Introduction to Metrics

CC7 Human Computer Interaction

Table of Contents

- Introduction to metrics
- Usability metrics

Introduction to Metrics

What is metrics? | Types of metrics

What is Metrics?

- General term for the measurement / criteria during testing
- Used to determine whether or not something is considered successful or not

Types of Metrics

Depending on what needs to be tested, the metrics change as well:

- Usability
- Performance
- Issue-based
- Self-reported
- Behavioral

Usability Metrics

What is usability? | Usability metrics

What is Usability?

- Measure of how well a specific user in a specific context can use a product/design to achieve a defined goal
- Goal must be achieved effectively, efficiently and satisfactorily
- Design usability is measured throughout the development process

Usability Metrics

- How easy a design is to use with a group of representative users
- Involves repeated observation of users completing tasks through different types of designs

Performance Metrics

What is performance metrics? | Tasks success | Time on task | Errors | Efficiency | Learnability

What is Performance Metrics?

Used to determine:

- Magnitude of a specific usability issues
- How many people are likely to encounter the same issue

Basic Measures for Performance

- Tasks success
- Time on task
- Errors
- Efficiency
- Learnability

Task Success

- Most common performance metric
- Looks whether participants can complete a task or not
  - If the participant cannot complete the task, then something needs to be fixed
  - Requires a clear end-state
- Ways to measure task success:
  - Participant verbally articulates the answer to the task
  - Participant provides answers in a more structured way

Time on Task

- Also known as task completion time or task time
- Time it takes to perform a task
  - Faster is not always better
  - Important when tasks are performed repeatedly
- Ways to measure time on task:
  - Logging on start time and end time
  - Manual or automated timers
  - Rules for turning the timer on or off

Errors

- Possible outcome of a performance issue
- Incorrect action on the part of the user
- Any action that prevents the user from completing a task efficiently
- Ways to measure error:
  - Know what is the correct action / sequence of actions
  - Single or multiple opportunities for error

Efficiency

- Amount of effort required to complete a task
  - Cognitive – deciding what action to task
  - Physical – activity required to complete the action
- Steps to collect and measure efficiency:
  - Identify actions to be measured
  - Define start and end of action
  - Count the actions
  - Actions must be meaningful
  - Look only at successful tasks

Learnability

- How much time and effort is need to become proficient
- Steps to collect and measure learnability:
  - Time-on-task
  - Errors
  - Task successes per unit time
  - Track these over time

Issue-Based Metrics

What is issue-based metrics? | Possible issues | Identifying issues | Biases in metrics | Types of biases

What is Issue-Based Metrics?

Identifying:

- Understand what is and is not an issue
- Know both the product and the usability questions that arise
- Being very observant about participant behavior

Possible Issues

- Preventing task completion
- Takes the user "off-course"
- Confusion / misinterpretation of content / navigation
- Production of an error
- Visibility of items that should not be noticed
- Assumption of utilizing something correctly
- Assumption of task completion
- Conduct of the wrong action

Identifying Issues

In-Person Studies

- Requires interaction with the users
- Think-aloud protocols
- Observations

Automated Studies

- Allow participants to enter comments at the task level
- Looks into success, time, ease of use ratings, and verbatim comments

Biases in Metrics

- ***Bias –*** prejudice in favor or against a thing, person, or group compared to another
- Bias when identifying metrics:
  - Systematic error in development and testing
  - Skews results of testing
  - Issues may not be fully resolved
  - Limits the types of users who can use the system

Types of Biases

- Participant
  - Demographics
  - Role of the participant
  - Capabilities of the participant
  - Number of participants
  - Level of feedback from the participant
- Task
  - Clearly defined v. open ended
  - Self-generated v. manually operated

Types of Biases

- Method
  - Input and output devices
  - Data gathering procedure
  - Algorithms used
  - Features and constraints focused
- Artifact
  - Hardware specifications
  - Software specifications
  - System performance

Types of Biases

- Environment
  - Equipment in conjunction to the room
  - Size of the room
  - External forces
- Moderator
  - Establishment of rapport
  - Comfort of the user

Self-Reported Metrics

Why self-reported metrics? | Collection of self-reported data | Biases in self-reported data

Why Self-Reported Metrics?

- Important information about user's perception of the system and their interaction with it
- How users feel about the system
- Tasks:
  - Collection of self-reported data
  - Analysis of self-reports
  - Post-task ratings

Collection of Self-Reported Data

Collection Tools

- Rating scale
  - Likert scale – statement to which the response is the level of agreement
  - Presenting pairs of bipolar adjectives
- Open-ended questions
  - Contains a set of premade questions
  - Helps guide the participant to deliver a relevant answer
  - Not meant to derive an answer the moderator wants to hear

Collection of Self-Reported Data

Collection Methods

Method	*Pros*	*Cons*
Oral	Easiest method for quick ratings	Observer has to take notes
Written	Good for quick ratings and longer surveys	Need to be manually entered
Online	Good for quick ratings and longer surveys	Needs computer access

Biases in Self-Reported Data

- Selection bias
  - Selection of participants with specific attributes for testing
  - Does not demonstrate a truly randomized sample
- Reporting bias
  - Selective revealing or suppression of information
- Social desirability bias
  - Tendency to provide more positive feedback in person or on the phone

Biases in Self-Reported Data

Avoiding Bias

- Person collecting should be different from the test monitor
  - Prevents *observer-expectancy effect* – collector of results can subconsciously influence the people participating in an experiment
- Make the survey anonymous
  - Prevents prejudices and selection bias
  - Examples of prejudices: ageism, classism, racism, sexism, etc.

Analyzing Self-Reports

Quantitative Data

- Assign a numeric value, compute for averages
- Focus on top 2 and bottom 2 responses

Qualitative Data

- Summarizing responses from open-ended questions
- Identify keywords through apriori coding
- Group keywords into themes

Types of Post-Task Ratings

- Ease of use
- After-scenario questionnaire
- Expectation measure
- Usability magnitude estimation

Behavioral and Psychological Metrics

Observing and coding overt behaviors | Affective states | Data capture equipment | Other concepts

Observing and Coding Overt Behaviors

Verbal behaviors

- Usually expressed by the user during oral testing
- Accompanied with non-verbal behaviors

Non-verbal behaviors

- Usually described by affective states
- Requires observation and specialized equipment

Affective States

- Boredom
- Confusion
- Delight
- Frustration
- Engaged concentration (flow)
- On-task and off-task solitary
- On task receiving and giving answers

Data Capture Equipment

- Facial expressions
  - Eye tracking
    - Proportion of users looking at specific element
    - Time to notice an element
  - Scan paths
  - Skin conductance
- Posture
  - Pressure
  - Brain signal
  - Motion

Other Concepts

- Incidence of affective states
- Persistence of affective states
- Likelihood to games

Pointers for Review

Topic	Number of Items	Number of Points
Interaction frameworks	5	10
Assessing interactions	5	10
Design principles	20	40
Usability testing	10	20
Metrics	10	20
Total	**50**	**100**