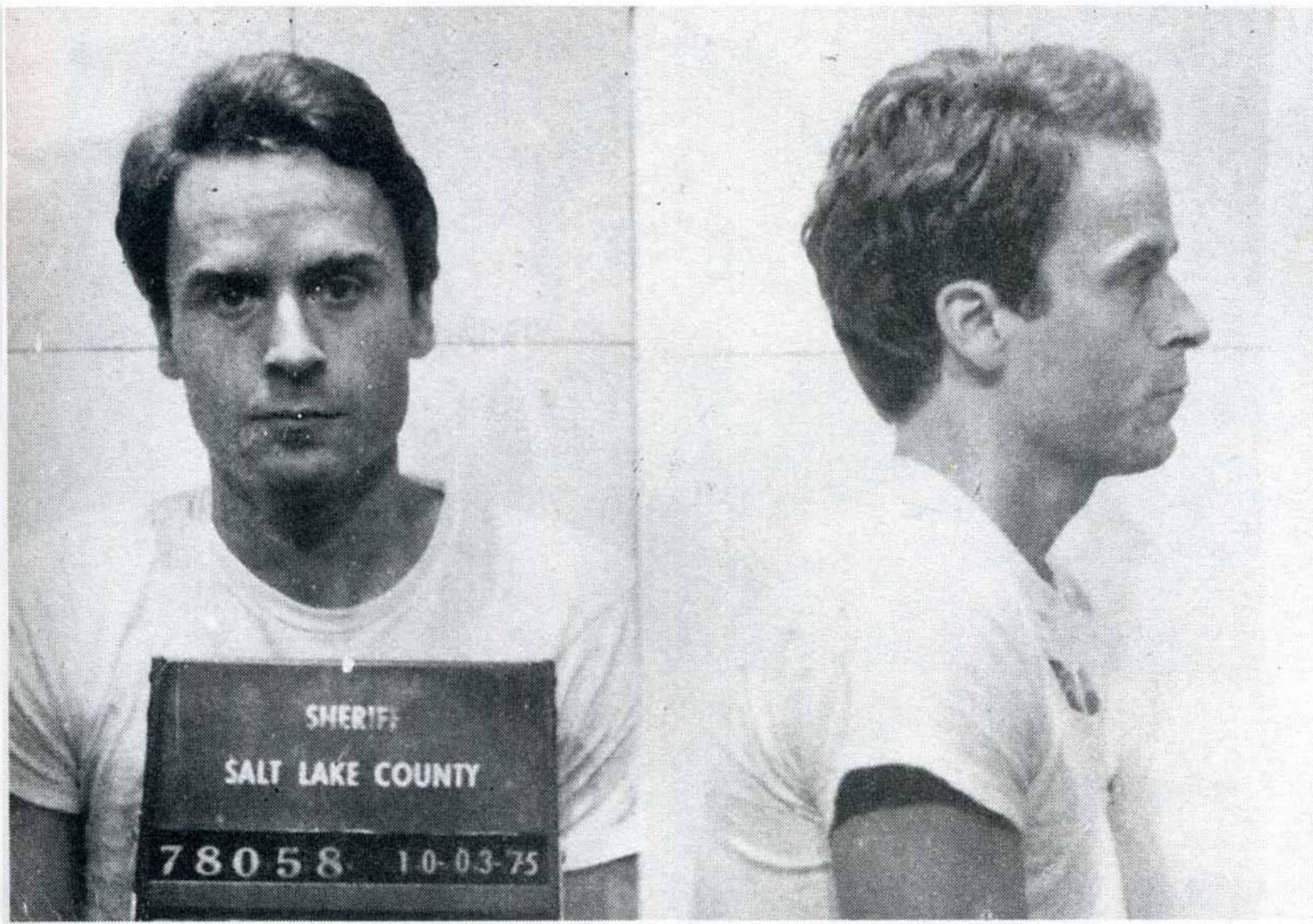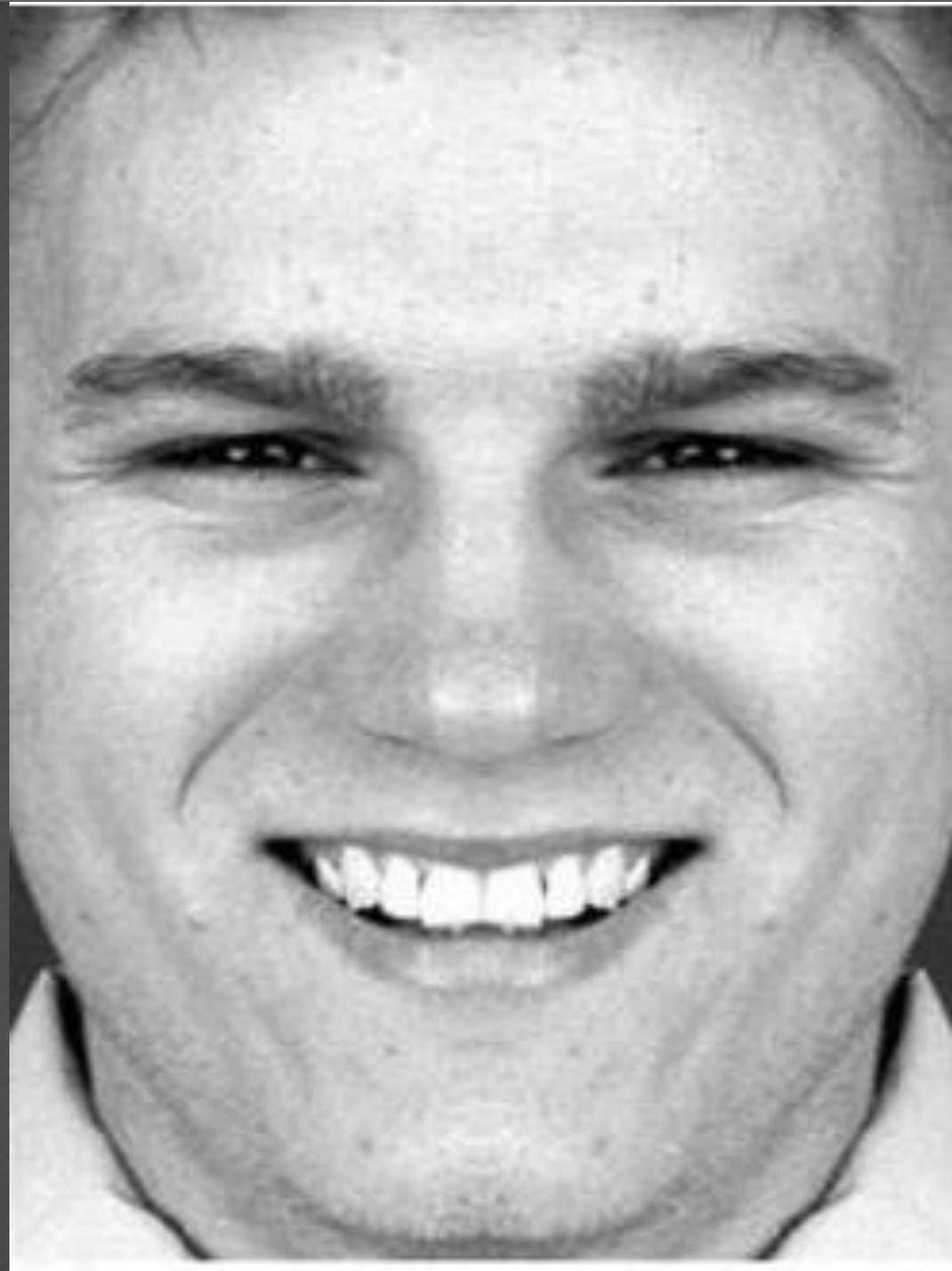
SHERIFF
SALT LAKE COUNTY
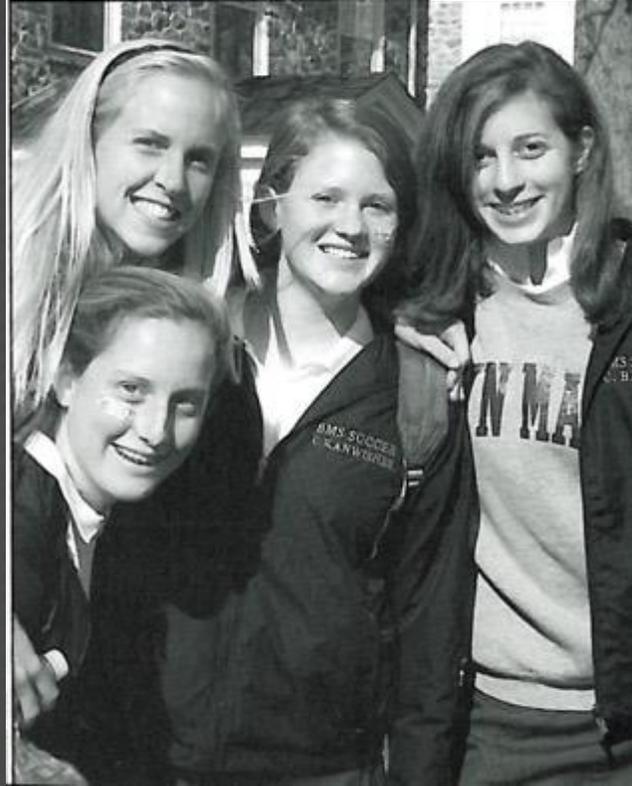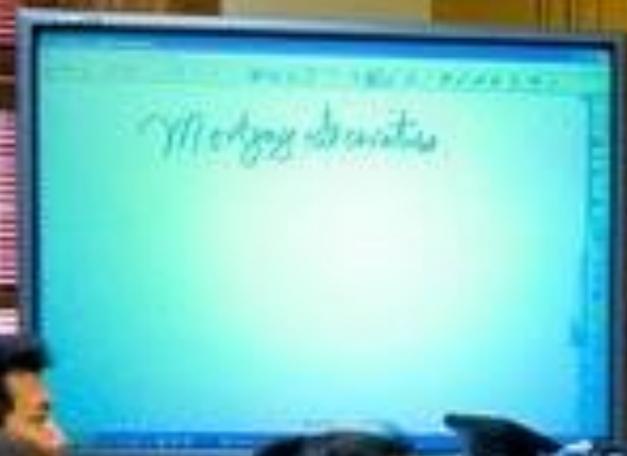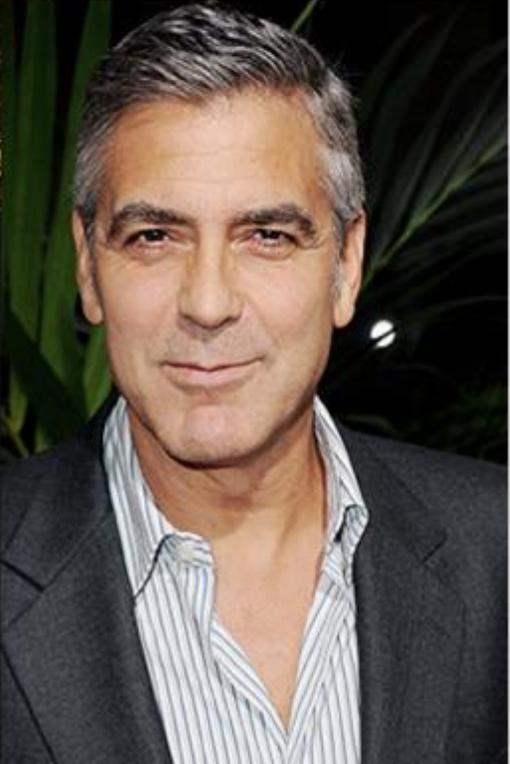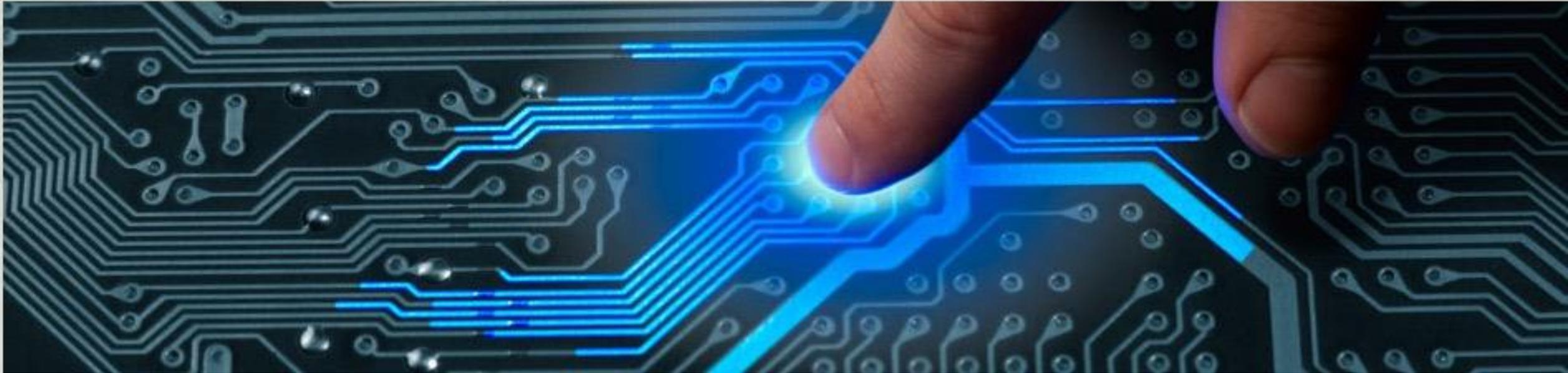
78058   10-03-75

Usability

ISO 9241 PT 11 Definition of Usability

COLLEGE GRADES

STANDARDIZED TEST SCORES

Attitudes
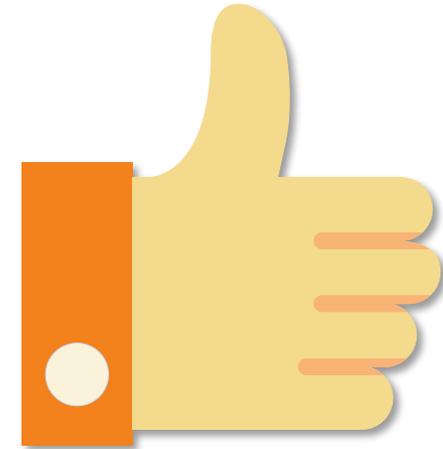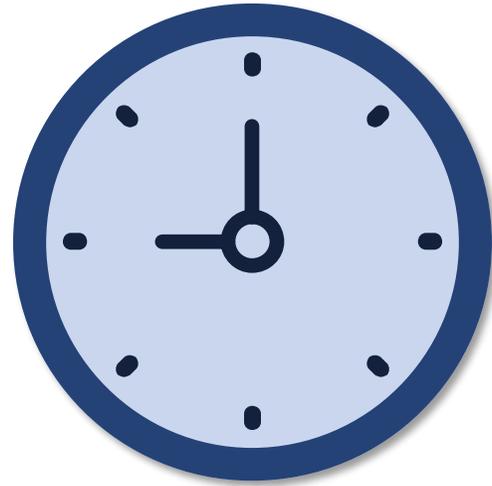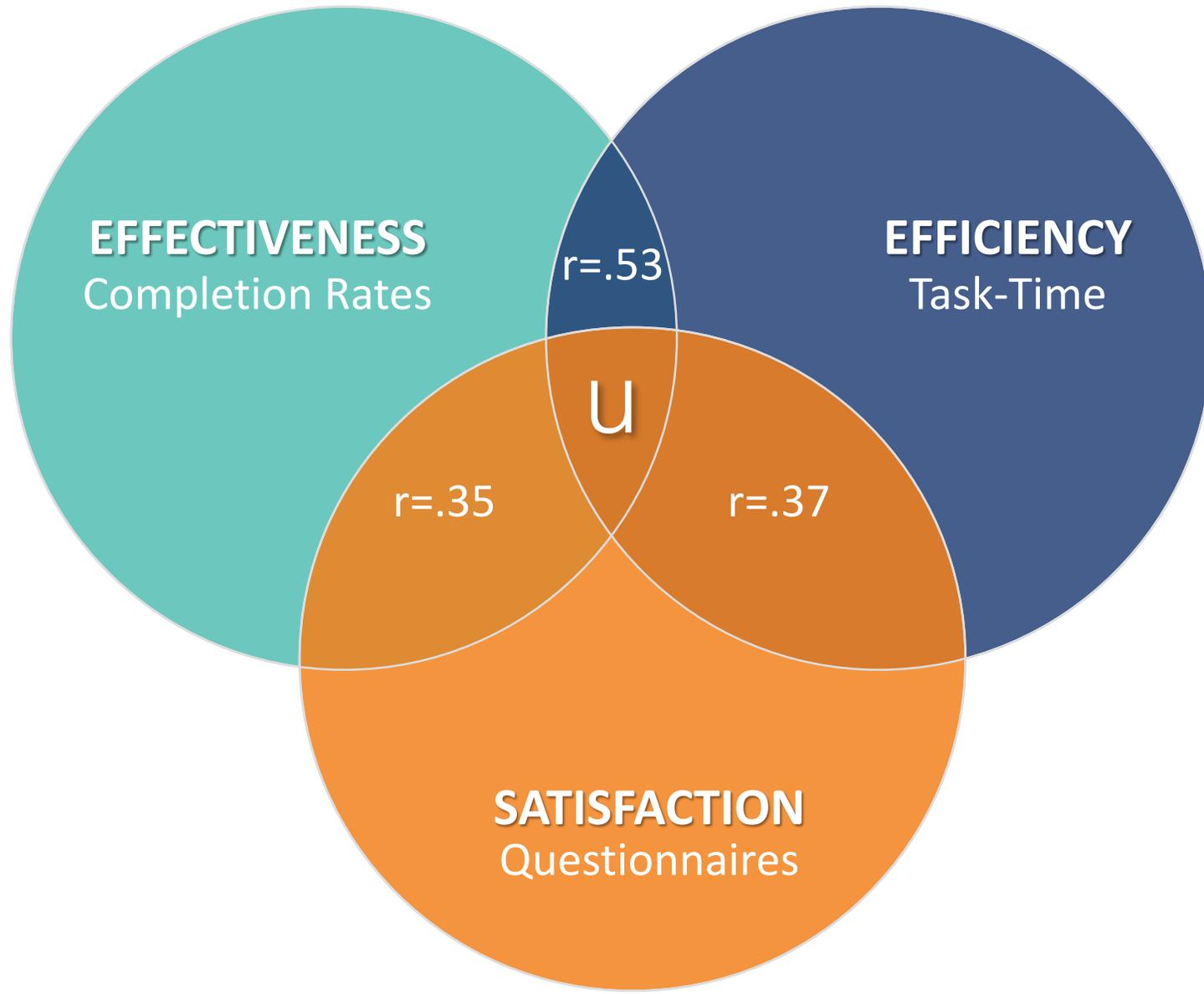
SUPR-Q Items & Factors

USABILITY

- This website is easy to use

- It is easy to navigate with this website

CREDIBILITY (TRUST)

- The information on this website is credible

- The information on this website is trustworthy

LOYALTY

- How likely are you to recommend this website to a friend or colleague? [NPS]

- I will likely visit the website in the future

APPEARANCE

- I found the website to be attractive

- The website has a clean and simple presentation

Selection of Websites in SUPR-Q Database

SUPR Q

Retail
Amazon
eBay
Zappos
Apple
Target
JC Pennys
Wal-Mart
Container Store
Crate & Barrel
Crumpler
Oriental Trading
Sears
LL-Bean
Michaels
Pier 1 Imports
Macys
Netflix
BestBuy
OfficeMax
OfficeDepot

Airlines
United Airlines
Frontier Airlines
American Airlines
Southwest
Jet Blue
Delta
Frontier

3rd Party Travel
Travelocity
Expedia
Orbitz
Kayak

Government
Illinois
Colorado
California
CDC
USA.gov
New York State

Floral Service
FTD
1-800 Flowers
ProFlowers

3rd Party Automotive
Edmunds
Kelly Blue Book
Cars
Autotrader

Financial Services
PayPal
Fidelity
TD Ameritrade
Vanguard
Bellco
US Bank
Chase
eTrade
Scrottrade
Wells Fargo

News & Information
Wall Street Journal
CNN
FoxNews
New York Times
IMDB
Yelp
Craigslist
Motley Fool
Yahoo

Social Networking
Facebook
Twitter
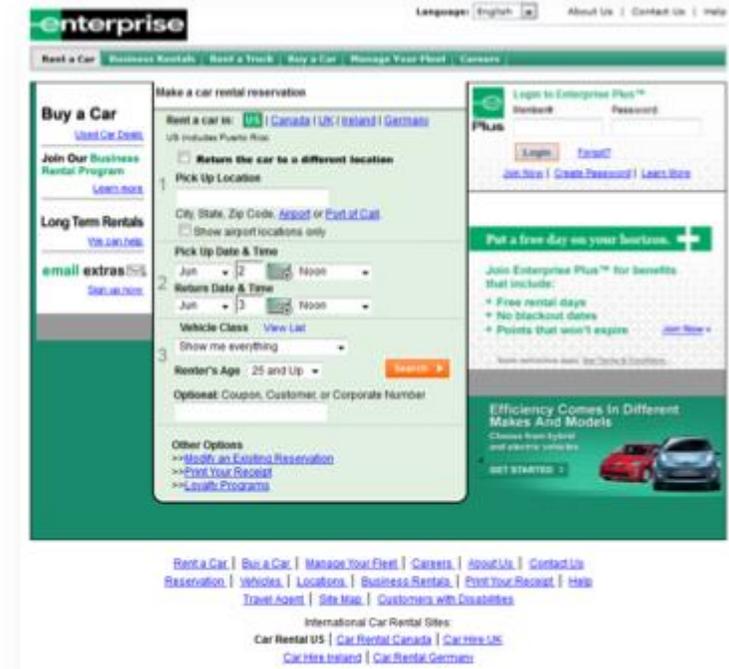LinkedIn
Flickr

Relationship/Dating
Match.com
eHarmony
PhentyofFish.com

Hotels
Hyatt
Hilton
Cell Phone Carriers
AT&T Wireless
Sprint
T-Mobile

200+
Total Websites

10,000
Responses

30-200
per Website
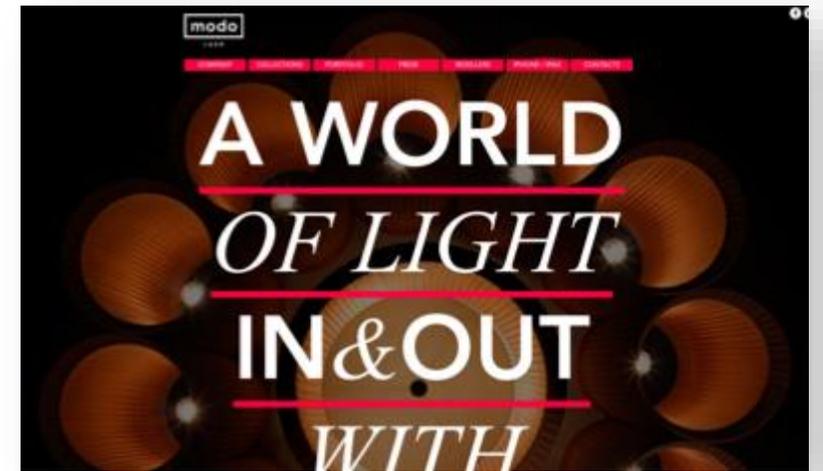
Remote Testing Setup Using Unmoderated Testing
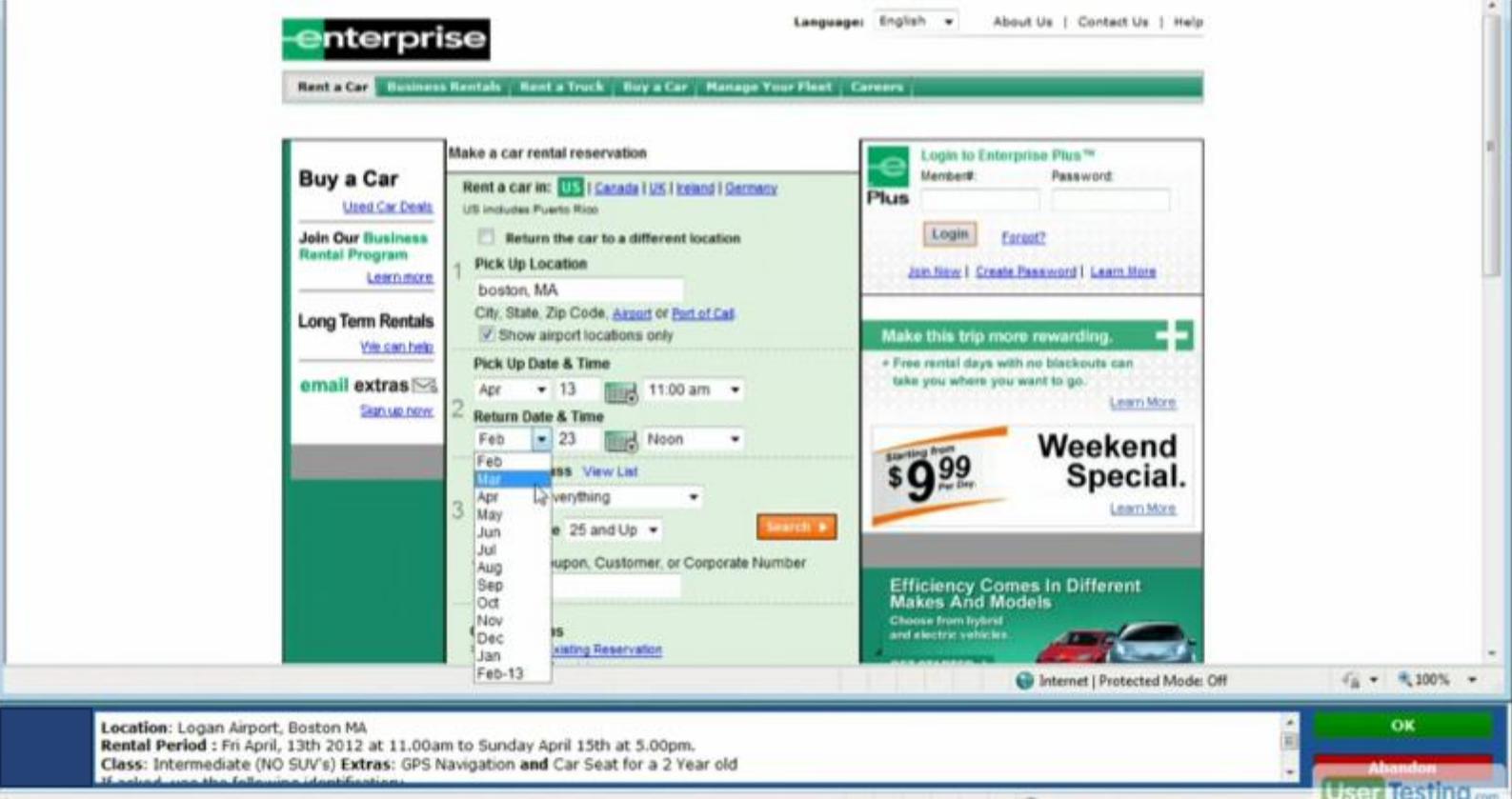
Task Interruption Penalty

USABILITY SCORE

20

15

10

5

0

14
16
16

5 Seconds   60 Seconds   No Limit
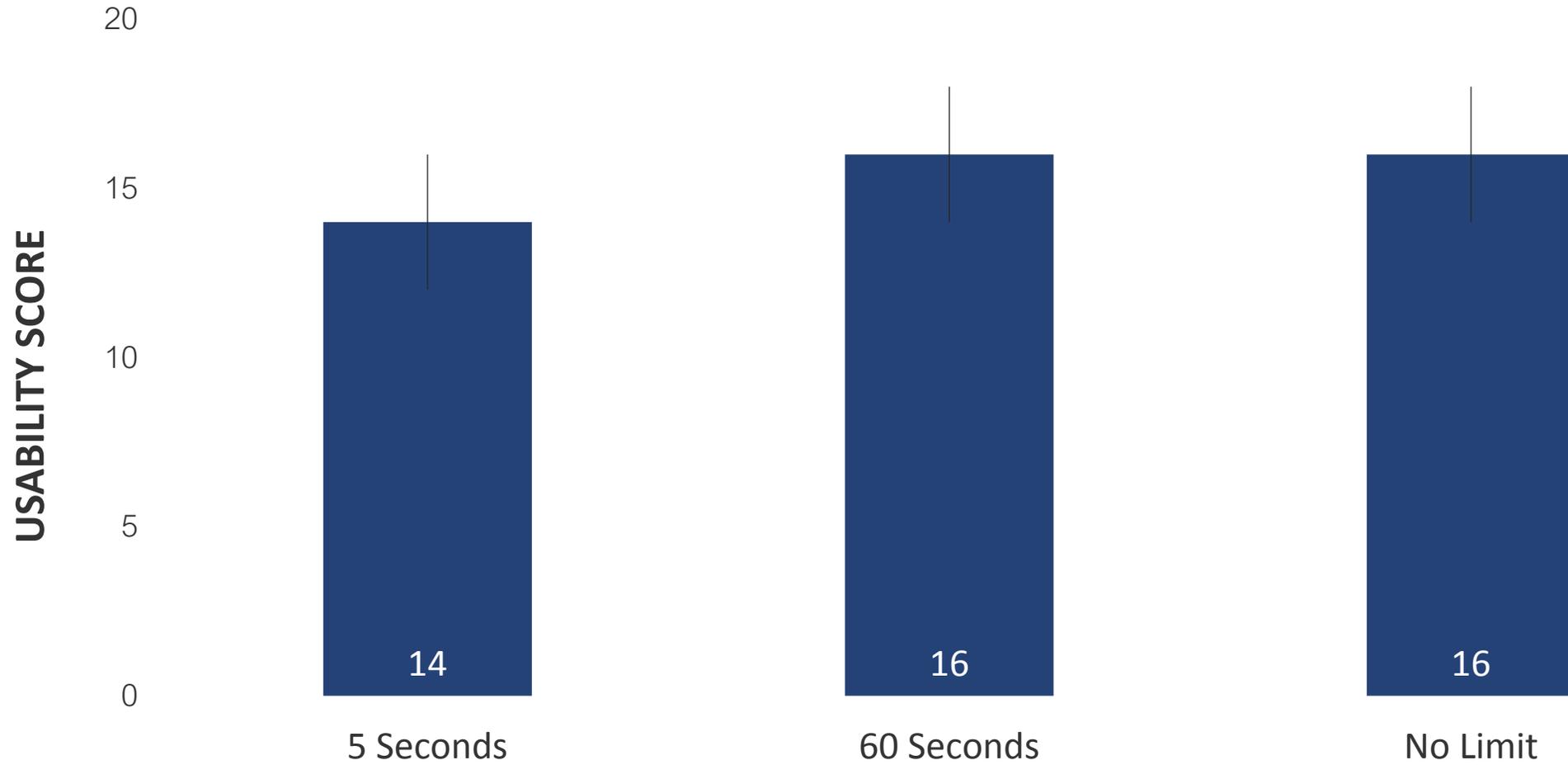
The 5 Second Penalty

	No Limit Task	5 Second (No Task)
Usability	39%	13% ⬇
Credibility	34%	13% ⬇
Appearance	35%	16% ⬇
Loyalty	28%	18% ⬇

Lasting Impressions Across 200 Websites

	During Usability Test		Retrospective
Usability	30%	⬆	62%
Credibility	33%	⬆	55%
Appearance	29%	⬆	58%
Loyalty	23%	⬆	64%

Can the first few seconds tell us anything reliable?

Attitudes & Actions

First Click Analysis: Websites

Wolfson, et al (2008)

If a user's first click is
down the wrong-path, only

46% eventually succeed

If a user's first click is
down the right-path,

87% eventually succeed

*Wolfson, et al (2008) Contextual card sorting (or FirstClick testing):
A new methodology for validating information architectures, UPA Proceedings*

First Click Analysis: Navigation

You are on an online department store website. Where would you go to find a Nano Tech Shaver?

Next

PERSONAL CARE
deodorant
eye care
feminine products
foot care
incontinence
shaving & grooming   Select
SKIN CARE
acne & blemish care
cleansers
exfoliants & masks
facial moisturizers
hair removal
lip care
lotions & creams
sun care
toners
SPA & MASSAGE
accessories
aromatherapy & relaxation

> **6x**

more likely to succeed

if

1st choice

is down right path

Framework to Improve the User Experience

Define

Measure, Analyze & Improve

Users

Tasks

Hypothesis

Metrics

Methods

Measure Before

Make Changes

Measure After
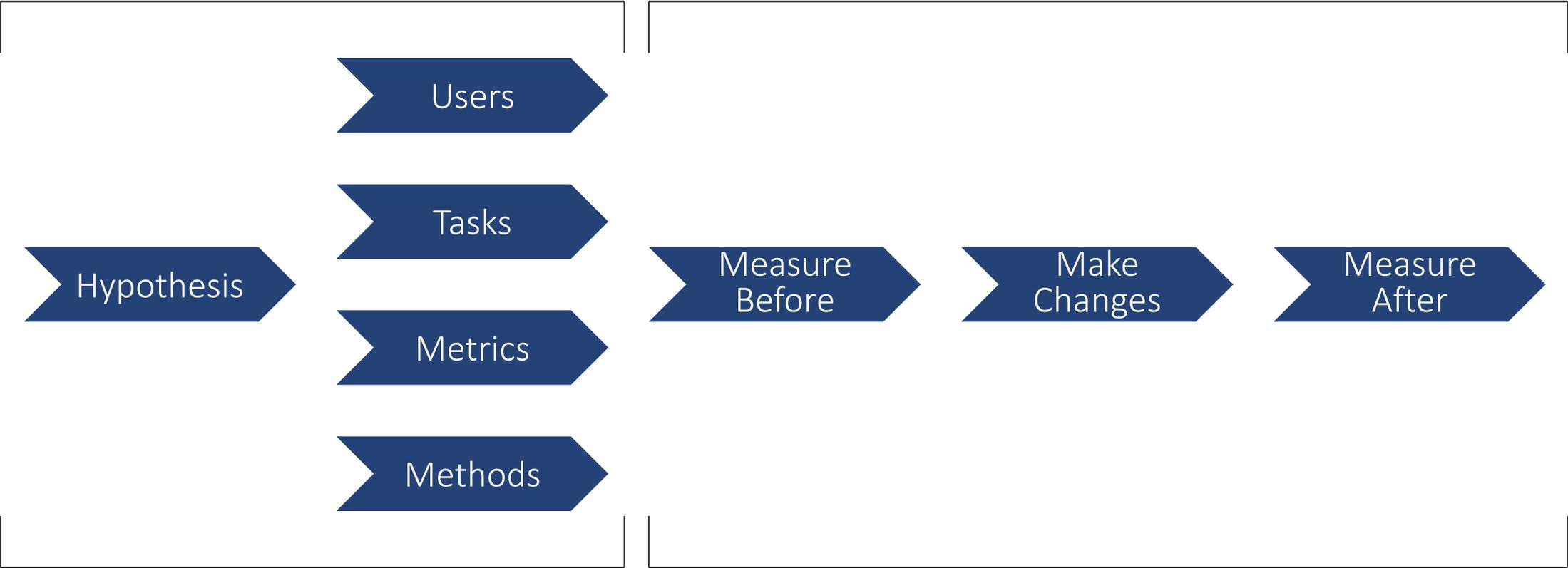
5 Key Takeaways

1. Usability is attitudes + actions—measure both.

2. If you only have 5 seconds, focus on the click.

3. Preference doesn't always equal performance. Attitudes don't always equal actions.

4. Context matters: retrospective vs. in-context.

5. Use a framework: define metrics, tasks then measure before and after.
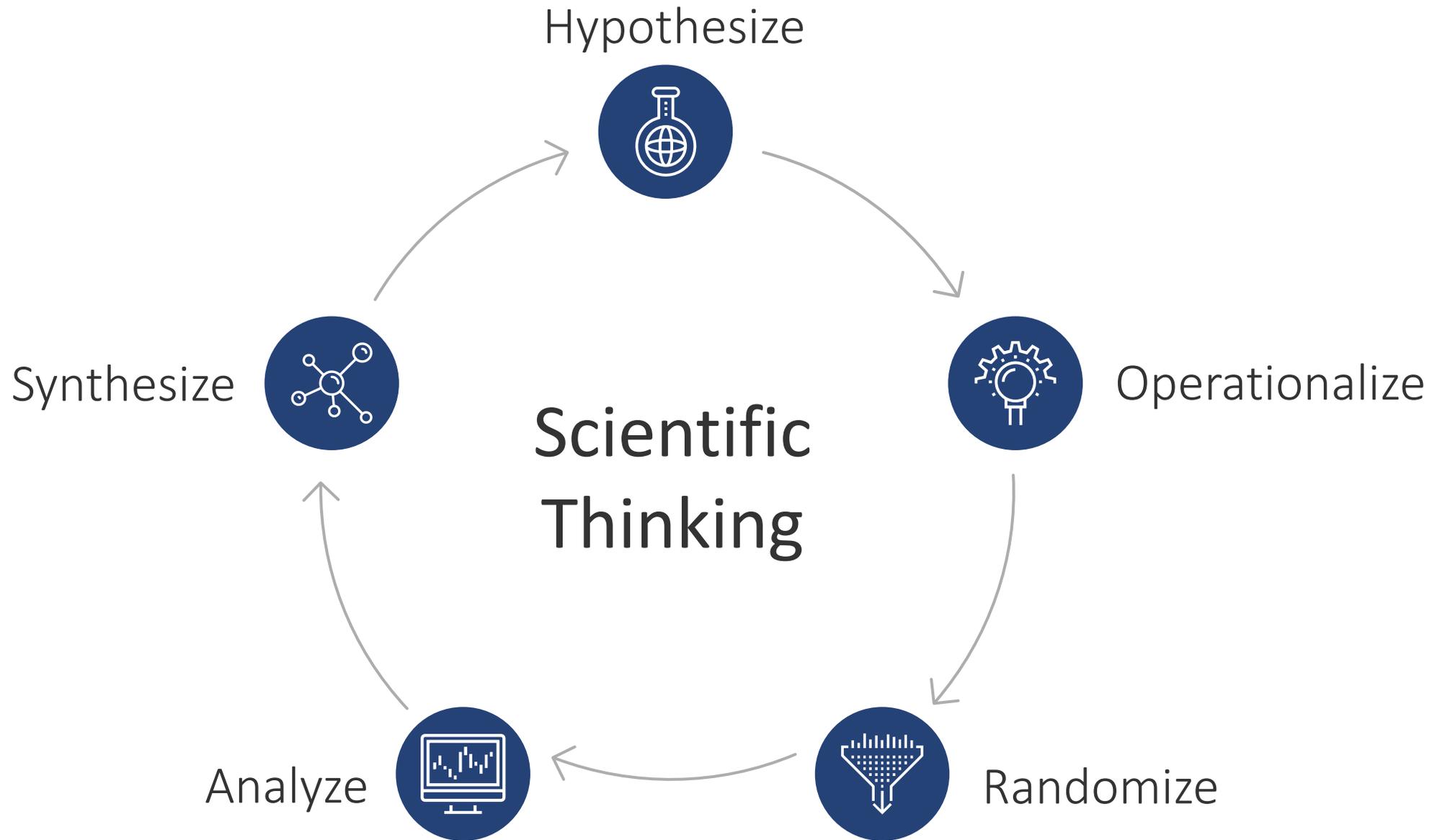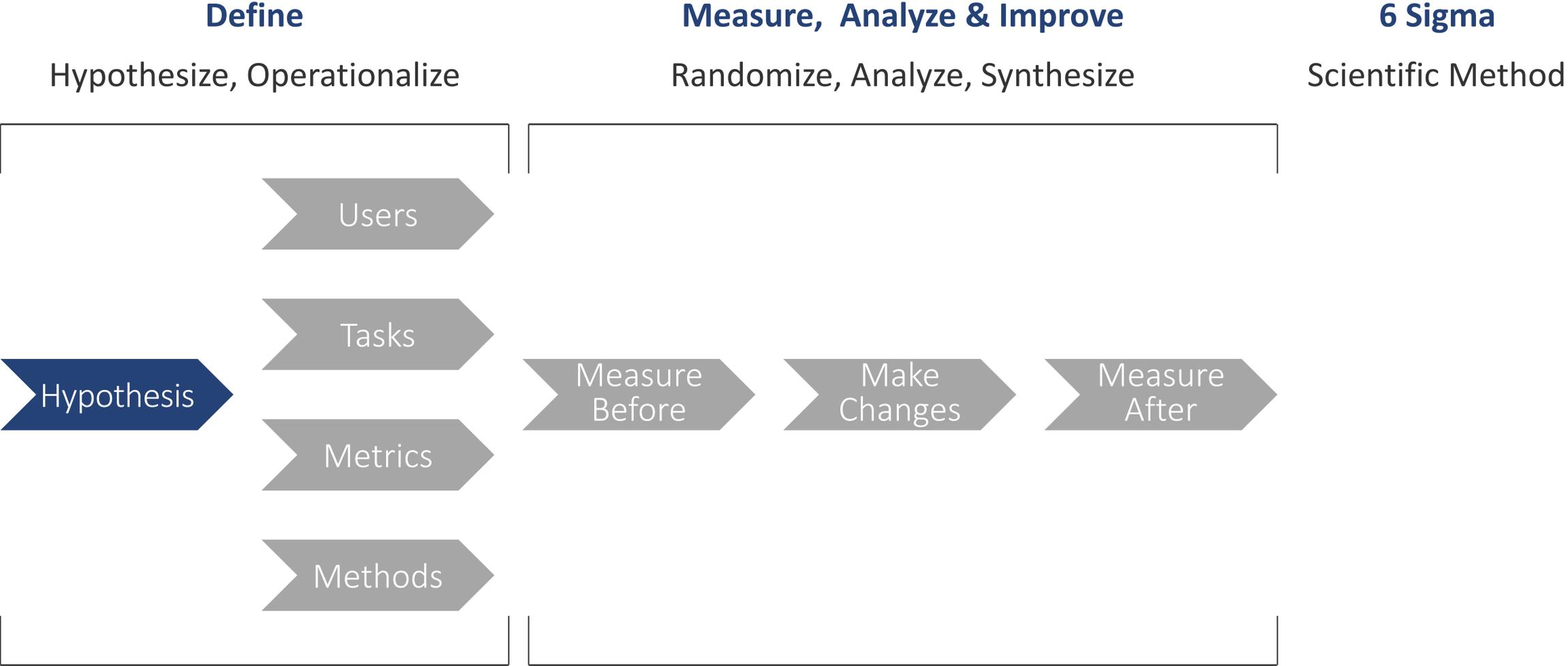
A Framework to Improve the UX

U I

U X

Hypothesize

Operationalize

Randomize

Analyze

Synthesize

Scientific Thinking

Framework to Improve the User Experience

Define

Hypothesize, Operationalize

Measure, Analyze & Improve

Randomize, Analyze, Synthesize
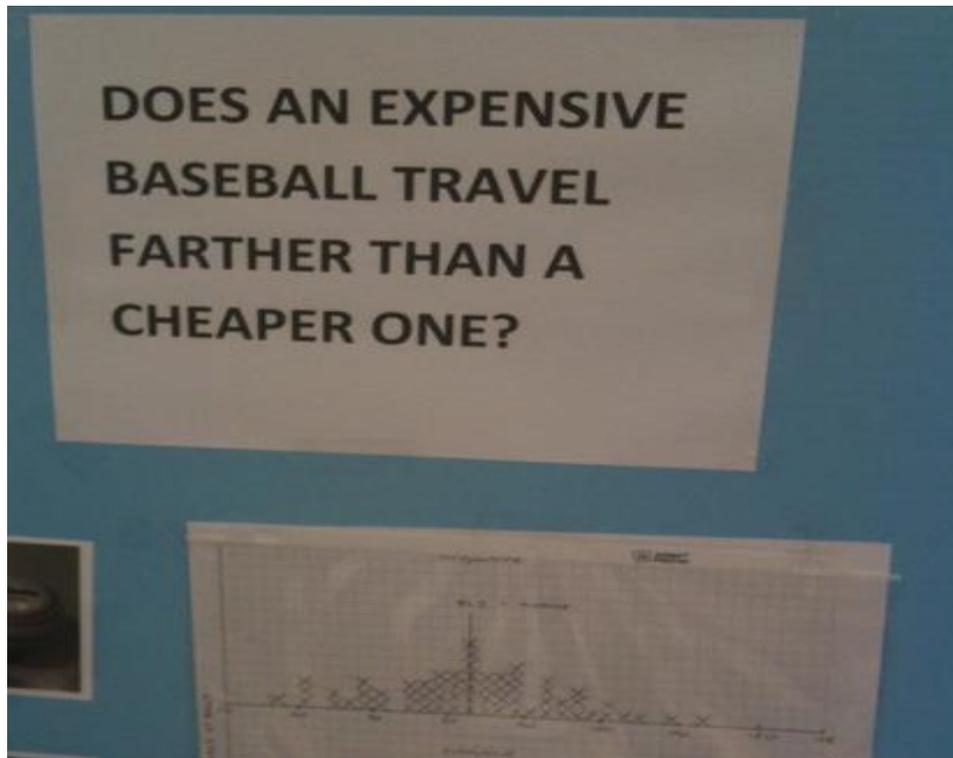
6 Sigma

Scientific Method

Hypothesize

Write your hypothesis in a way you can test.

Come up with a testable hypothesis

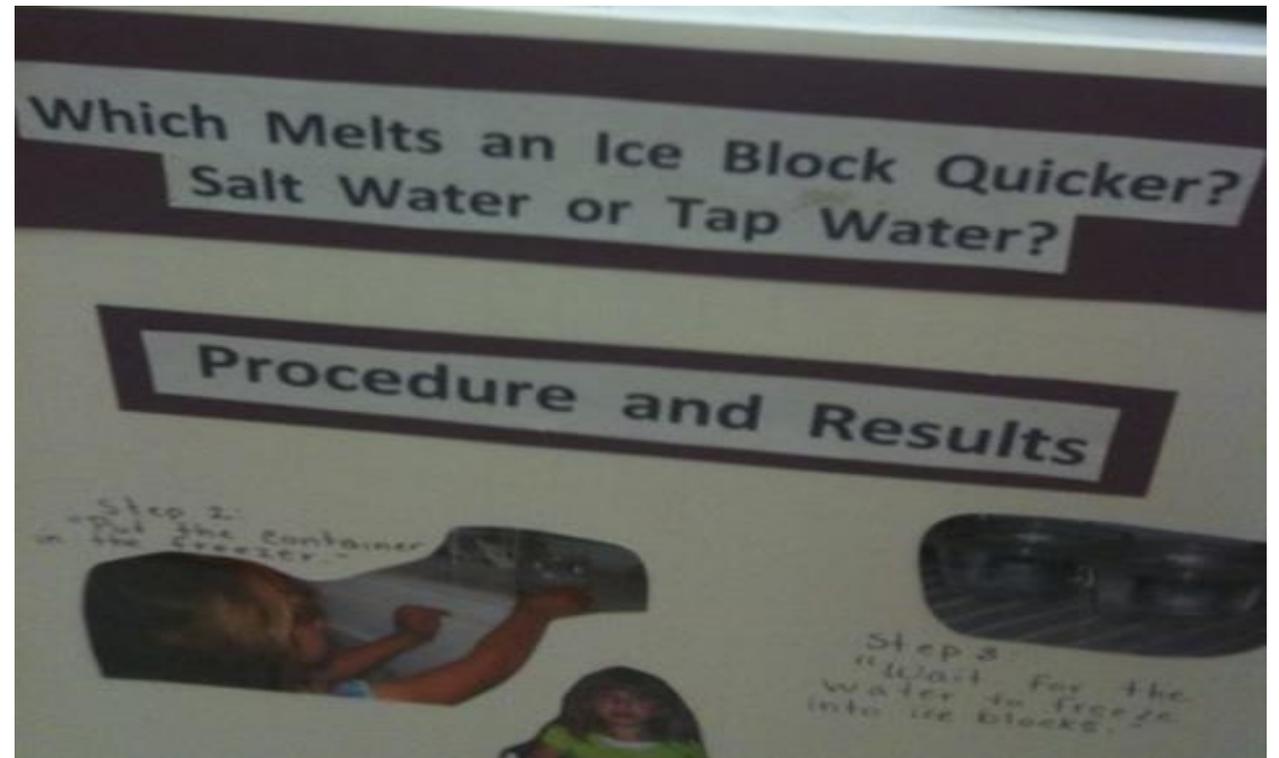DOES CHANGING THE AMOUNT OF VINEGAR AFFECT THE COLOR WHEN DYING EASTER EGGS?

DOES AN EXPENSIVE BASEBALL TRAVEL FARTHER THAN A CHEAPER ONE?

Which Melts an Ice Block Quicker? Salt Water or Tap Water?

Procedure and Results

Hypothesis:

The layout of the buttons (order and color) leads to accidental form cancelling.

Operationalize

Users › Metrics › Methods › Tasks

Who are Your Users?

LifeLock Needs Your Feedback

Thank you for visiting our website. Your opinion is very important to us. You have been randomly chosen to take a brief survey to help us improve your experience.

If you agree to participate, you can continue browsing the site and then you will be redirected to answer a few questions about your experience.

[Yes, I will participate] [No Thanks]

Powered by UserZoom - Privacy Policy

Close [x]

Who
- Gender
- Age
- Income
- Education
- Occupation
- Marital Status
- # of children

What they do
- Most recent purchase
- Total transactions
- Product experience (new vs. repeat)
- Years of experience
- Total revenue or profit

When
- Weekends vs. weekdays
- Life Events: after a baby, marriage, or move
- Daytime vs. evening
- Periodic activities: haircuts and toothpaste every five weeks

Where
- Rural vs. urban
- Domestic vs. Intl.
- City names and market size
- Regions & states
- Zipcodes

What they think
- Lifestyles: traveler vs. homebody
- Values: frugal vs. spendthrift
- Technology: early adopter vs. laggard
- Personalities: risk seeking vs. risk averse
- Overall product satisfaction: low vs. high

What they are likely to do
- Likelihood to recommend to friends
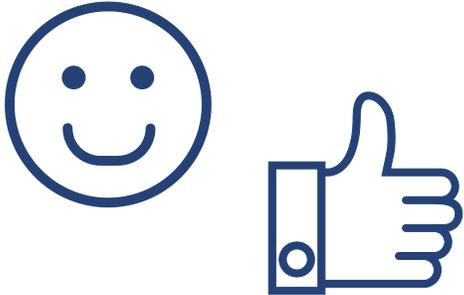- Likelihood to repurchase

Operationalize

Users → **Metrics** → Methods → Tasks
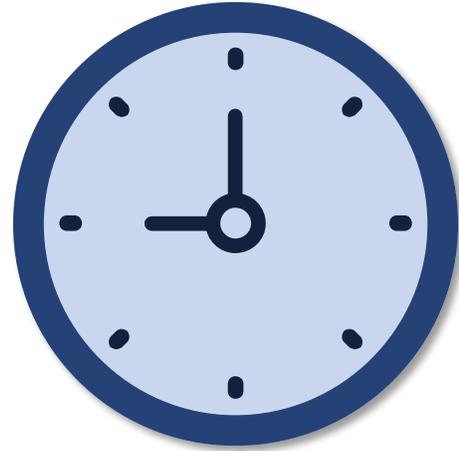
Measure the User Experience: Attitudes & Actions

Attitudes

Ease

Satisfaction

Likelihood to Recommend

Actions

Completion Rates

Time

Errors

Task Based "Action Metrics"

Completion & Findability Rates

- CODE TASK SUCCESS AS 1
- CODE TASK FAILURE AS 0

- Define a success criteria and do not give partial credit in the computations

Task – Level Difficulty

SINGLE EASE QUESTION (SEQ)

Overall, this task was?

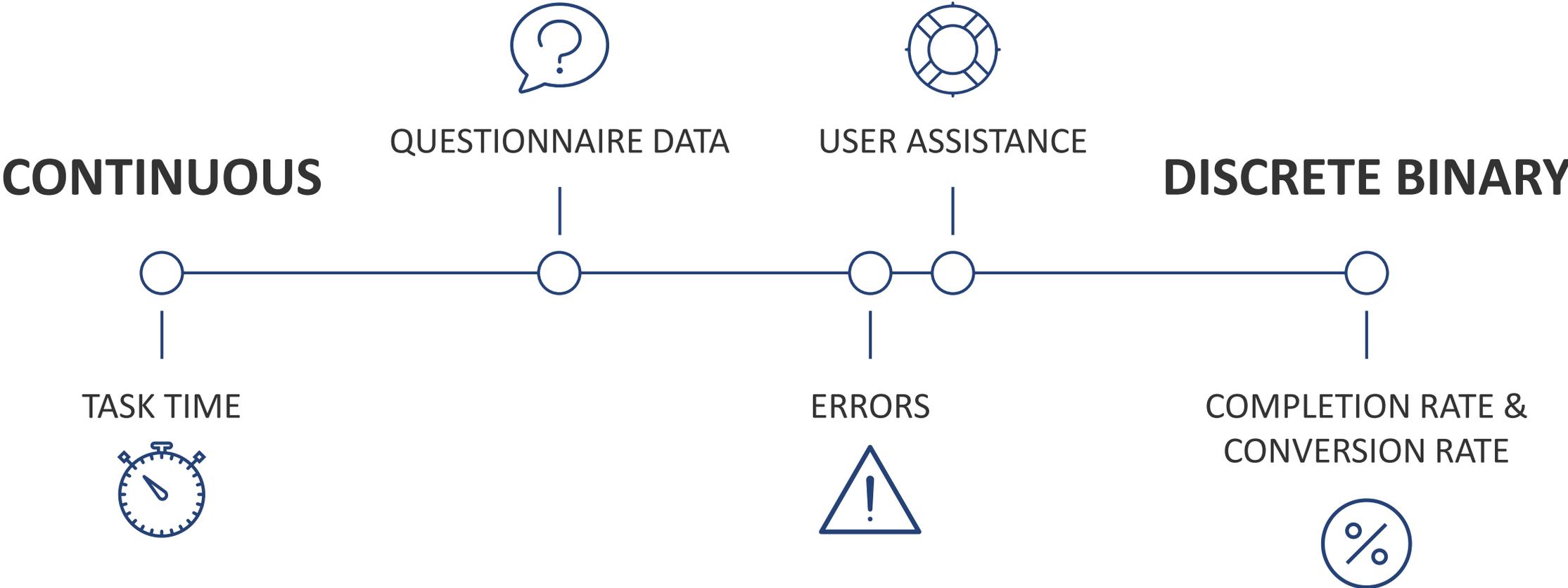Very Difficult   Very Easy

    1   2   3   4   5   6   7

    ○   ○   ○   ○   ○   ○   ○

Task Time

- AVERAGE TASK COMPLETION TIME

- MEAN TIME TO FAILURE

- AVERAGE TIME ON TASK

Metric Types

QUESTIONNAIRE DATA

USER ASSISTANCE

CONTINUOUS

DISCRETE BINARY

TASK TIME

ERRORS

COMPLETION RATE &
CONVERSION RATE

Attitude Metrics

Satisfaction: General & Specific

Ask General

Overall how satisfied are you with the [Brand][Product]?

Specific

How satisfied are you with :

- Customer service
- Setup and customization
- Purchasing process

Not at All Satisfied	Slightly Satisfied	Somewhat Satisfied	Very Satisfied	Completely Satisfied

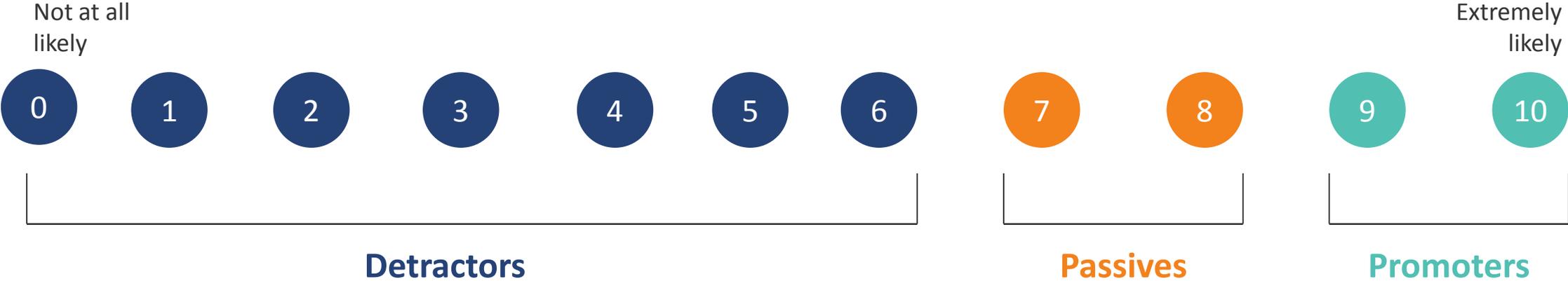# Satisfied but not Loyal

>90%

of car customers are
satisfied or very satisfied

<40%

Repurchase the same brand

Loyalty: Net Promoter Score

How likely are you to recommend to a colleague or friend?

Not at all
likely

Extremely
likely

(0) (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)

Detractors   **Passives**   **Promoters**

NPS = **% of Promoters** - **% of Detractors**

(9s & 10s)   (0-6)

Net Promoter, NPS, and Set promoter Score are trademarks of Satmetrix Systems, Inc., Bain & Company, and Fred Reichheld

Systems Usability Scale (SUS)

I think that I would like to use this system frequently

I found the system unnecessarily complex

I thought the system was easy to use

I think that I would need the support of a technical person to be able to use this system

I found the various functions in this system were well integrated

I thought there was too much inconsistency in this system

I would imagine that most people would learn to use this system very quickly

I found the system very cumbersome to use

I felt very confident using the system

I needed to learn a lot of things before I could get going with this system

1   **2**   **3**   **4**   **5**

Strongly Disagree

Strongly Agree

Using SUS

- Items are rescaled from 0 to 100 instead of from 0 to 40

- SUS scores are not percentages

0   20   40   60   80   100

Poor   Average   Excellent

Raw SUS Scores & Percentile Ranks
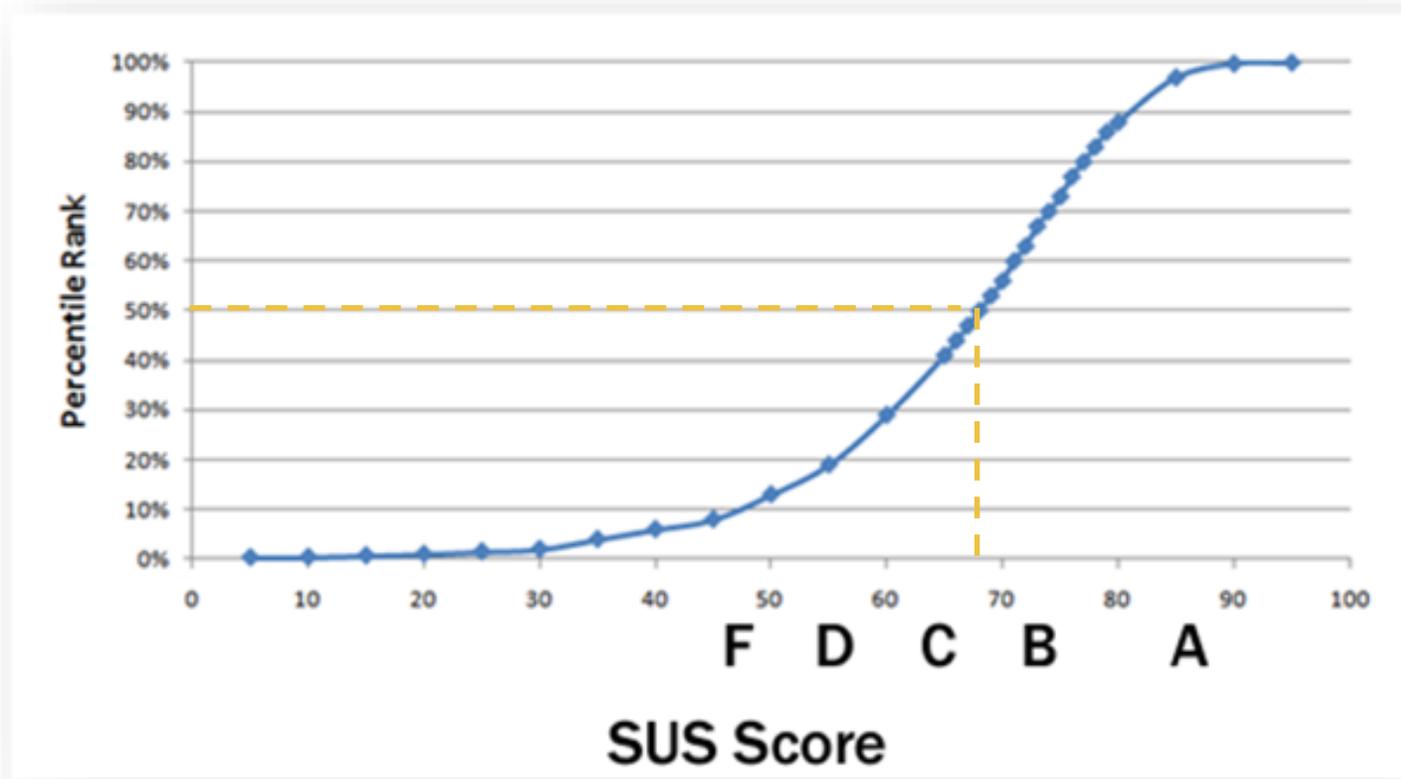
Average SUS Score

68

(500 products)

Consumer Software: SUS & NPS

S U S

50   60   70   80   90   100

N P S

-20   -10   0   10   20   30   40   50
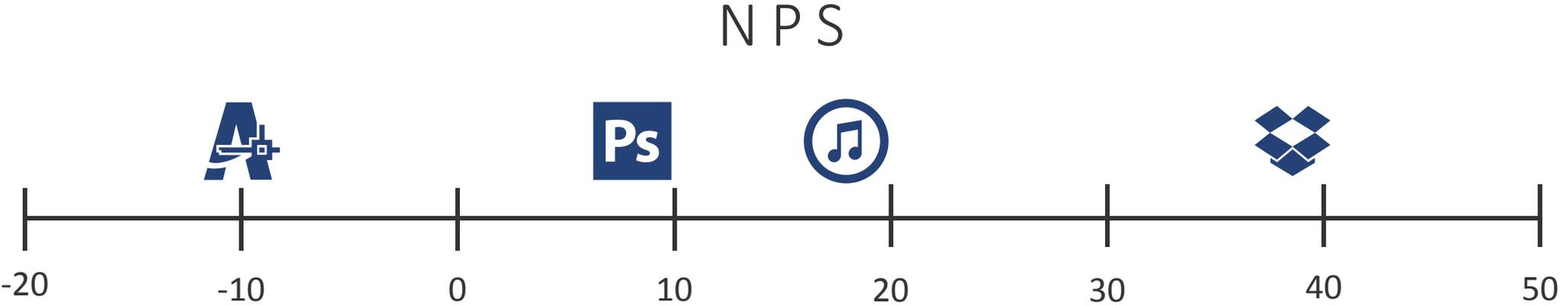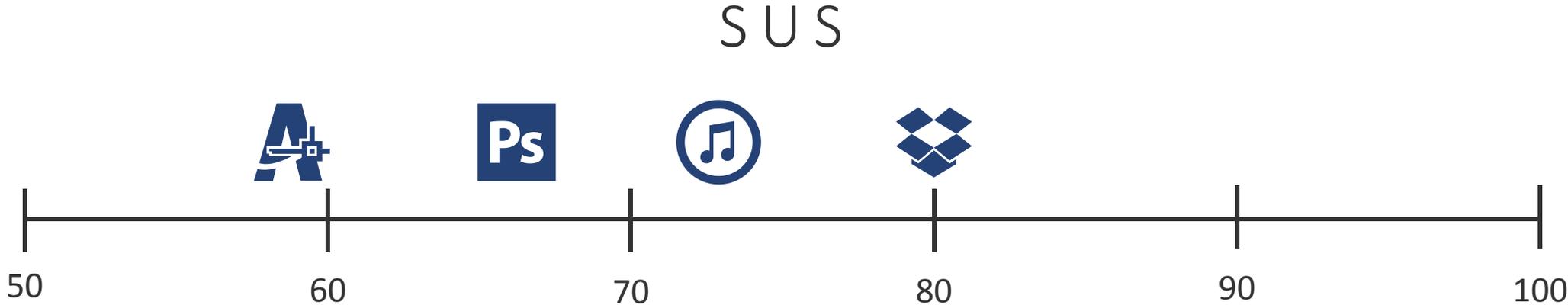
SUS explains 30%-50% of NPS

Quick Quiz: Continuous or Discrete Binary?

? Percentage of users completing a task

? Percent of users purchasing a monthly subscription

? Average order value

? Time to Register

? Average difficulty rating on a 1-7 scale

Attitude Does Not Always Equal Action

Which Form do you Prefer?

Operationalize: Errors and Hesitations

B2B Measurement

- Same Metrics Work

- Less emphasis on NPS, public facing

- More on task metrics

Operationalize

Users ▸ Metrics ▸ **Methods** ▸ Tasks

Evaluation Methods

Actions/Doing Things

Actual Use

Google Analytics

A/B Testing

Observation

Simulated Use

Usability Testing

Tree-Testing

1st Click Testing
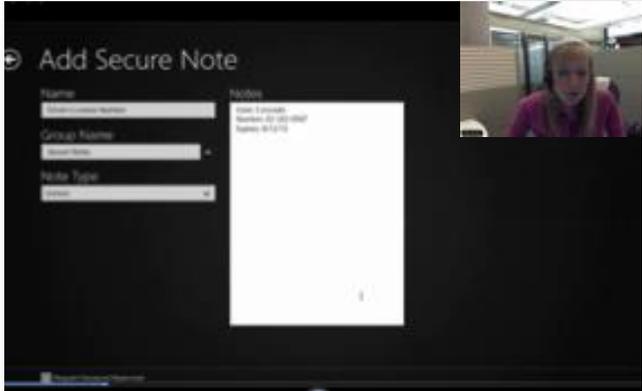
Attitudes/Saying Things

Surveys

Interviews

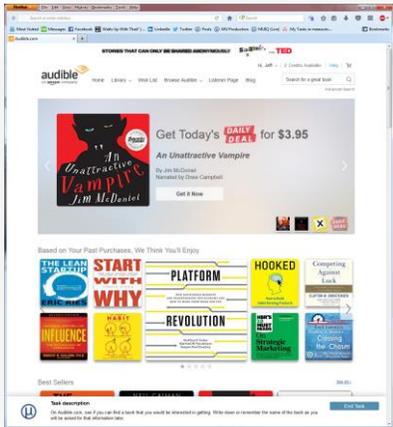# Usability Testing Modes

Remote Moderated

Remote Unmoderated

Remote Think Aloud

Operationalize

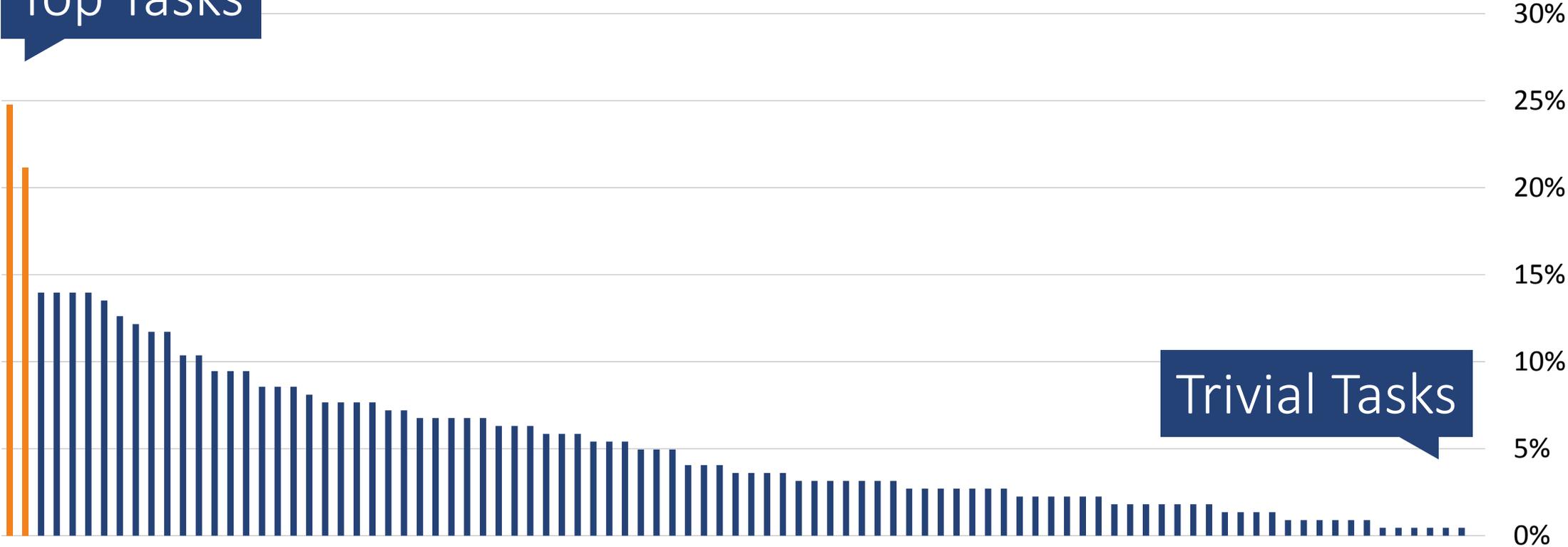Users → Metrics → Methods → **Tasks**

Doing Things: Measuring Task Behavior

Framework to Improve the User Experience

Define

Measure, Analyze & Improve

Users

Tasks

Hypothesis

Metrics

Methods

Measure Before

Make Changes

Measure After
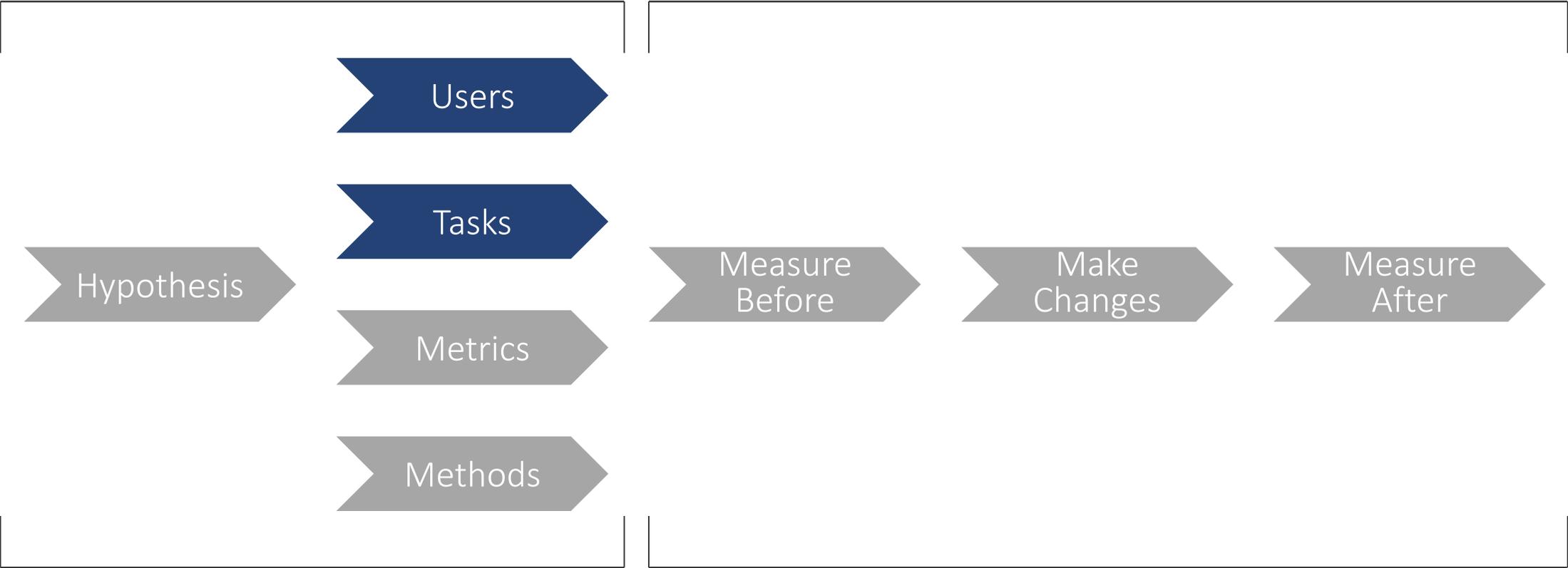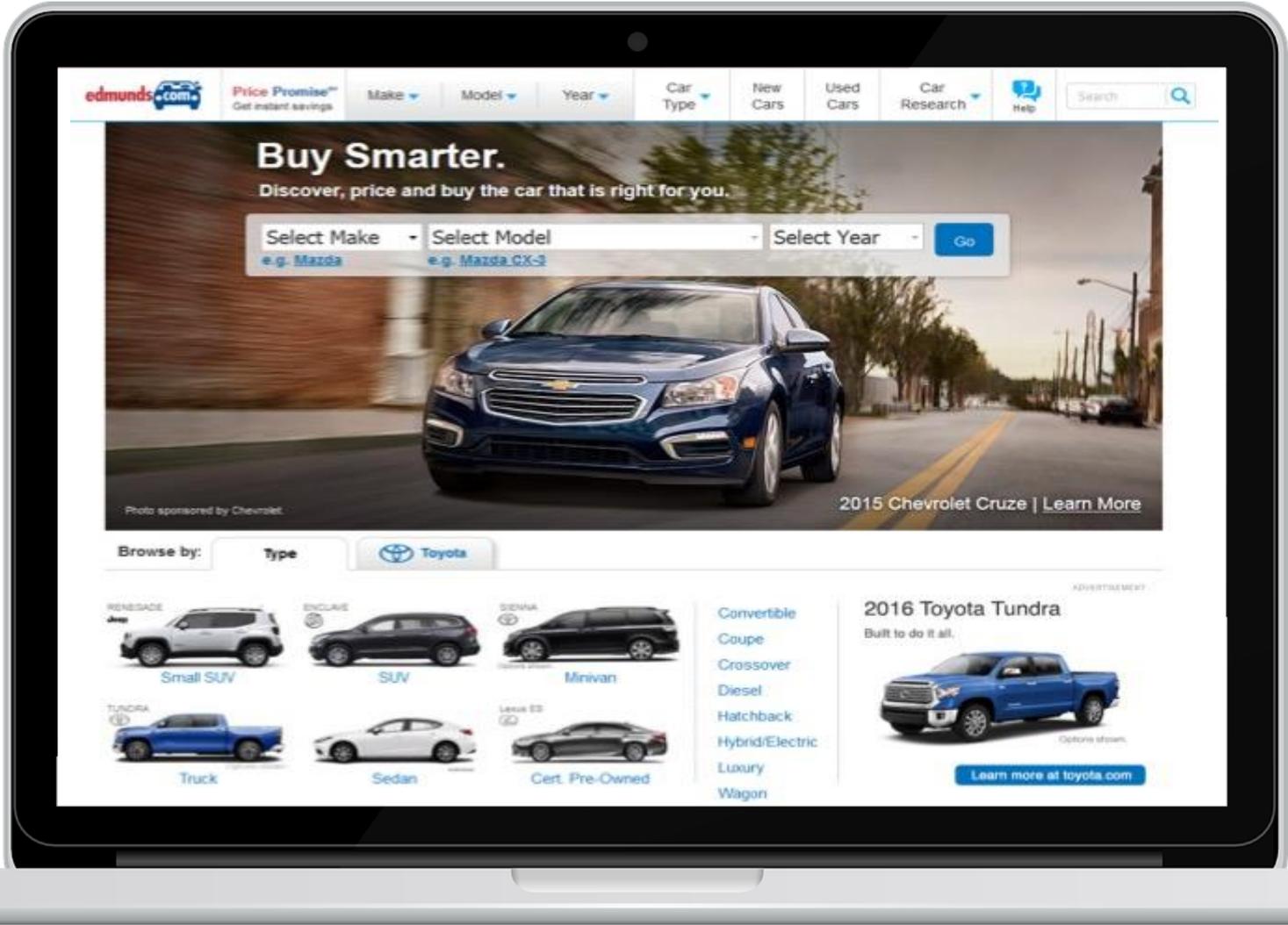
Automotive Websites Top Tasks

Pick Only Five

Automotive Website: 94 Tasks

How many miles per gallon will a car get?
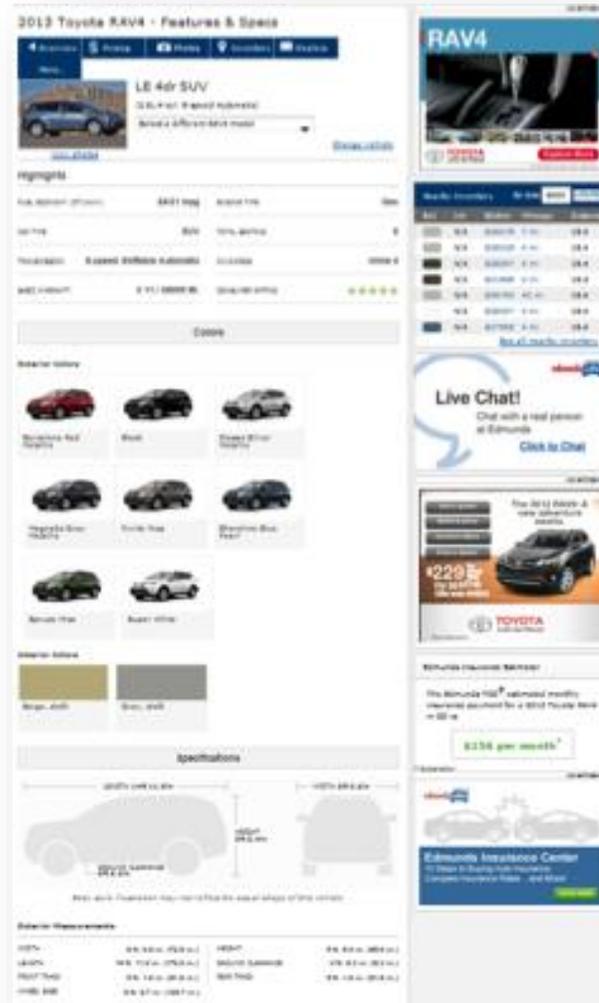
Top Tasks: Health Insurance Website

"Find someone close to my home/office"

"See if my healthcare professional will accept my insurance"

"See if a particular doctor is in my plan"
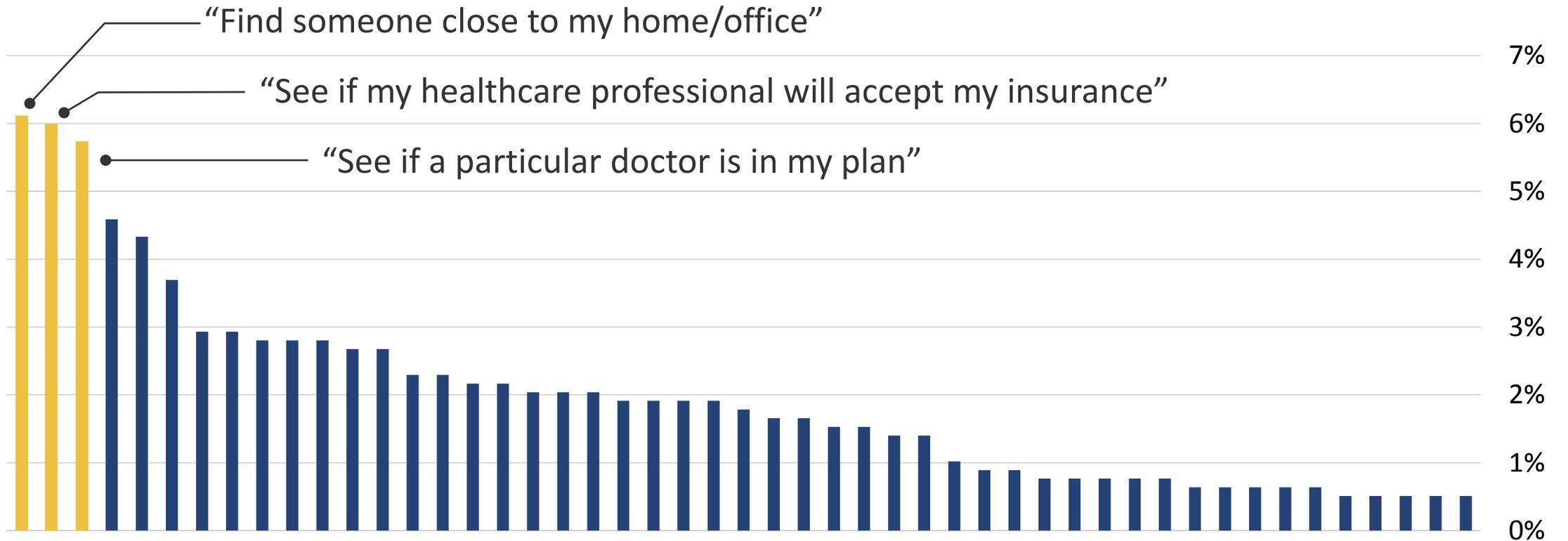
Top Tasks: Cable Provider Portal

35%

30%

Pay Bill

25%

View available bill statements

20%

View account summary

15%

See what is currently playing on TV(s)

10%

5%

0%
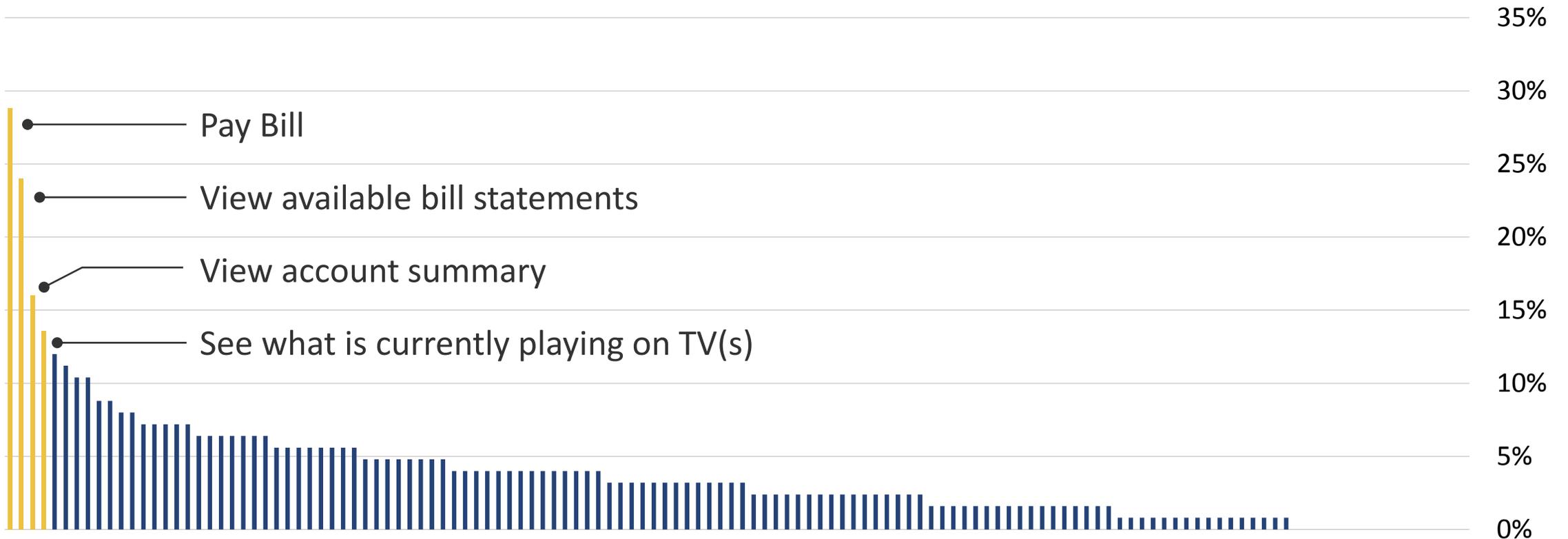
Top Tasks Exercise

Take a top tasks survey

Short URL : https://goo.gl/oHwQ8q

Also on Course Page: measuringu.com/denverux2017/

Task: Fill out The Form

Observe 20 Users Fill Out the Form in a Moderated Usability Test

Analyze

Quantify and assess the experience.

Analyze:

9/20 users (45%) mouse towards the cancel button before submitting (90% CI 28% - 63%).

Synthesize

Interpret the results, think of alternative explanations
and rule them out if you can.

Synthesize

Good Evidence Button Layout Causes Problems

Does a moderated study artificially make users more error-prone?

Redesign of Buttons & Form

Hypothesize

Operationalize

Scientific Thinking

Synthesize

Analyze
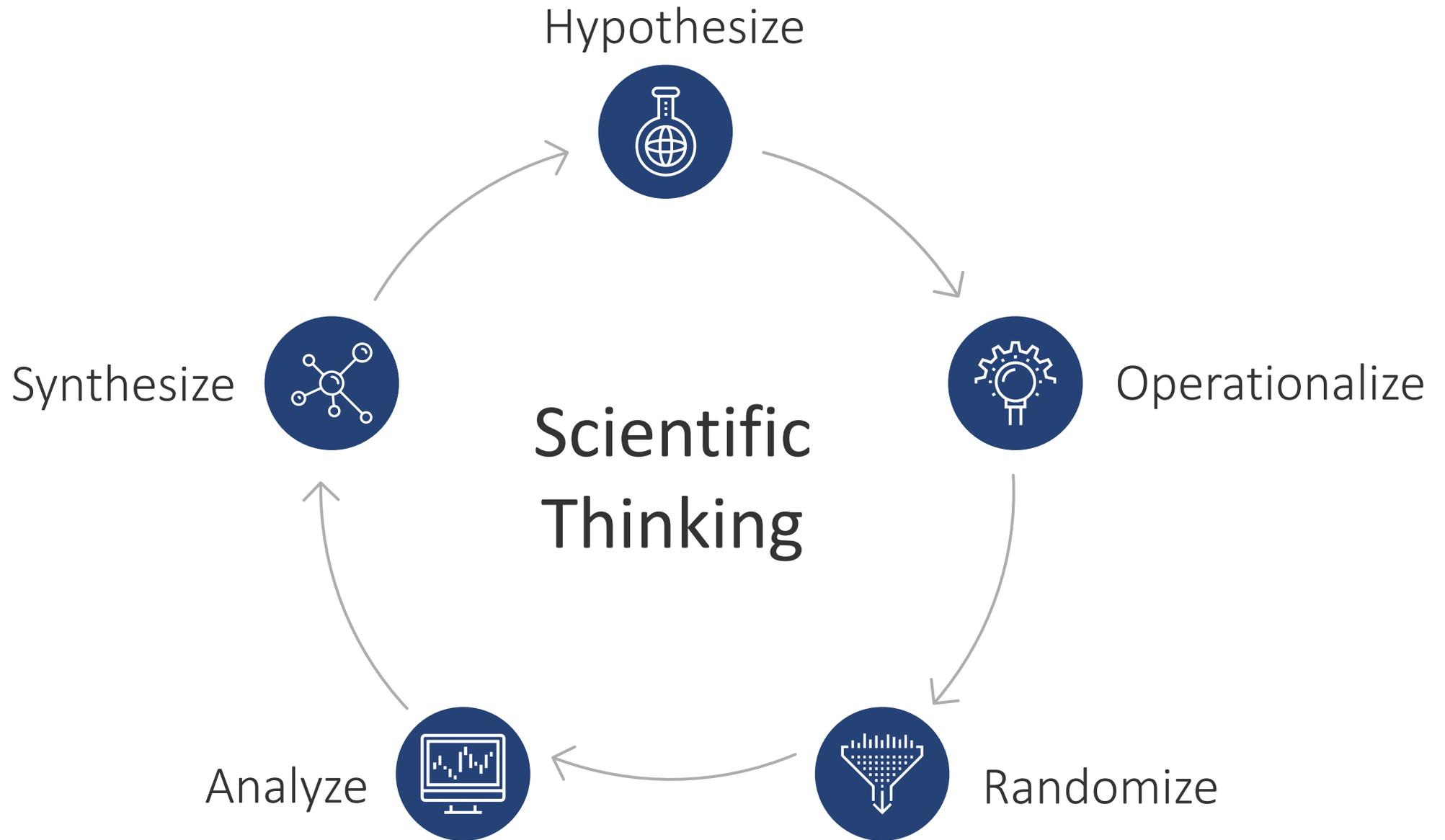
Randomize

Hypothesize & Operationalize

Hypothesize

A new button layout that modifies the color, location and treatment will reduce the number of hesitations.

Operationalize

Metrics	Methods	Tasks
Hesitations	Unmoderated Test	Submit Form
Time	40 Participants	
Errors		
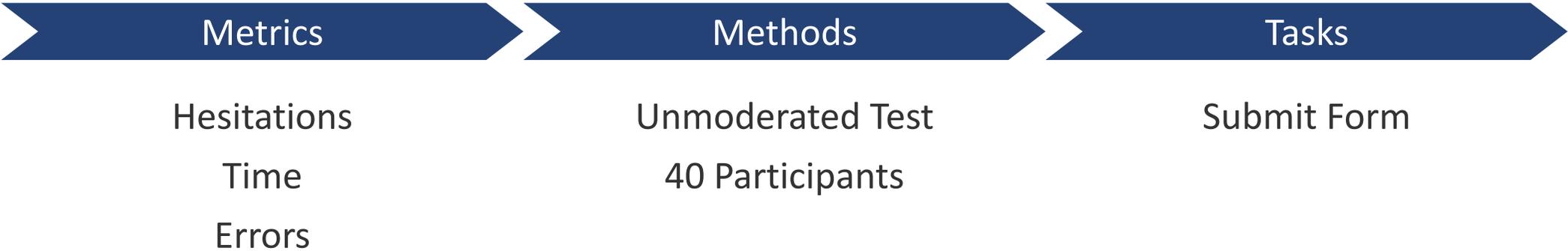
Randomize

Which of the following features would you look for in a house?

A

- ☐ Granite counters
- ☐ Bay windows
- ☐ Finished basement
- ☐ Vaulted ceilings
- ☐ Ranch style floorplan
- ☐ None of these

[Search] Cancel

What kinds of cities are you looking for?

B

- ☐ English speaking
- ☐ Large cities
- ☐ Country getaways
- ☐ Coastal
- ☐ Inland
- ☐ None of these

Cancel [Search]

Which of the following features would you look for in a house?

C

- ☐ Granite counters
- ☐ Bay windows
- ☐ Finished basement
- ☐ Vaulted ceilings
- ☐ Ranch style floorplan
- ☐ None of these

[Search] Cancel

Which of the following car colors are you looking for?

D

- ☐ White
- ☐ Black
- ☐ Blue
- ☐ Green
- ☐ Red
- ☐ Any color

Cancel [Search]

Analyze & Synthesize

Analyze

Number of hesitations was cut by more than half for all forms (p <.05)

D

Which of the following car colors are you looking for?

- [ ] White
- [ ] Black
- [ ] Blue
- [ ] Green
- [ ] Red
- [ ] Any color

Cancel   Search

Synthesize

Better button but other form attributes could affect generalizability of study.

Improving Findability Example

Design Challenge: Retail Browsing
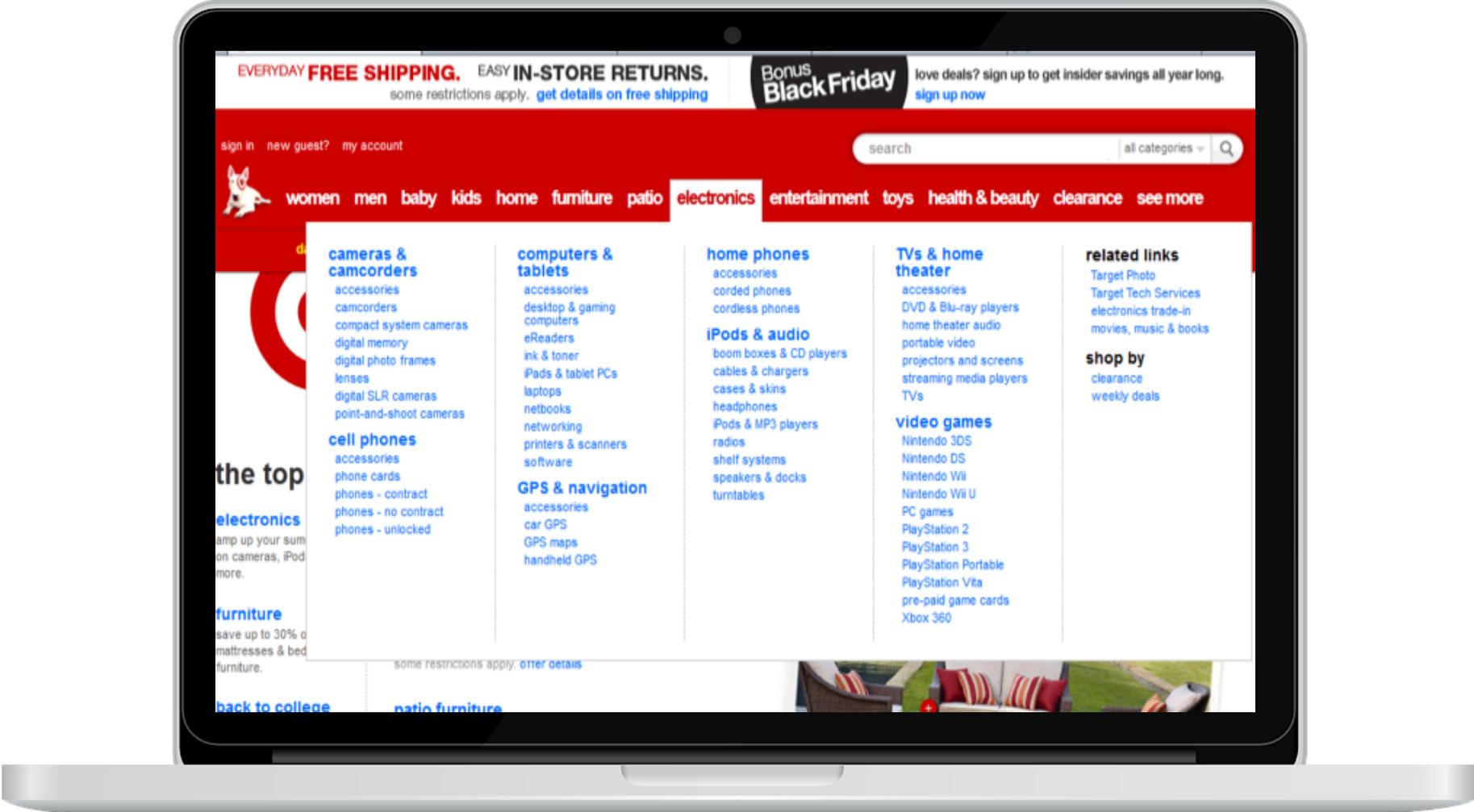
Hypothesize, Operationalize & Randomize

Hypothesize

The navigation inhibits users from finding key products.

Operationalize

Metrics	Methods	Tasks
Findability	Unmoderated Tree Test	Find Select Products
Time	75 Participants	
Ease		

PERSONAL CARE
deodorant
eye care
feminine products
foot care
incontinence
shaving & grooming   Select
SKIN CARE
acne & blemish care

cleansers
exfoliants & masks
facial moisturizers
hair removal
lip care
lotions & creams
sun care
toners
SPA & MASSAGE
accessories
aromatherapy & relaxation

Analyze

Analyze & Synthesize

In Men's

Only In Gifts

In Guys & Juniors

Synthesize: What Are Better Categories?

No one used the term "Juniors" to sort items into

Hypothesize

Operationalize

Randomize

Analyze

Synthesize

Scientific Thinking
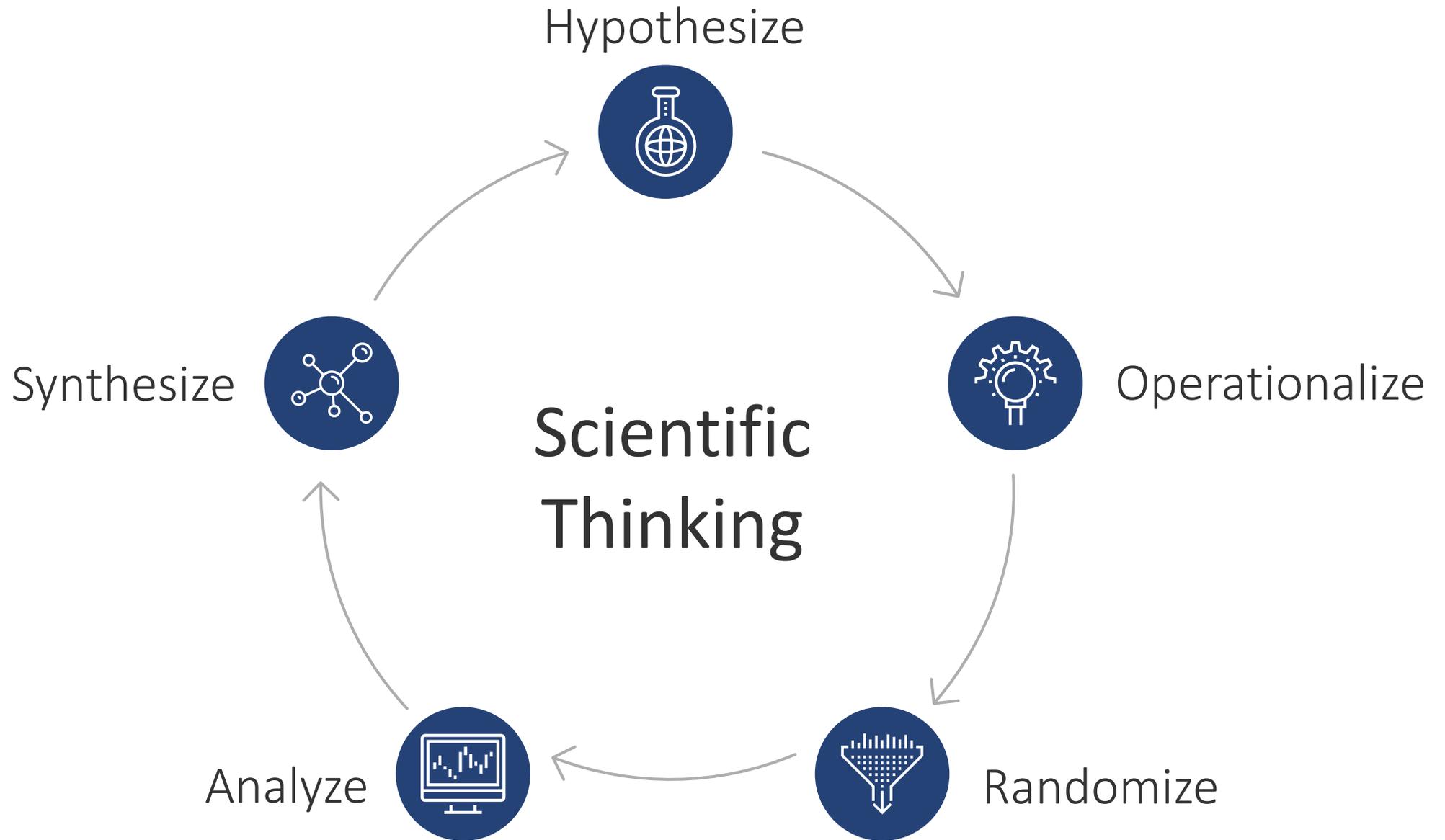
Hypothesize, Operationalize & Randomize
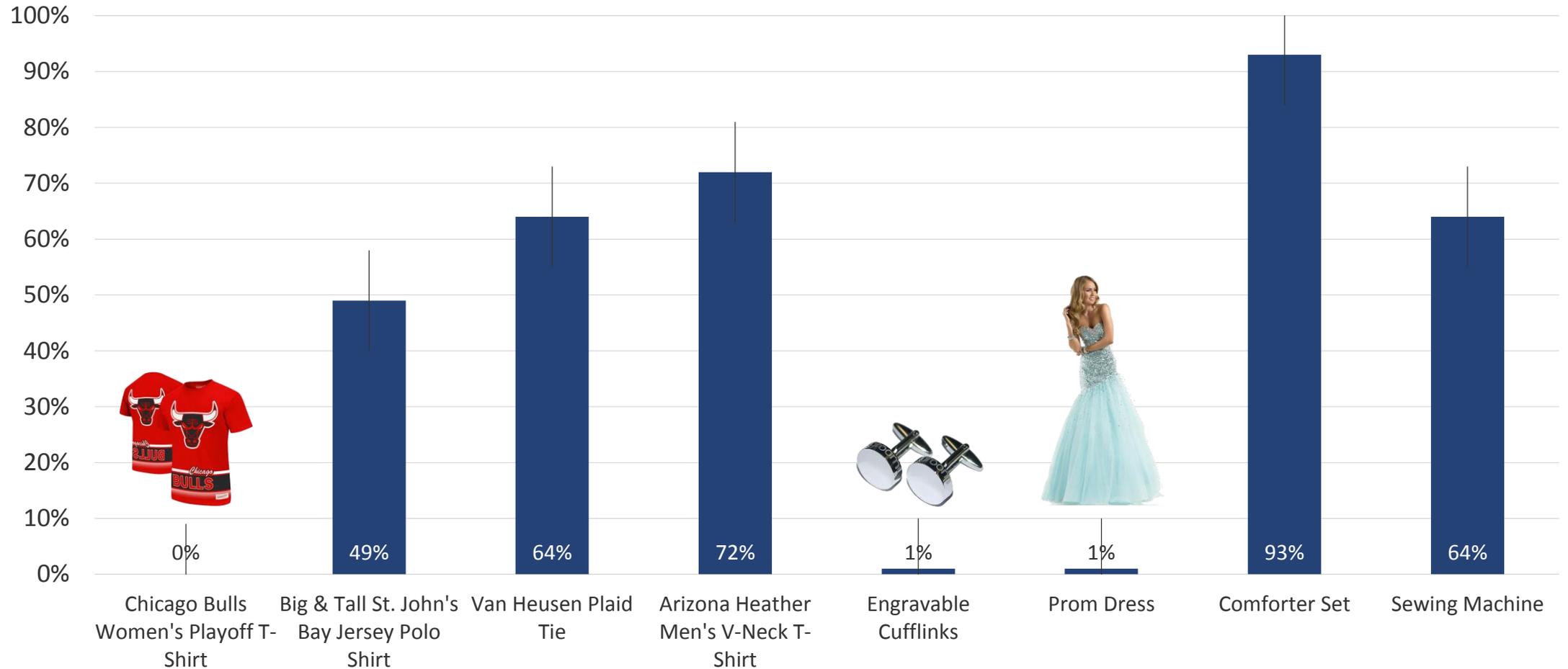
Hypothesize

Navigation changes will improve the findability of key products.

Operationalize

Metrics	Methods	Tasks
Findability Time Ease	Unmoderated Tree Test	Find Select Products

PERSONAL CARE   cleansers

deodorant   exfoliants & masks

eye care   facial moisturizers

feminine products   hair removal

foot care   lip care

incontinence   lotions & creams

shaving & grooming [Select]   sun care

    toners

SKIN CARE   SPA & MASSAGE

acne & blemish care   accessories

    aromatherapy & relaxation

Analyze: Successfully Located Items – Before vs After
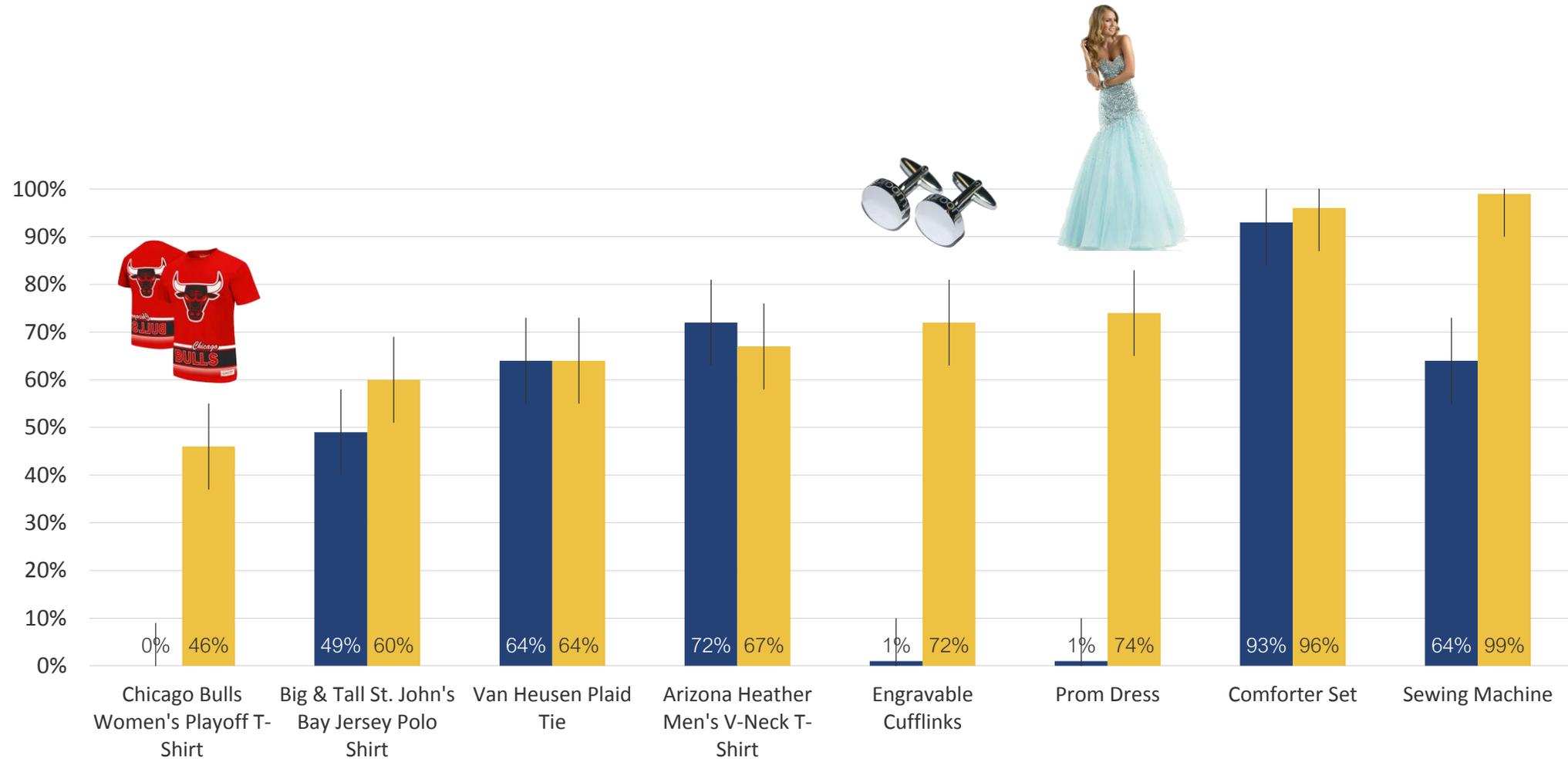
Analyze & Synthesize: Task Ease also Increased
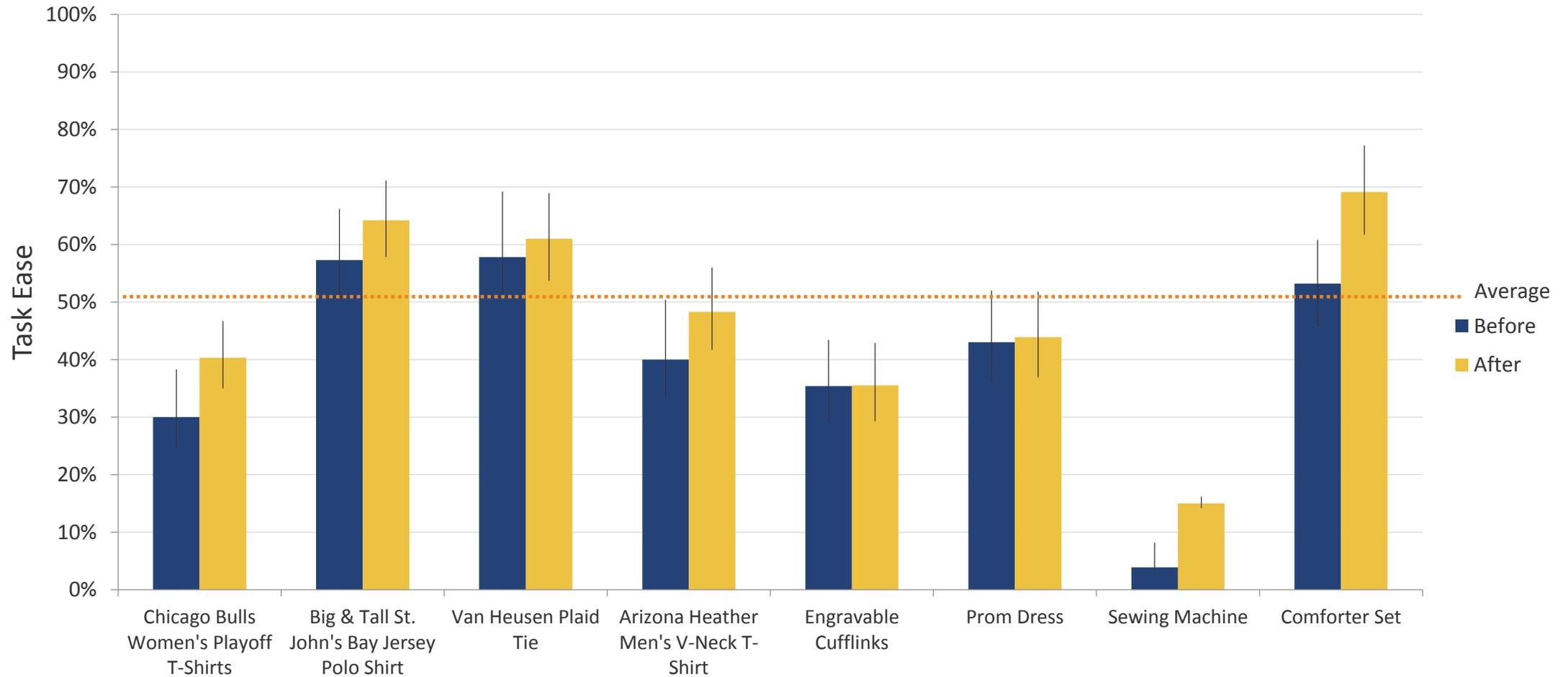
Picking the Right Method

Can users find the function in the software?

Tree Test   Card Sort   Survey   Usability Test   Click Test

Where do users click first on the homepage?

Tree Test   Card Sort   Survey   Usability Test   Click Test

How likely are customers to re-use the product?

Tree Test   Card Sort   Survey   Usability Test   Click Test

Use the Booking Form

Think About:

◎ Goals

💡 Methods

📈 Metrics

📋 Tasks

Project Booking Form

Hypothesize
What is the Hypothesis/Research Question(s)?

Operationalize

1. What type of method are you using?

   a. Are there tasks? If so, what are they.
   b. How will this be tested? [e.g., live site? Product, Prototype? Wireframes?]
      i. Desktop, Mobile (Android & iOS?]

2. What are your metrics (how will success be measured?)?

3. Is this a stand-alone or comparative study?
   a. If this is a comparative study, is it between (different users in each group) or within subjects (same users in each group)?

4. Who are the participants? [e.g., Geo, Membership, Age, Income, State (logged in or out) Tenor, etc.]

   a. Are there distinct subgroups [e.g. Account Holders vs. Prospects?]

5. What sample size will you use and why? Use the tables on the next pages to help with sample size planning.

Group Activity

Scenario 1: You want to know how easy or difficult the current version of the company's expense reporting software is to determine if it needs to be redesigned.

Scenario 2: Call center agents are complaining that it takes too long and is too difficult to fill out caller information. IT wants to investigate more before they invest in a new solution.
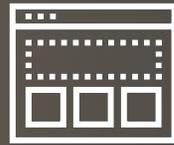
Use the Project Booking Form Online

About MeasuringU

MeasuringU is a quantitative research firm based in Denver, Colorado focusing on quantifying the user experience.
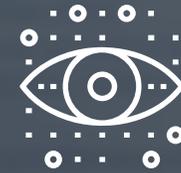
MUIQ

Remote UX Testing Platform
(Desktop & Mobile)

UX Research

SUPR

Measurement
& Statistical Analysis

Eye Tracking & Lab
Based Testing

Denver UX BOOTCAMP

UX Boot Camp Aug 16th-18th
denverux.com

MeasuringU.com
@MeasuringU