



2018

# USABILITY TESTING

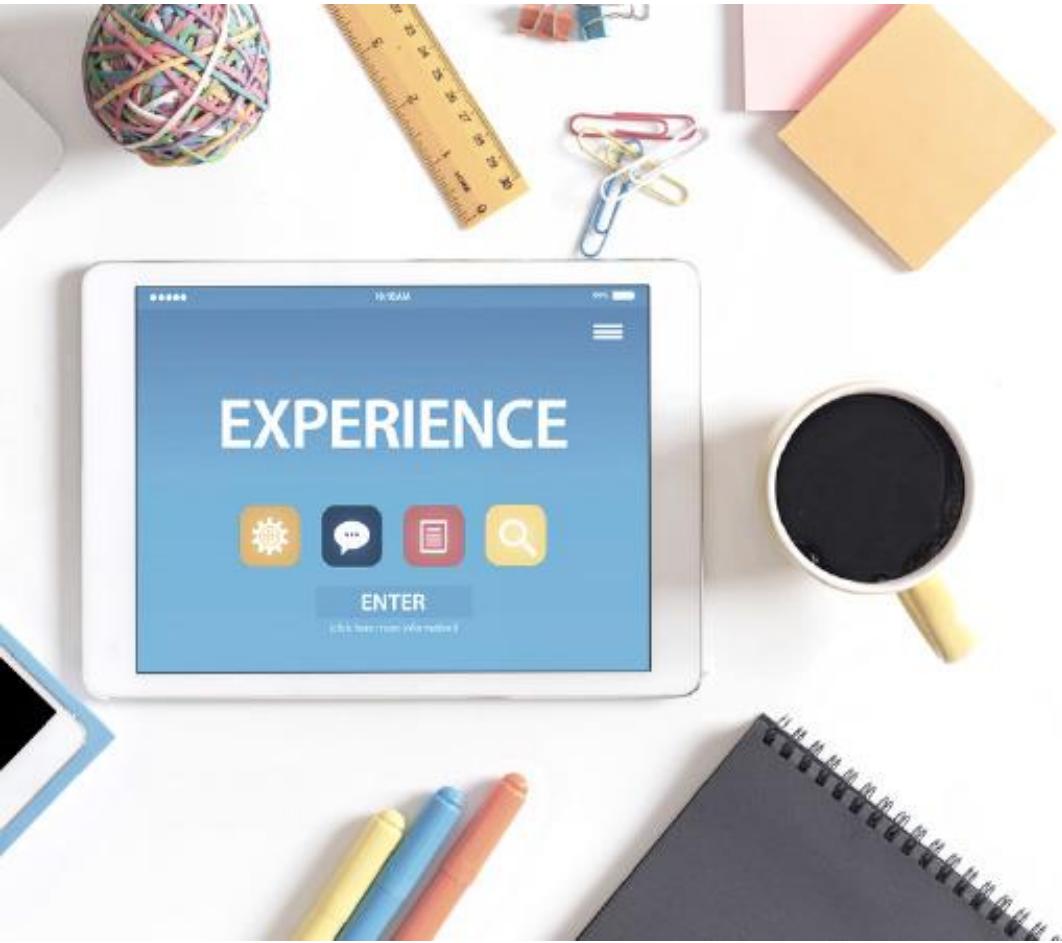
Deep Dive Training

IT | HUE

Honeywell Internal

**Honeywell**  
THE POWER OF CONNECTED

# Agenda



-  Introduction
-  Plan a Usability Test
-  Run a Usability Test
-  Analyze Results
-  Report on a Usability Test

A photograph showing a person's hands working at a desk. A laptop is open on the left, and the person is holding a pen over an open notebook on the right. The background is blurred, showing what appears to be a window with a view of greenery.

Usability Testing

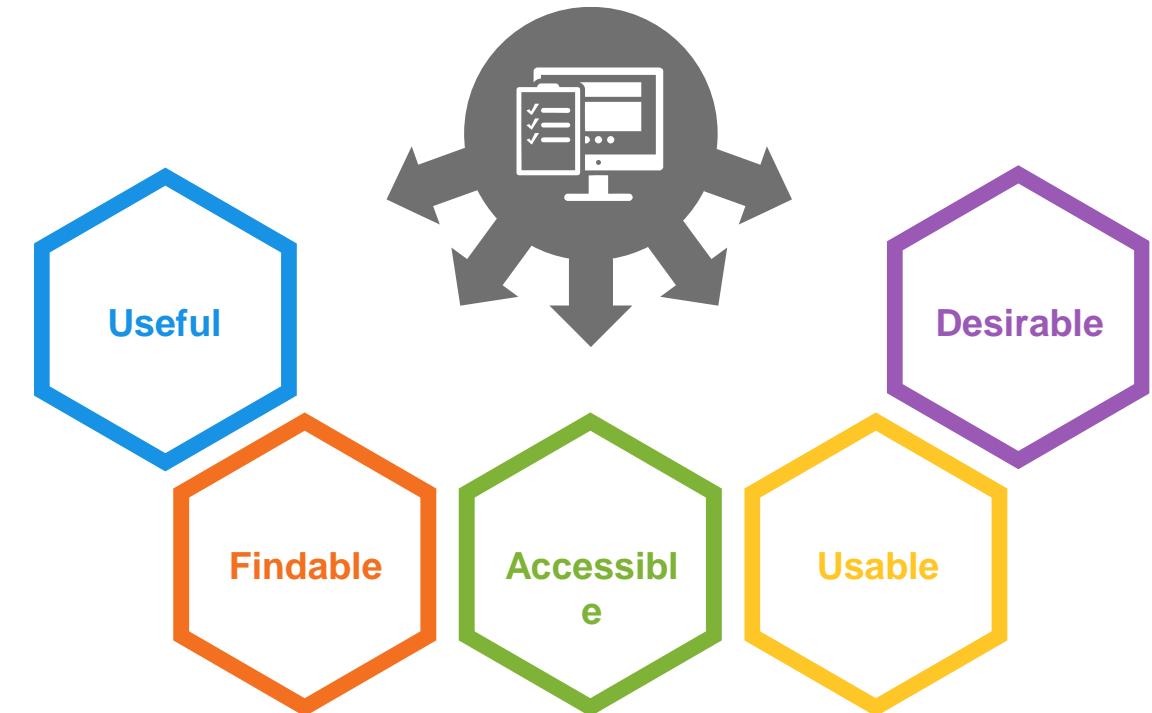
---

## Introduction

# What is Usability Testing

- **Usability testing** is a way to see how easy to use something is by **testing** it with real users
- Users are asked to complete tasks, typically while they are being observed, to see where they encounter problems and experience confusion
- If more people encounter similar problems, recommendations will be made to overcome these usability issues

**Usability Testing**  
Determines whether an application is



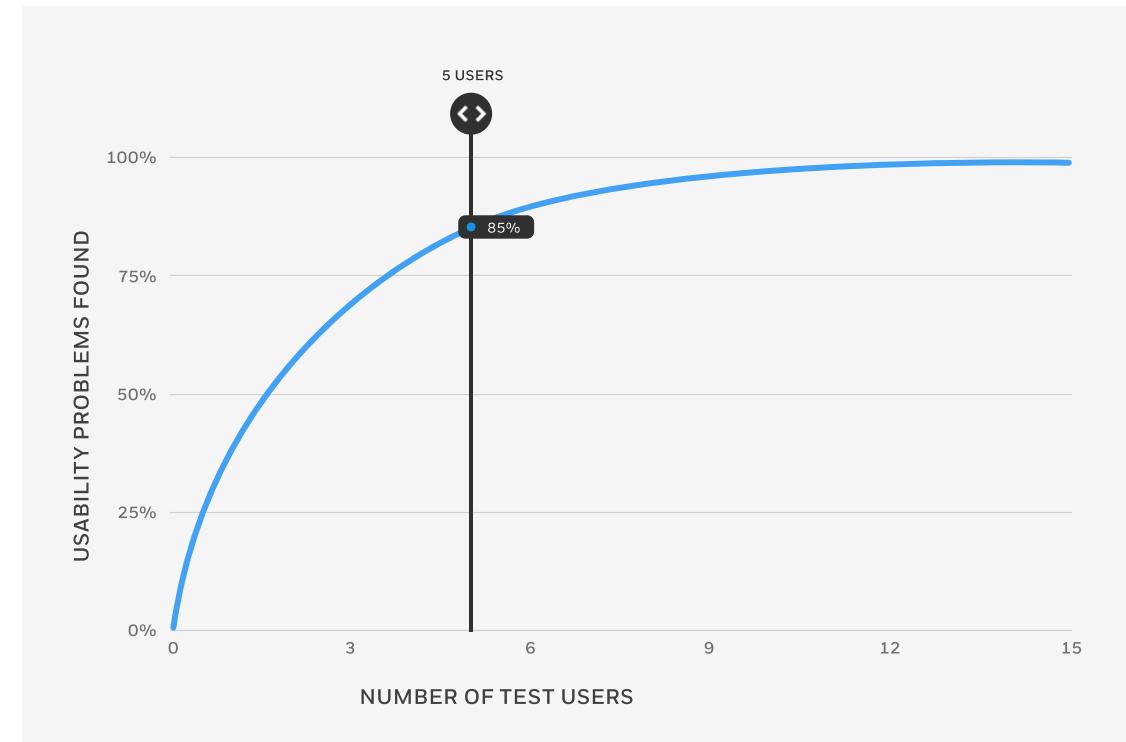
# Why Usability Testing

Testing an offering with potential users to measure ease of use and efficiency by walking through core scenarios as they think aloud.

- Real Users
- Real Tasks
- “Thinking Aloud”
- Prototypes & Real products
- Qualitative & quantitative data
- Lab vs. onsite vs. remote
- Moderated vs. unmoderated

## Research Shows

Testing with 5 users can identify up to 85% of usability issues



## Usability Testing

---

### Plan a Usability Test



# Select Areas to Test

Where is the project in the product life cycle?

## Early

Testing can have a big impact, but make sure you're exploring relevant issues

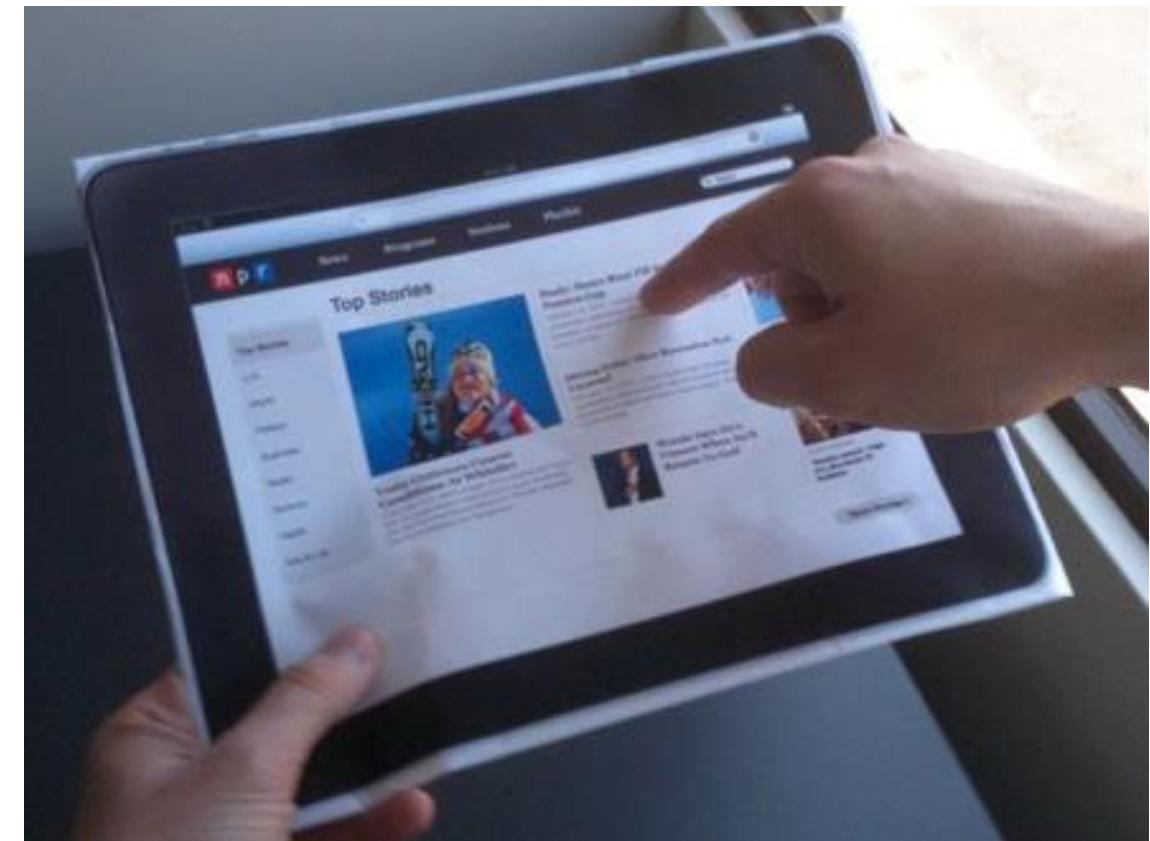
## Middle

Testing can have moderate impact, Fast turnaround important. Choose issues where you'll be listened to

## Late

Be Realistic

1. Consider risks if don't fix enough
2. What can be fixed later
3. Process to not repeat this mess next time

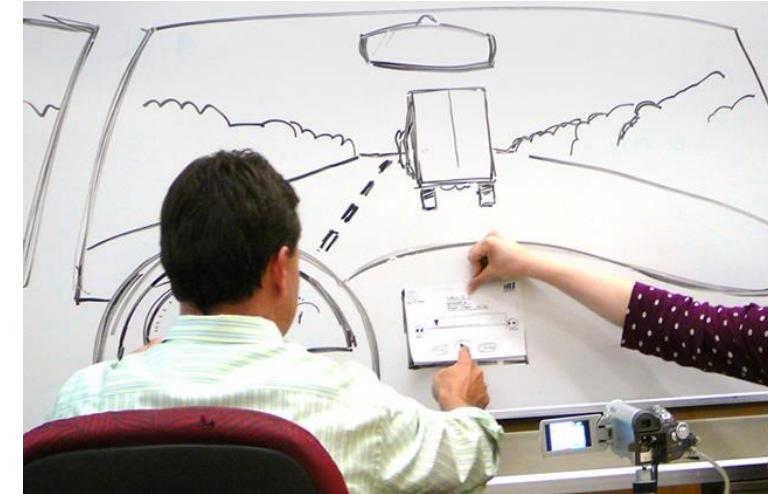


# Select Areas to Test

## What will you test with?

### Paper

- Cheap, usually easy to create, conceptual
- Not authentic experience
- Paper does communicate “conceptual”, room for change
- Best suited for low-complex scenarios



### Powerpoint

- I'm not a “designer” but I can design fairly well in Visio
- Key point: Hyperlinks, so it looks like it's functioning
- More about screens than functionality (won't function well, i.e., fields cannot be filled in)



### Clickable Prototype (InVision, Axure, Adobe XD)

- Hard-coded → Limited data richness

### Production Code

- Authentic, but late in the game for usability testing

# Who to Test

Determine what personas are needed as test participants.

- Age, IT literacy, gender, role, ...all this the persona definition should tell you
- Test each persona separately

## Who to recruit

- Professional recruiters: For many types of consumers, and some target audiences (like MDs)
- Lists of customers, you contact yourself
- Referrals from sales staff, or from technicians who know their customers
- Subscriber lists, Professional Society members (especially valuable if want special expertise)



## Develop Test

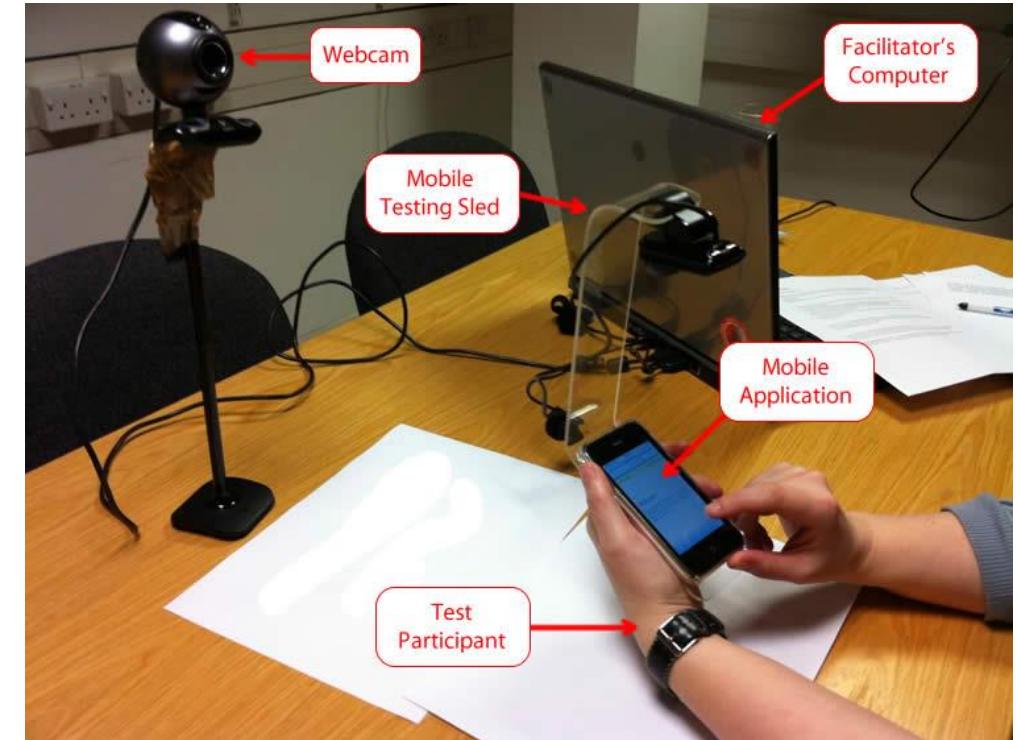
Develop test scenarios and metrics to assess, and update the usability test booklet accordingly

### Examples for metrics:

- Binary task success
- Task execution time
- Errors
- Help provided by moderator

### How much time per session?

- If testing in a public place keep sessions really short (5-10 minutes)
- Remote tests should be 15-30 minutes
- Usually hard to go longer than 90 minutes



# Develop Test

## What are the research questions?

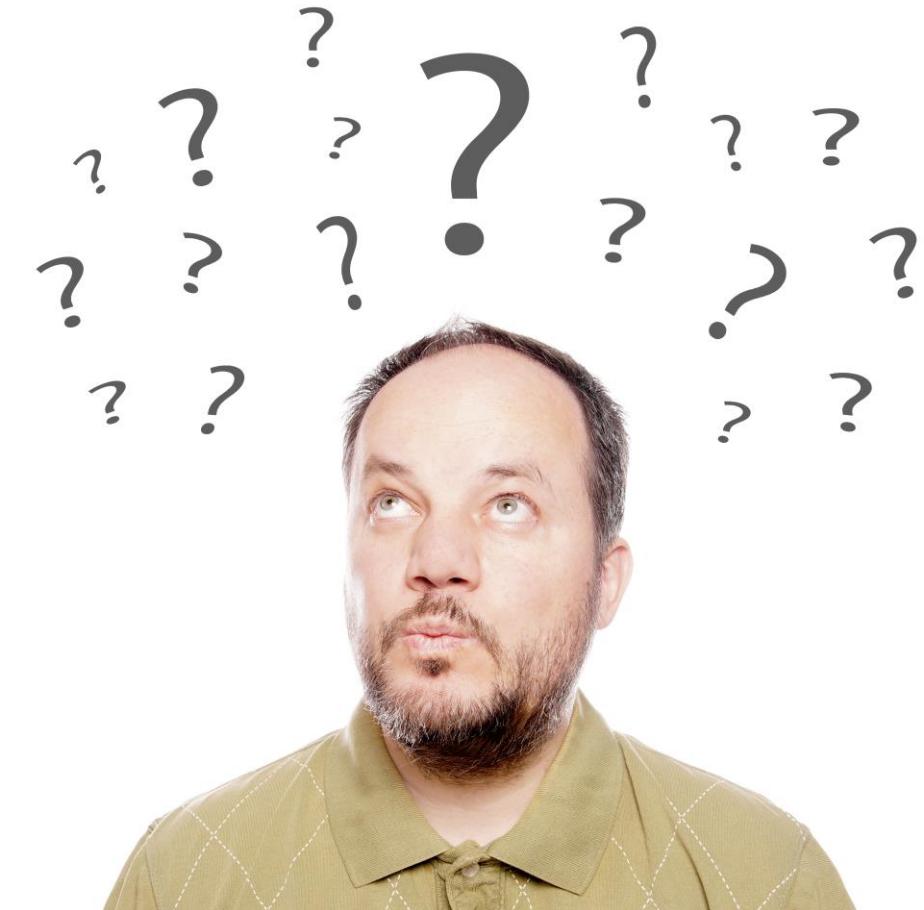
- What do we know?
- What don't we know? Are there other methods to find answers, e.g. interviews, surveys?
- Research questions are not tasks

## Research:

- What are biggest problems people have placing orders?
- Will radiologists be able to efficiently diagnose cases, without errors?
- Will consumers like this better than their current product

## Tasks:

- How long will it take customers to find the due date for their current statement?
- Find any gas stations, with "Regular" gas under \$2/gallon, within about 2 miles



# Recruit Participants

## How many participants to run?

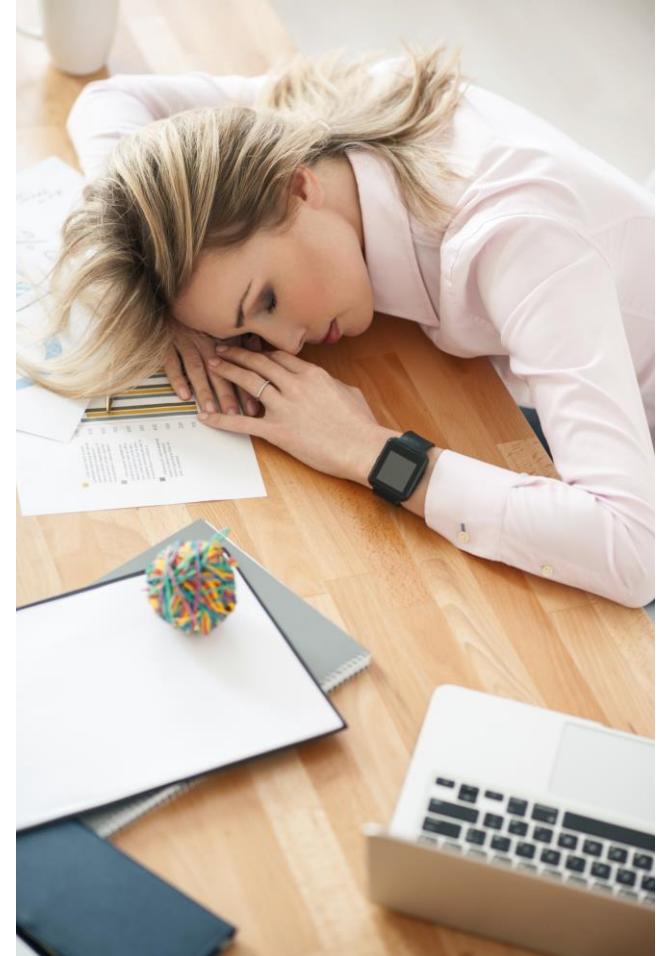
- 5 or a few more. Go higher if test has truly different groups
- Consider more rounds of testing if prototyping and time are available
- Bump up numbers if only one round of testing
  - Shoot for at least 3 people per demographic/persona

## Run how many sessions per day?

- Don't make testing days too hard on anyone (including the moderator)
- Breaks between sessions gives flexibility
  - 60 minute sessions: up to 5 per day (breaks of at least 30 min)
  - 90 minute sessions: up to 4 per day (breaks of at least 30 min)
- Schedule lightly on first testing day, to allow updating of everything, and to allow for the unexpected

## Forms for participants

- Informed consent form
- Media Release
- When in doubt, check with legal!



# Testing Logistics

## What will you record?

- Screen – see what happened
- Audio – hear what happened
- Very useful – The participant's face, to see emotions
- Also useful – hands on keyboard, on artifacts, etc

## Moderator Guide

- Introduction
- Why you're there
- This is usability so
  - Please speak aloud
  - Testing UI, not you
  - Don't worry about errors
  - Not my design, pointing out issues will help me

Book Room	<input type="checkbox"/>
Invite Participants	<input checked="" type="checkbox"/>
Staff Roles	<input type="checkbox"/>
Recordings	<input type="checkbox"/>
User Testing Software	<input type="checkbox"/>



# Testing Logistics

## Moderator Guide (continued)

- Ask Questions
- Can stop at any time, for any reason
- Plan the task list
- Think about time. Consider optional tasks
  - Have an extra task or two prepared
- Have probing questions ready and suggested things to observe
- Questions to Ask:
  - Per task: Make sure to ask paraphrase of task
- Debriefing questions
  - What do you like best
  - What did you like least
  - Any additional comments
- SUS [System Usability Scale]

### Usability Testing Facilitation – 45-Minute Training Session

1. Review Excel template – **10 min**
  - Using the Task Timer
    - Start, pause, stop, select recorded before resetting, then move to next task
  - System Usability Scale
    - Scale of 1 to 5 – the higher the rating, the higher the participants agreement with the statement
  - How to edit Participants in first tab (if necessary)
  - How we'll aggregate data across sessions of the same Transaction type (copy/paste up to 20)
    - Be sure to upload Facilitators Guides to SharePoint once sessions are completed.
2. Review Email to be sent day of testing, 1-2 hours prior – **5 min**
  - Attach the right Tasks to each email
3. Usability testing session demo – **15 min**
  - Let's watch a video from our Global Compensation Planner testing
    - Usability Testing Facilitation - 30-Minute Training Session.mp4
  - We may pause at the following times to talk about a few tips:
    - 02:14 – You may need to prompt to move participant to next task.**
      - Sometimes a user will clearly state they have completed, sometimes they won't.
      - It is okay to ask the participant, "Given the completion criteria for this task, how are you doing?" We would never say, "You have completed the task, right?"
      - This is especially helpful if the participant has been on the task for a long time and is expressing some frustration.
    - 02:45 – Prompt for open-ended feedback, at the right moments.**
      - Only prompt once task is noted as Complete by the user
      - Use open-ended prompts, such as "I heard you say...tell me more about that."
      - Always stop the timer when side conversation occurs
    - 05:40 – Sometimes a user will re-read a task, or part of a task.**
      - Pause the timer when a user re-reads, and restart timer when they resume the task
    - 09:32 – You may realize the participant needs you to clarify a task.**
      - If a user seems to be floundering or seems unclear, it is okay to pause the timer and have the user reference the task to ensure the completion criteria is clear.
  - **JUMP TO 15:00 – System Usability Scale Prompts**
    - Read them, listen and record the participant's rating and comments
    - Prompt for more if a user gives an extreme score of 1 or 5, especially if their rating seems out of sync with the way they reacted or commented during the tasks
    - It can be helpful to paste the rating scale into the Lync chat, for the participant to reference
      - ✓ 1=Strongly Disagree, 5=Strongly Agree
    - **16:10 – Tell me more about that...**
      - If there is time left in the session, the SUS prompts are a great opportunity to gather more open-ended feedback. Say, "You mentioned...tell me more about that."
4. 5 common situations that arise during testing – **10 min**
  - A participant asks you how to complete a task or asks for other help.
  - A participant gets interrupted during the session (phone call, Lync, Etc).
  - A participant gets very frustrated during a task, and expresses this.
  - A system error prevents the participant from completing a task.
  - A participant is on an early task and has already used 1/4 of the session time to try to complete it.
5. Questions?

A facilitator's job is to impartially observe, record and facilitate, with golden silence, to gather unbiased feedback.

Usability Testing

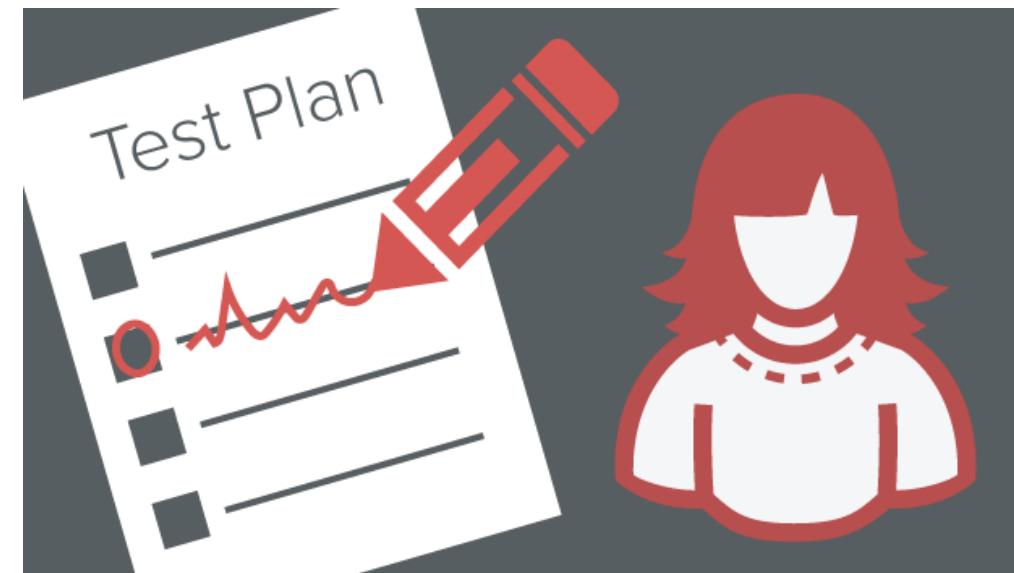
Run a Usability Test



## Pilot the Test

Run pilot test to ensure proper functioning of testing procedure

- Pilot testing (a session of two before the real test) helps fin-tune usability studies, leading to more reliable results
- It provides an opportunity to validate the wording of the tasks, understand the time necessary for the session, and may even supply an additional data point for your study
- To get the most out of a pilot session, **schedule it at least 1 day in advance** of the scheduled study



# Usability Testing

## Typical usability issues uncovered

- Anything that prevents task completion
- Anything that takes someone “off course”
- Anything that creates some level of confusion
- Anything that produces an error
- Not seeing something that should be noticed
- Assuming something is correct when it is not
- Assuming a task is complete when it is not
- Performing the wrong action
- Misinterpreting some piece of content
- Not understanding the navigation

## What is looked at:

- User sentiment from thinking aloud, body language, mimic
- Task success rates
- Task execution times
- Number of errors
- Number of time users asked for help
- Answers to de-briefing questions and questionnaires (e.g. System Usability Scale - SUS)

## Execute

- Welcome participant
- Introduce yourself and the project
- Have them Sign NDA and consent forms
- Explain test procedure
- Go through test scenarios, reminding users to verbalize thoughts
- Apply post-test questionnaires **SUS**
- Thank participant for joining

### Taking notes during session:

- Excel template for all key questions
- Take notes with computer
- Take notes at a constant pace
- Note-taker can be great but everyone looks for different things



## Execute

### What to do if:

- Participant no-show
- If remote, can't get to internet
- Prototype fails
- Computer fails
- Participant arrives quite late
- Running very behind schedule

### After a test

- Debriefs are great. Consider recording debriefs if you didn't take good notes
- Write down first impressions of results while they're fresh in your mind
- Push self for rapid turn-around





Usability Testing

Analyze Results

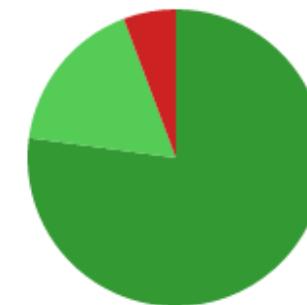
# Performance Metrics

- Every type of user behavior is measurable in some way
- Performance metrics rely on the use of scenarios or tasks. With the help of tasks we can gather quantitative data on such items as:
  - Task success
  - Time on Task
  - Errors
  - Efficiency
  - Learnability
- Performance metrics are one of the most valuable tools used in usability evaluation

## Task 7

You've accidentally 'friended' someone, and you'd like to cancel your connection to them.

Using Yelp → Friends and fans → How do I unfriend someone?



<span style="color: green;">■</span> Direct Success	27	77%
<span style="color: lightgreen;">■</span> Indirect Success	6	17%
<span style="color: red;">■</span> Failure	2	6%
<span style="color: black;">■</span> Skip	0	0%

[View the Pietree for this task](#)

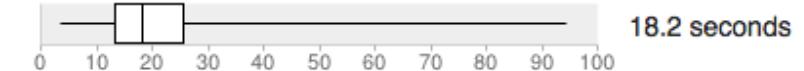
⊕ Success



⊕ Directness



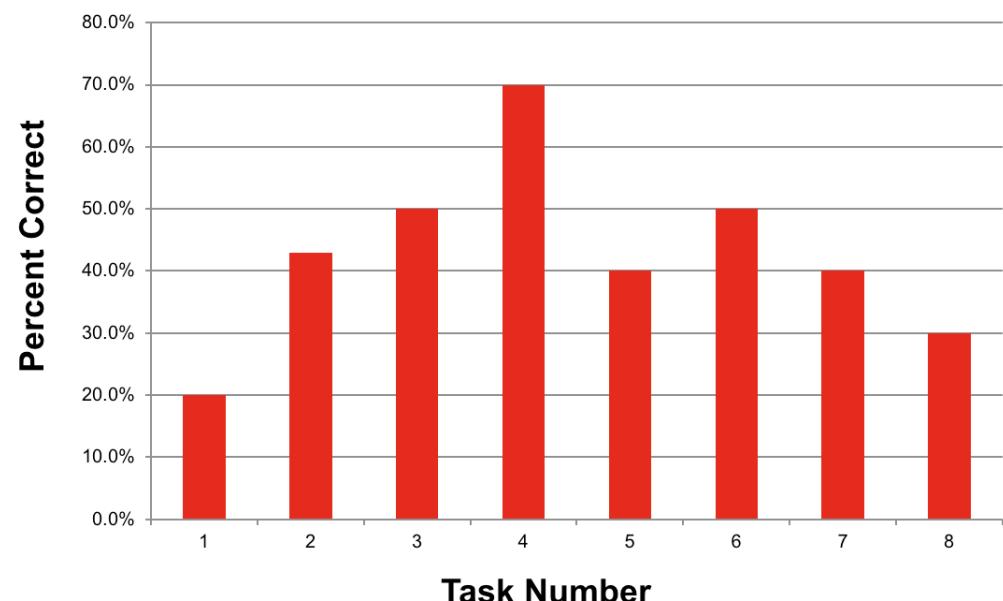
⊕ Time Taken



# Task Success

Measures how effectively users are able to complete a given set of tasks

- Perhaps the most widely used performance metric
- What is needed:
  - Well defined tasks
  - Well defined definition of task success
- Two methods of presenting success measures
  - Binary success
  - Levels of success
- Binary Success
  - Participants complete the task successfully or they do not
  - Zero for user who does not complete the task and a one for a user who completes the task
- Typically presented in a bar graph



# Task Success

Levels of Success can be dependent on the extent to which the participant finished the task

- Complete Success
- Partial Success
- Failure
- Various methods for scoring success
  - Complete Success = 1.0
  - Partial Success = 0.5
  - Gives up or wrong answer = 0.0
- Key issue in measuring level of success is defining what is meant by each level before the study has started
- The point is to be consistent in one's judgement for each participant

## Four point scoring system for levels of success\*

### **0 = no problem**

The participant successfully completed the task without any difficulty or inefficiency

### **1 = cosmetic problem**

Need not be fixed unless extra time is available on project

### **2 = minor problem**

The participant successfully completed the task but took a slight detour. He or she made one or two small mistakes but quickly recovered and was successful

### **3 = major problem**

The participant successfully completed the task but had a major problem. He or she struggled and took a major detour in eventual successful completion of the task

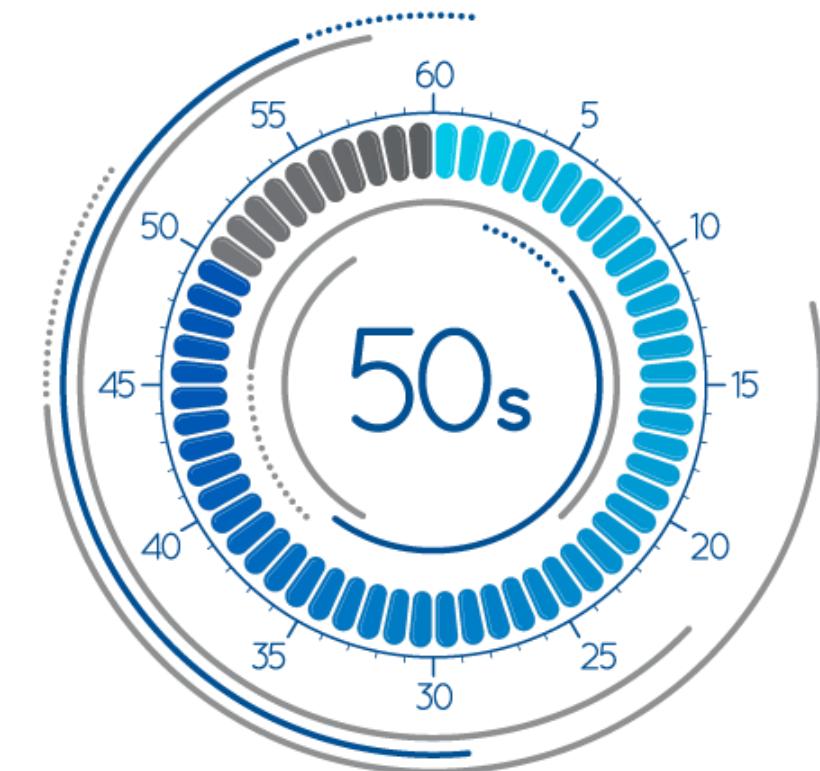
### **4 = failure / gave up**

The participant provided the wrong answer or gave up before completing the task, or the moderator moved on to the next task

\* Adapted from "Severity Ratings" (from Nielsen & Mack '94)

# Time on Task

- The amount of time spent on a task is important for tasks that are performed repeatedly
- Measuring time on task
- Analyzing time on task data
  - Average time to complete task
  - Range of times taken to complete task
  - Percentage of tasks exceeding or falling below a threshold
- Time on task will vary because:
  - Participants may be having trouble completing the task
  - Participants have been asked to “think aloud” slowing down the process of doing the task
  - Participants will respond to the study differently
    - Some may finish as quickly as possible
    - Others may be as careful as possible



# Metrics: Errors

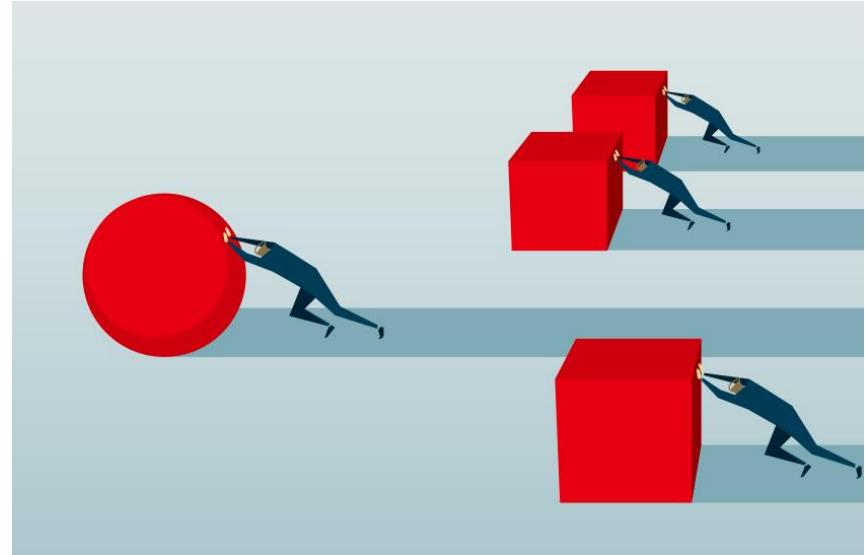
- Types of errors
  - Slips
  - Mistakes
  - User Interface problems
  - Scenario Errors
- Analyzing and presenting errors for tasks with a **single** possibility for error
  - Plot the frequency of errors for each task
  - Divide the number of errors by the number of participants for each task
  - Average the error rates for each task into a single error rate
  - Average all the tasks that had a certain number of errors
  - Determine which tasks have an error rate above a set threshold



- Analyzing and presenting errors for tasks with **multiple** possibilities for error
  - Divide the total number of errors by the number of error opportunities.
  - Calculate the average number of errors made by each participant for each task
  - Determine how many tasks fall above a threshold error rate
  - Code errors as: (produce an average for each)
    - 1 = Trivial
    - 2 = Moderate
    - 3 = Serious

# Metrics: Efficiency

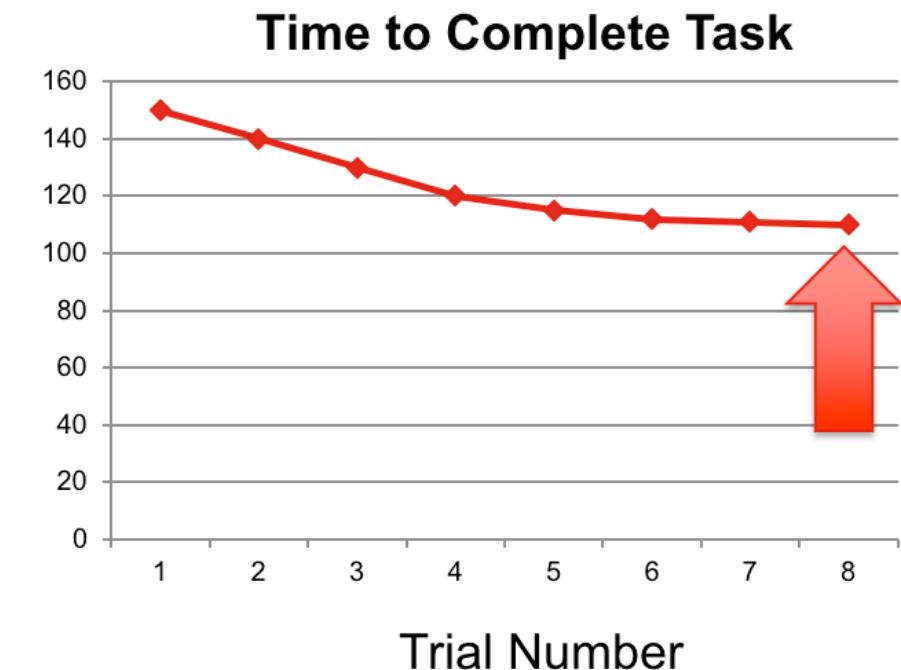
- The resources expended in relation to the accuracy and completeness with which users achieve goals
- Efficiency is often measured as the amount of time taken to complete a task
- Collecting and Measuring Efficiency
  - Identify actions to be measured
  - Define start and end of an action
  - Count the actions
  - Make sure the actions are meaningful
  - Examine only tasks that are successful



- **Percent efficiency** = Min number of actions to achieve task divided by the # of actions taken
- **Efficiency** = # of tasks successfully completed divided by the total time spent on all tasks
  - Number of tasks completed per minute

# Metrics: Learnability

- Learnability is a measure of how much time and effort it takes for someone to become proficient with something
- How to collect and measure learnability data:
  - Hold multiple usability sessions spaced by a number of days
  - Run trials all at once but separated by a distractor task
  - Run trials one after the other with no breaks
  - Look at time on task, errors or number of steps
  - **Plot data for each trial and look at the slope**
- People don't learn by individual trials, they learn continuously
- People also learn in leaps and bounds
  - Something finally "**gels**" with them



# Issue Based Findings: Usability

- Usability Issues
  - Usually qualitative
  - Description of an observed problem encountered by one or more participants that was observed during the usability study
  - Usually found via the participant "thinking aloud"
  - Can often be an unexpected surprise

**Positive finding are important to report so that redesigns do not “break” the interface in the future**

- Identifying Usability Issues
  - In person studies
    - Verbal expressions
    - Participants not saying or doing something they should have done
    - Nonverbal behaviors such as facial expressions
  - Automated/Remote studies
    - Verbatim comments required by participant at the end of page or task
    - Feedback requirements for low ease of use scores



# Issue Based Findings: Severity Ratings

Severity ratings focus attention on issues that really matter

## Low

Any issue that annoys or frustrates participants but does not play a role in the task failure

## Medium

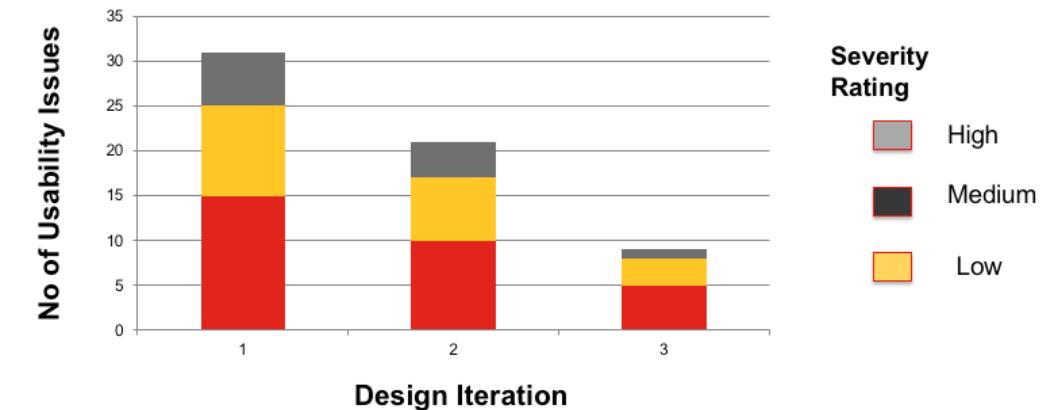
Any issue that contributes to but does not directly prevent task completion

## High

Any issue that directly leads to task failure

Another category suggested by Nielsen:

**4 – CATASTROPHIC** – Product cannot be launched unless fixed



- 0** - don't agree that this is a usability problem
- 1** - cosmetic problem
- 2** - minor usability problem
- 3** - major usability problem; important to fix
- 4** - usability catastrophe; imperative to fix

# Recommendations

Derive conclusions and recommendations from results and document in report, include:

- Metrics
- Demographic Data
- UX Strengths
- UX weaknesses with severity ratings and mitigation direction
- SUS interpretation

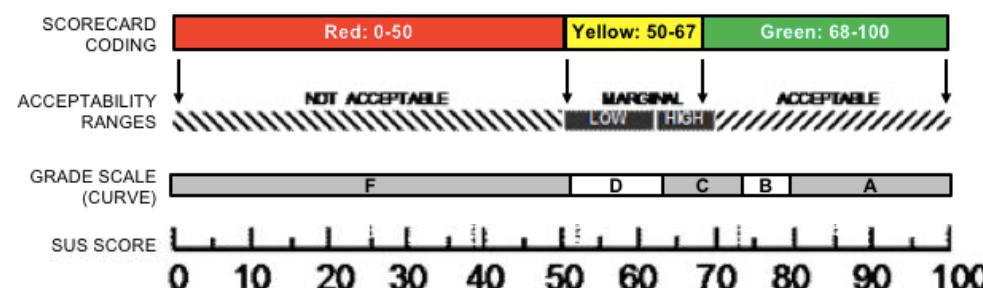
**Affinity sorting** is an effective tool for organizing ideas and data. Findings are grouped against common themes/trends. This helps show what users needs and wants are, and why; which helps to brainstorm effective recommendations.



# SUS Evaluation Process

SUS = System Usability Scale

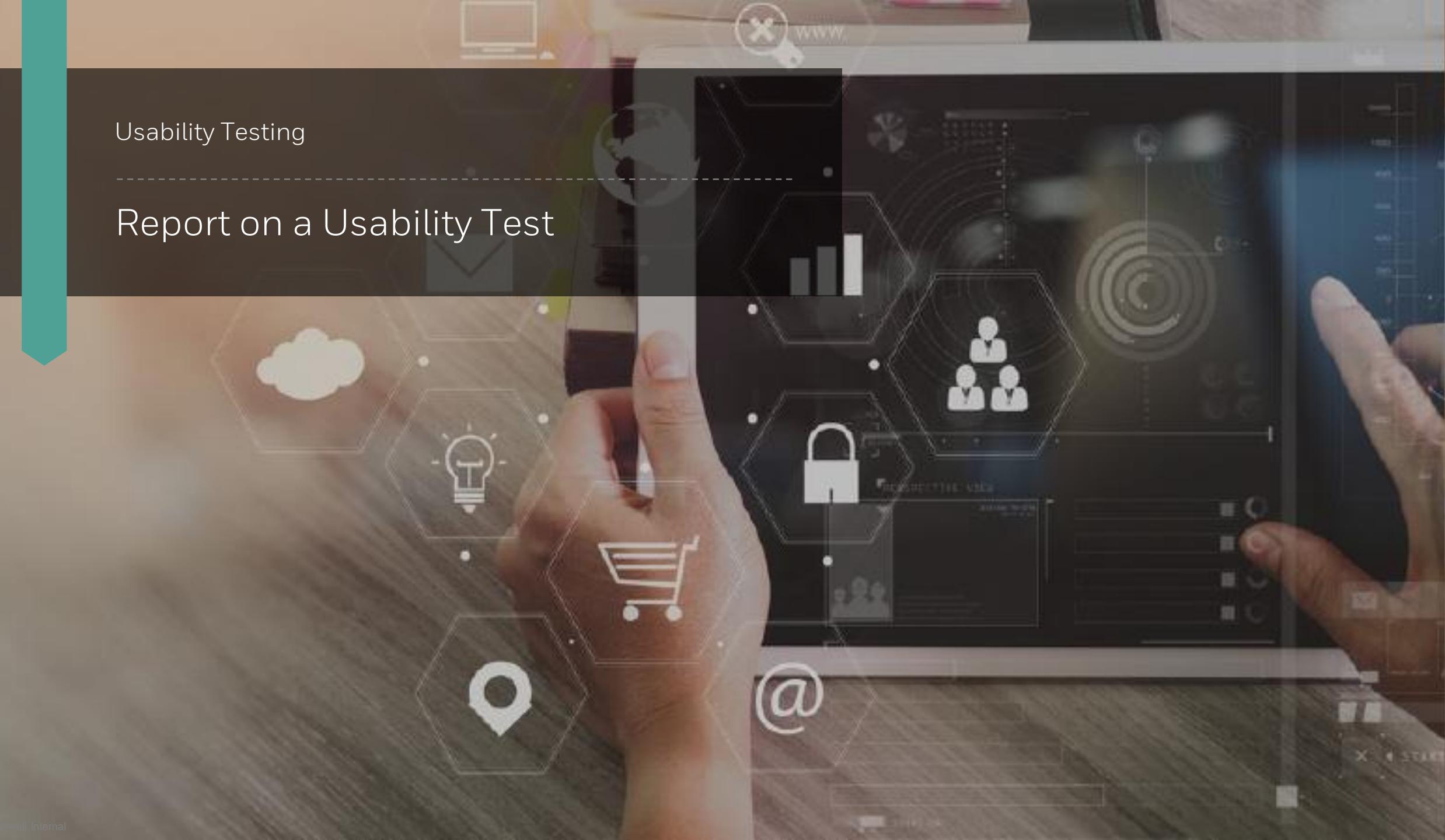
- Focusing on users' experiential attitudes towards an interactive system
- The SUS Question spreadsheet is distributed to users, filled out, and collected immediately after the testing
- The results are then entered into the SUS Calculation workbook and scores are automatically calculated
- SUS scores may tell us that something is wrong – but not what is wrong and how it can be fixed



Participant ID: _____	Site: _____	Date: ____/____/____					
<b>System Usability Scale</b>							
<b>Instructions:</b> For each of the following statements, mark <u>one</u> box that best describes your experience with the software today.							
	Strongly Disagree	1	2	3	4	5	Strongly Agree
1. I think that I would like to use this software frequently.	<input type="checkbox"/>						
2. I found the software to be simple.	<input type="checkbox"/>						
3. I thought this system was easy to use.	<input type="checkbox"/>						
4. I think that I could use this software without assistance.	<input type="checkbox"/>						
5. I found the various functions in this software were well integrated.	<input type="checkbox"/>						
6. I thought there was a lot of consistency in this software.	<input type="checkbox"/>						
7. I would imagine that most people would learn to use this software very quickly.	<input type="checkbox"/>						
8. I found this software very intuitive.	<input type="checkbox"/>						
9. I felt very confident using this software.	<input type="checkbox"/>						
10. I could use this software without having to learn anything new.	<input type="checkbox"/>						

Usability Testing

## Report on a Usability Test



# Review Results

## Review results with project team and determine action items

- Write up issues on sticky notes and sort them into priority lists
- Ask for top 10 items – base on data and observations, not opinion
- Take a vote for top priorities



# Report Out

- Objectives and Goals
- Executive Summary
- Detailed Summary
- Participants
- Scenarios & Tasks
- Results
- Key Findings
- Notable quotes
- Recommendations
- Conclusion



Thank You!