

Week 6: Evaluation Methods

Monday lecture
Uta Hinrichs & Tara Capel

mid-semester feedback
keep doing

13 responses

- Flipped classroom style
- In-class activities
- Practical nature of lectures and activities
- In-class discussions to hear different perspectives
- Interactive atmosphere

mid-semester feedback
stop doing

4 responses

- Ratio between required videos/readings and what is covered in class is off

mid-semester feedback start doing

12 responses

- Show length of the videos (done)
- Guest lectures
 - This is coming up!
- More examples from industry
- More hands-on activities/tutorials
- More clarity on coursework
 - Clarification of terms such as “justification”
 - More examples
 - Clarification on structure
 - We are trying to provide a clear framing, but please let us know if you would like more details on certain aspects

mid-semester feedback start doing

12 responses

- For future HCI courses in upcoming years
 - Longer lectures once a week
 - Consistent lecture halls
 - More practice-oriented sessions
 - Weekly & bi-weekly coursework
 - Field trip (?)

any questions about Assignment 2 (so far)?

today

- Re-cap: Think Aloud & hands-on activity
- Re-cap: Questionnaires/survey

re-cap “think aloud”

think aloud

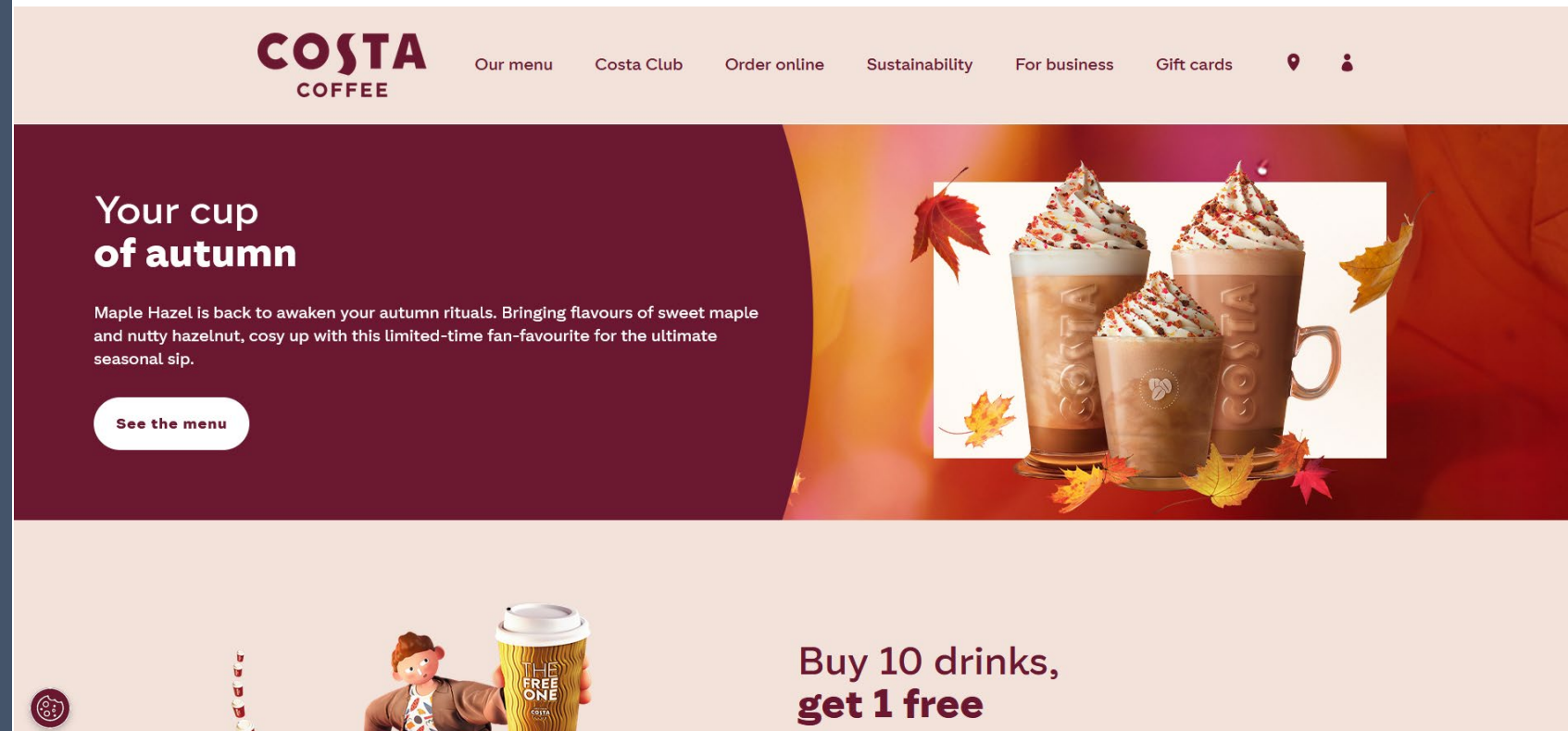
- Provide a set of tasks to solve using the interactive system in focus
- Ask the participant to “think aloud” while solving the task
- Only the participant should talk – debrief happens later
- Pros
 - Can help to uncover issues and sources of confusion and frustration with interactive systems
 - Can help to understand what people are doing, and the reasoning behind people’s interactions

think aloud

- Provide a set of tasks for participants to complete
- Ask the participant to verbalise what they are thinking, doing, feeling while they are completing the tasks
- Only the participant should talk – debrief happens later
- Cons
 - Requires “training” participants
 - Quality of results can vary dramatically between participants
 - “Think aloud” can influence people’s interactions with the system

think
aloud
example

- Ordering a Latte from the Costa website



think aloud activity

think aloud activity



- Get into groups of 2
 - Open the [think aloud tutorial](#)
 - Link on Open Course, HCI Week 6
1. Assign roles
 - Researcher/Observer
 - Study participant
 2. Go through the tutorial [5-10min]
 3. Swap roles
 4. Go through the tutorial in swapped roles [5-10min]
 5. Discuss
 - What was hard to do?
 - What was easy?

re-cap questionnaires/surveys

questionnaire

- Online forms or on-paper
- Questions should be directly related to the questions you are exploring
 - See study questions in Week 5
- Question types
 - Open-ended questions
 - Single- and multiple-choice questions
 - Matching/ranking
 - Rating scales (e.g., Likert Scales or Semantic Scales)
 - e.g., Likert Scales or Semantic Scales
 - Typically, on a 5 to 7 point scale

System Usability Scale (SUS)

- Post-test assessment of usability
- 10 questions addressing usability and learnability of a system

1. I think that I would like to use this system frequently.

1. Strongly Disagree

2.

3.

4.

5. Strongly Agree

☐☐☐☐☐

2. I found the system unnecessarily complex.

1. Strongly Disagree

2.

3.

4.

5. Strongly Agree

☐☐☐☐☐

3. I thought the system was easy to use.

1. Strongly Disagree

2.

3.

4.

5. Strongly Agree

☐☐☐☐☐

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

other established usability scales

- Single Ease Question (SEQ)
 - Post-task

Overall, this task was?

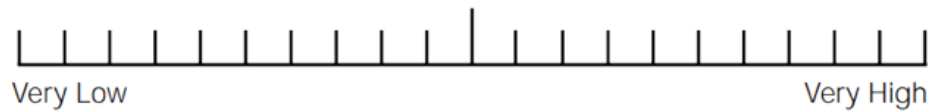
1. Very Difficult 2. 3. 4. 5. 6. 7. Very Easy



- NASA TLX (Task Load Index)
 - 6 questions
 - Unlabelled 21-point scale

Mental Demand

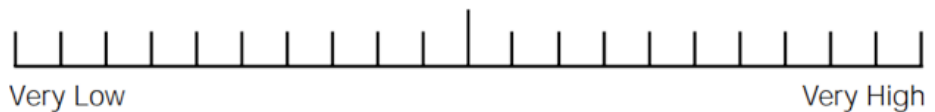
How mentally demanding was the task?



Very Low Very High

Physical Demand

How physically demanding was the task?



Very Low Very High

established usability scales pros & cons

- Pros
 - Extremely established and widely used
 - Experimentally valid and reliable
 - Large amounts of industry data available to help benchmark scores (especially for SUS)
- Cons (limitations)
 - Self-reported data can be unreliable
 - Participants may have a wildly different sense of the meaning of the scales (what does a “5” out of “7” mean?)
 - Require large participant samples to be meaningful
 - Measurement of subjective user experience rather than performance
 - A combination of satisfaction and performance can provide a clearer story
 - Metrics do not tell you the story behind participants’ rating or what the strengths or weaknesses of the system are

next steps

- Wednesday, Oct. 22; 14:10 – 15:00
- **Teviot Lecture Theatre - Doorway 5, MST**
- Questionnaire/Surveys
- Study variables