

Usability Evaluations

- Think Aloud
- Heuristic Evaluation
- SUS
- Eye Tracking
- Cognitive Walkthrough
- ...

Usability Evaluations

- Think Aloud
- Heuristic Evaluation
- **SUS**
- Eye Tracking
- Cognitive Walkthrough
- ...

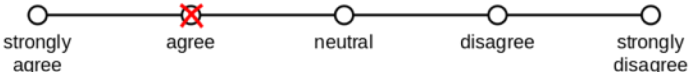
System Usability Scale (SUS)

- Subjective, self-reported
- Ten-item **Likert scale**

From Wikipedia Likert Scale article

Website User Survey

1. The website has a user friendly interface.



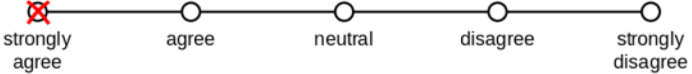
2. The website is easy to navigate.



3. The website's pages generally have good images.



4. The website allows users to upload pictures easily.



5. The website has a pleasing color scheme.



System Usability Scale (SUS)

- Subjective, self-reported
- Ten-item **Likert scale**
- Developed by John Brooke at Digital Equipment Corporation (DEC) in the UK in 1986
- Yields a single score on a scale of 0–100

System Usability Scale (SUS)

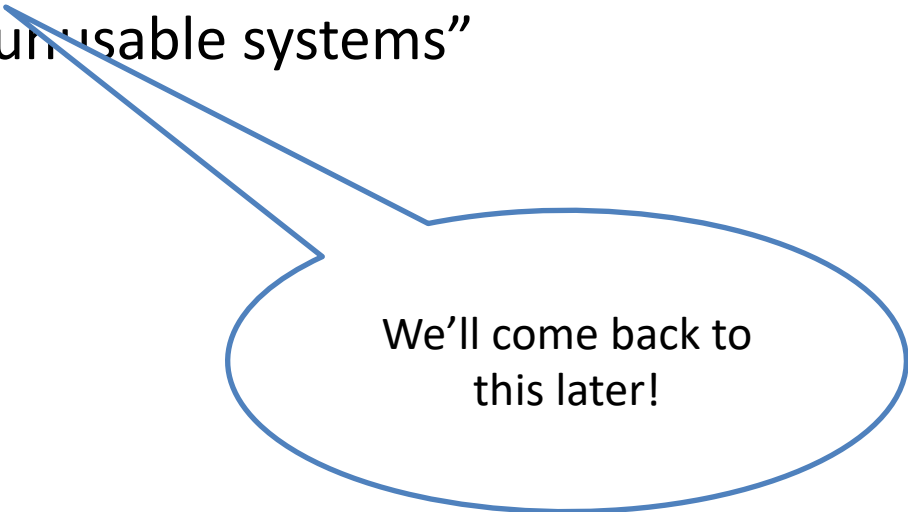
- Measurements of usability can include:
 - **effectiveness** (can users successfully achieve their objectives)
 - **efficiency** (how much effort and resource is expended in achieving those objectives)
 - **satisfaction** (was the experience satisfactory)
- -
 -
 -

System Usability Scale (SUS)

- Measurements of usability can include:
 - **effectiveness** (can users successfully achieve their objectives)
 - **efficiency** (how much effort and resource is expended in achieving those objectives)
 - **satisfaction** (was the experience satisfactory)
- Other things may also be of interest...e.g.,
 - **Learnability**
 - **Memorability**
 - ...

System Usability Scale (SUS)

- “SUS has become an industry standard, with references in over 1300 articles and publications. The noted benefits of using SUS include that it:
 - Is a very easy scale to administer to participants
 - Can be used on small sample sizes with reliable results
 - Is valid – it can effectively differentiate between usable and unusable systems”



We'll come back to
this later!

System Usability Scale (SUS)

- Participants score the following 10 items with one of five responses that range from Strongly Agree to Strongly Disagree:
 1. I think that I would like to use this system frequently.
 2. I found the system unnecessarily complex.
 3. I thought the system was easy to use.
 4. I think that I would need the support of a technical person to be able to use this system.
 5. I found the various functions in this system were well integrated.
 6. I thought there was too much inconsistency in this system.
 7. I would imagine that most people would learn to use this system very quickly.
 8. I found the system very cumbersome to use.
 9. I felt very confident using the system.
 10. I needed to learn a lot of things before I could get going with this system.

System Usability Scale (SUS)

- Participants score the following 10 items with one of five responses that range from Strongly Agree to Strongly Disagree

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

Note, this one is phrased “positively”

Note, this one is phrased “negatively”

(Scoring SUS FYI)

- The participant's scores for each question are converted to a new number, added together and then multiplied by 2.5 to convert the original scores of 0-40 to 0-100. Though the scores are 0-100, these are not percentages and should be considered only in terms of their percentile ranking.
- Based on research, a SUS score above a 68 would be considered above average and anything below 68 is below average, however the best way to interpret your results involves "normalizing" the scores to produce a percentile ranking.

...like the SUS but not: NASA-TLX

NASA Task Load Index (NASA-TLX)

“Rates perceived workload in order to assess a task, system, or team's effectiveness or other aspects of performance.”

NASA Task Load Index (NASA-TLX)

- Mental Demand
- Physical Demand
- Temporal Demand
- Performance
- Effort
- Frustration

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7 point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date
------	------	------

Mental Demand How mentally demanding was the task?

Very Low Very High

Physical Demand How physically demanding was the task?

Very Low Very High

Temporal Demand How hurried or rushed was the pace of the task?

Very Low Very High

Performance How successful were you in accomplishing what you were asked to do?

Perfect Failure

Effort How hard did you have to work to accomplish your level of performance?

Very Low Very High

Frustration How insecure, discouraged, irritated, stressed, and annoyed were you?

Very Low Very High

Effort or Performance	Temporal Demand or Frustration
Temporal Demand or Effort	Physical Demand or Frustration
Performance or Frustration	Physical Demand or Temporal Demand
Physical Demand or Performance	Temporal Demand or Mental Demand

15

Frustration or Effort	Performance or Mental Demand
Performance or Temporal Demand	Mental Demand or Effort
Mental Demand or Physical Demand	Effort or Physical Demand
Frustration or Mental Demand	

16