User Metrics

1. **Understanding UX Metrics:** UX metrics are like specialized tools that help us measure and understand different aspects of how users experience a product or system. Each metric focuses on a specific part of this experience.
2. **Using Multiple Metrics:** Sometimes, one tool isn't enough to give us the full picture. We use several metrics to look at the user experience from different angles.
3. **Objective Metrics:** Some of these tools are all about hard facts. We collect data while users are performing specific tasks. For instance, we might measure how long it takes for users to complete a task or how many mistakes they make.
4. **Subjective Metrics:** Other tools focus on how users feel about the experience. We often gather this information through surveys or questionnaires, and then turn those feelings into numbers, usually by calculating averages.
5. **More Examples of Objective Metrics:** Beyond the examples mentioned, we can also track how often users need help, how much time is spent on errors and correcting them, and how many times users repeat failed actions.
6. **Benchmark Level:** This is like our starting point, a reference. It represents the level of performance for the current system or version. We use it as a basis for comparison to see how well any new system performs in relation to the old one. It's the point of comparison, whether it's measured automatically or manually.

UX TARGET

1. **Target Level in UX Metrics:** The target level in a UX metric is the value that shows we've achieved a successful user experience. It's like our goal for each specific aspect we're measuring.
2. **Improvement Focus:** When our UX metrics don't meet these target levels during evaluation, it's a sign that we need to improve. Designers should focus on these areas.
3. **Balanced Target Levels:** We need to make sure that the target levels for all our UX measurements are achievable at the same time. We shouldn't set goals in a way that achieving one makes it really hard to achieve another related goal.
4. **Setting Target Levels:** So, how do we decide on these target levels? Usually, we set them to be better than the starting point (baseline level). After all, why build a new system if it's not going to be an improvement? But we also need to allow time for users to get used to the system, especially at the beginning.

In simple terms, target levels are what we aim for in UX metrics to ensure a successful user experience. When we don't reach them, it's a sign that we need to improve. These goals should be balanced and better than what we started with. But we also need to be realistic about giving users time to adapt.

- Through years of working with real-world UX practitioners and doing our own user experience evaluations, we have refined the concept of a UX target table.

Our UX target table, as evolved from the Whiteside, Bennett, and Holtzblatt (1988) usability specification table

Work Role: User Class	UX Goal	UX Measure	Measuring Instrument	UX Metric	Baseline Level	Target Level	Observed Results

- For convenience, one row in the table is called a "UX target."
- The first three columns are for the work role and related user class to which this UX target applies, the associated UX goal, and the UX measure.
- The three go together because each UX measure is aimed at supporting a UX goal and is specified with respect to a work role and user class combination.

1. **Developing UX Targets:** Over time, as we've worked with UX professionals and conducted our own user experience assessments, we've improved the idea of a "UX target table."
2. **UX Targets in the Table:** Think of the table like a list. Each row in this list is referred to as a "UX target." It's like a specific goal we want to achieve.
3. **Table Columns:** This list has columns, like a chart. The first three columns tell us:
    - Who the goal is for (a specific work role or type of user)
    - What the goal is (the desired user experience)
    - How we measure the goal (the specific metric we use)

Work Roles, User Classes, And UX Goals

Choosing a work role, user class, and UX goal for a UX target

Work Role: User Class	UX Goal	UX Measure	Measuring Instrument	UX Metric	Baseline Level	Target Level	Observed Results
Ticket buyer: Casual new user, for occasional personal use	Walk-up ease of use for new user						

> # UX Measures
>
> Some common UX measures that can be paired with quantitative metrics include:
>
> **Objective UX measures (directly measurable by evaluators)**
>
> - Initial performance
> - Long-term performance (longitudinal, experienced, steady state)
> - Learnability
> - Retainability
>
> **Subjective UX measures (based on user opinions)**
>
> - First impression (initial opinion, initial satisfaction)
> - Long-term (longitudinal) user satisfaction

UX Measures are ways to gauge how good a user's experience is. We use these measures to understand the user's perspective.

Objective UX Measures are things we can directly measure, like:

- How well a user performs initially (at the start)
- How well they perform over a longer time (once they've used the system for a while)
- How easy it is for them to learn to use the system
- How well they retain what they've learned

Subjective UX Measures are based on what users think:

- Their first impressions (what they feel when they first start using the system)
- How satisfied they are with the system over a longer time

In simple terms, we use these measures to see how well users are doing and how happy they are with the system. Some measures are based on facts, while others are based on what users think and feel.

Choosing initial performance and first impression as UX measures

Work Role: User Class	UX Goal	UX Measure	Measuring Instrument	UX Metric	Baseline Level	Target Level	Observed Results
Ticket buyer: Casual new user, for occasional personal use	Walk-up ease of use for new user	Initial user performance					
Ticket buyer: Casual new user, for occasional personal use	Initial customer satisfaction	First impression					

UX Evaluation Techniques

Formative vs. Summative Evaluation:

- **Formative Evaluation:** This is like a chef tasting soup while cooking it. The chef is trying to figure out what's missing or what needs fixing. Similarly, in design, formative evaluation involves collecting feedback and qualitative data to spot and solve issues in the user experience (UX). It's like troubleshooting and making improvements during the design process.
- **Summative Evaluation:** Now, think of when the guests taste the soup at dinner. This is the final judgment, the moment of truth. In design, summative evaluation is about summing up the

overall quality of the design. It collects quantitative data to see how well the design performs. It's like the final grade for the design.

Formal Summative Evaluation: This is like a serious scientific experiment. It's a structured and rigorous way to compare different aspects of a design. It's almost like conducting a controlled test to see which design elements work best.

Informal Summative Evaluation: This is a more relaxed approach. It's often used alongside formative evaluation. Instead of a strict experiment, it's like casually checking to see how well the design is doing. It uses metrics, like how quickly users complete tasks, to track progress and see if the design is meeting its goals.

Types Of Formative And Informal Summative Evaluation Methods

• Dimensions for Classifying Formative UX Evaluation Methods

– empirical method vs. analytic method

– rigorous method vs. rapid method

Rigorous Method vs. Rapid Method: These are two different approaches for evaluating user experience (UX), and each has its own use.

Rigorous Method:

- Choose this when you want a thorough and effective evaluation. It's like a deep dive.
- It's more expensive and takes more time, just like conducting detailed research.

- Use it for assessing specific UX measures and metrics, like how fast users complete tasks and how many errors they make. This helps in performance-oriented scenarios.
- Ideal when you need a controlled environment without distractions, much like a lab setting.
- Opt for rigorous lab-based testing when you want reliable data and can invest in it.

Rapid Method:

- This is the quick and cost-effective way to evaluate UX. It's like a speed check.
- It might not be as effective as the rigorous method but can still provide valuable insights.
- Choose rapid evaluation when things are changing rapidly, especially in the early stages of design. It's like a quick check to see if you're on the right track.
- Use it for design concepts and get initial feedback from the design team, customers, and potential users. It's like a fast brainstorming session.

In simple terms, if you want a deep, thorough evaluation, go for the rigorous method, even if it's more expensive and time-consuming. If you need a quick check to keep things moving, use the rapid method, especially when things are changing quickly in the design process.

Analytic Method vs. Empirical Method: These are two different ways to evaluate a design or product.

Analytic Method:

- It's all about looking at the design itself, focusing on its inherent attributes and qualities.

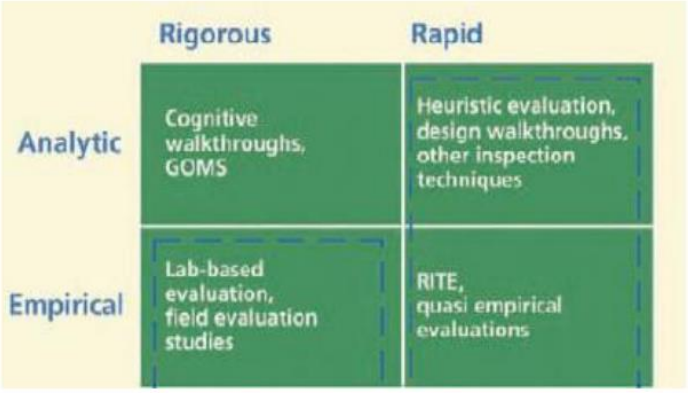
- You analyze the design based on its characteristics and features without involving real users.
- Think of it as a critical examination of the design on paper or on a computer screen, like a detailed review.

Empirical Method:

- This method involves real users actively using the design.
- You collect data by watching and recording how actual users interact with the design, often in a controlled lab environment.
- It's like a hands-on, real-world test of the design to see how it performs in practice.

In simple terms, the analytic method is like evaluating a recipe by reading it, while the empirical method is like testing the recipe by actually cooking the dish and tasting it. Analytic is about characteristics, and empirical is about real-world performance.

- **Where the Dimensions Intersect**



Sample UX evaluation methods at intersections between the dimensions of UX evaluation method types.

Types Of Evaluation Data

- **Objective Data vs. Subjective Data**

  - **Objective** UX data are data observed directly by either the evaluator or the participant.
  - **Subjective** UX data represent opinions, judgments, and other subjective feedback usually from the user, concerning the user experience and satisfaction with the interaction design.

- **Quantitative Data vs. Qualitative Data**

  Quantitative data are numeric data, such as data obtained by user performance metrics or opinion ratings. Quantitative data are the basis of a informal summative evaluation component and help the team assess U achievements and monitor convergence toward UX targets, usually in comparison with the specified levels set in the UX targets.

  Qualitative data are non-numeric and descriptive data, usually describing a UX problem or issue observed or experienced during usage. Qualitative data are usually collected via critical incident and/or the think aloud technique and are the key to identifying UX problems and their causes

Quantitative Data: These are numbers, like scores, times, or ratings, that give a measurable, numerical value to something. In UX, this might be user performance data or ratings from surveys. You use quantitative data to see how well you're meeting your goals.

Qualitative Data: This is more descriptive and non-numeric. It's about the qualities and experiences users have. Qualitative data helps identify problems and understand why they happen. It's often collected by listening to users' thoughts or incidents they describe.

In short, quantitative data is about numbers and measurements, while qualitative data is about descriptions and understanding user experiences

Some Data Collection Techniques

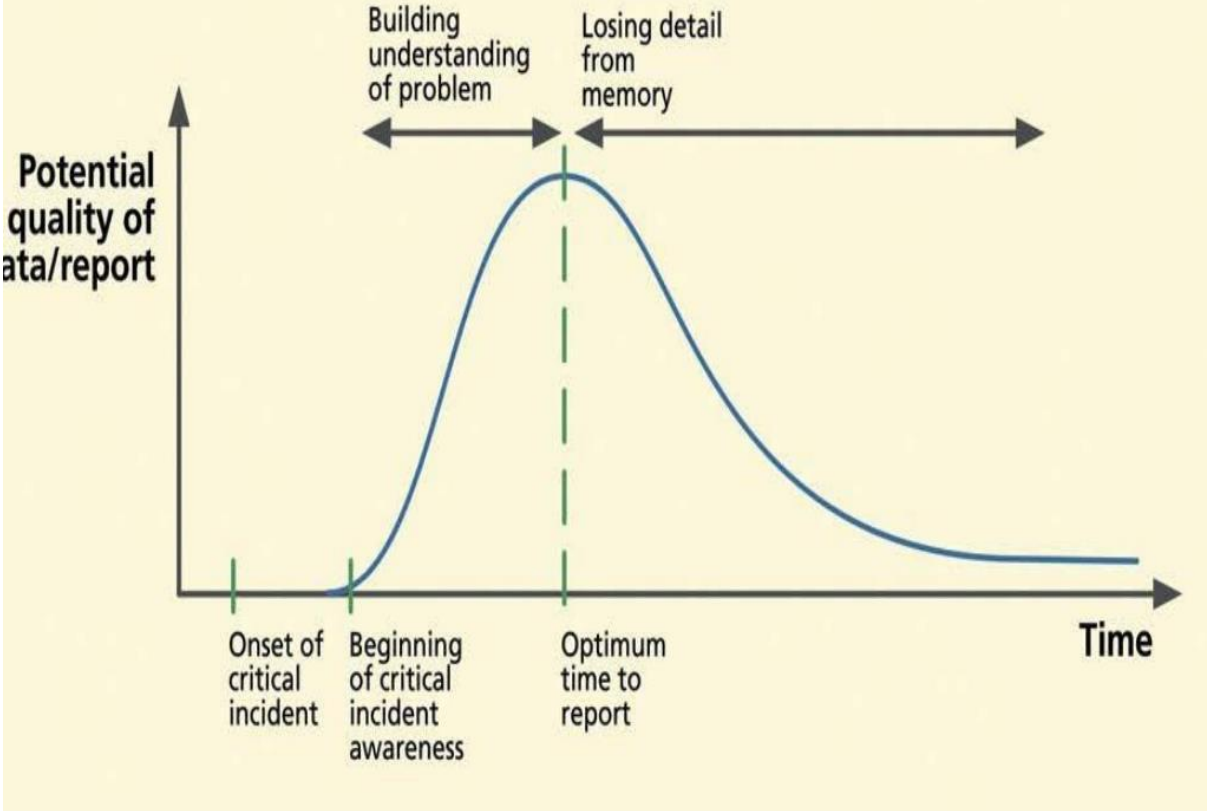
- **Critical Incident Identification**
  - the user's general activity or task
  - objects or artifacts involved
  - the specific user intention and action that led immediately to the critical incident
  - expectations of the user about what the system was supposed to do when the critical incident occurred
  - what happened instead
  - as much as possible about the mental and emotional state of the user
  - indication of whether the user could recover from the critical incident and, if so, a description of how the user did so
  - additional comments or suggested solutions to the problem

- Relevance of critical incident data
- History of critical incident data
- Mostly used as a variation
- Critical incident reporting tools
- Who identifies critical incidents?
- Timing of critical incident data capture: The evaluator's awareness zone

Critical incident identification involves documenting specific aspects of an incident that's crucial for understanding a problem or situation. Here's what you should focus on when identifying critical incidents:

1. The user's general activity or task: What were they trying to accomplish?
2. Objects or artifacts involved: What tools or objects were they using?
3. The specific user intention and action that led to the critical incident: What were they trying to do?
4. Expectations of the user about what the system was supposed to do when the incident occurred: What did they think was going to happen?
5. What happened instead: What was the actual outcome?
6. As much as possible about the mental and emotional state of the user: How were they feeling?
7. Indication of whether the user could recover from the incident and how: Could they fix the problem themselves?
8. Additional comments or suggested solutions to the problem: Any thoughts or ideas for improvement?

This information is important for understanding the relevance and history of critical incidents, and it's usually collected using specific tools and reported by evaluators who observe or work with users. The timing of data capture, referred to as the evaluator's awareness zone, is also crucial to ensure incidents are recorded when they happen.



The Think-Aloud Techniques

- The think-aloud technique is a qualitative data collection technique in which user participants, as the name implies, express verbally their thoughts about their interaction experience, including their motives, rationale, and perceptions of UX problems.
- The think-aloud technique is simple to use, for both analyst and participant. It is useful for when a participant walks through a prototype or helps you with a UX inspection.
- The think-aloud technique is also effective in assessing emotional impact because emotional impact is felt internally and the internal thoughts and feelings of the user are exactly what the think-aloud technique accesses for you.

Questionnaires

Semantic differential scales: These are questionnaires with opposites at each end of a scale, and users rate their experience somewhere in between. For example, rating a website from "difficult" to "easy."

The Questionnaire for User Interface Satisfaction (QUIS): A specific questionnaire designed to assess user satisfaction with software interfaces.

User Evaluation of Interactive Computer Systems

For each question, please circle the number that most appropriately reflects your impressions about this topic with respect to using this computer system or product.

1. Terminology relates to task domain	[distantly]	0 1 2 3 4 5 6 7 8 9 10	[closely]
2. Instructions describing tasks	[confusing]	0 1 2 3 4 5 6 7 8 9 10	[clear]
3. Instructions are consistent	[never]	0 1 2 3 4 5 6 7 8 9 10	[always]
4. Operations relate to tasks	[distantly]	0 1 2 3 4 5 6 7 8 9 10	[closely]
5. Informative feedback	[never]	0 1 2 3 4 5 6 7 8 9 10	[always]
6. Display layouts simplify tasks	[never]	0 1 2 3 4 5 6 7 8 9 10	[always]
7. Sequence of displays	[confusing]	0 1 2 3 4 5 6 7 8 9 10	[clear]
8. Error messages are helpful	[never]	0 1 2 3 4 5 6 7 8 9 10	[always]
9. Error correction	[confusing]	0 1 2 3 4 5 6 7 8 9 10	[clear]
10. Learning the operation	[difficult]	0 1 2 3 4 5 6 7 8 9 10	[easy]
11. Human memory limitations	[overwhelmed]	0 1 2 3 4 5 6 7 8 9 10	[are respected]
12. Exploration of features	[discouraged]	0 1 2 3 4 5 6 7 8 9 10	[encouraged]
13. Overall reactions	[terrible]	0 1 2 3 4 5 6 7 8 9 10	[wonderful]
	[frustrating]	0 1 2 3 4 5 6 7 8 9 10	[satisfying]
	[uninteresting]	0 1 2 3 4 5 6 7 8 9 10	[interesting]
	[dull]	0 1 2 3 4 5 6 7 8 9 10	[stimulating]
	[difficult]	0 1 2 3 4 5 6 7 8 9 10	[easy]

An excerpt from QUIS, with permission

The System Usability Scale (SUS): A widely used questionnaire to measure the usability of a system or product.

- **The System Usability Scale (SUS)**

The questions are presented as simple declarative statements, each with a five point Like scale anchored with "strongly disagree" and "strongly agree" and with values of 1 through 5. These 10 statements are:

- 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

The Usefulness, Satisfaction, and Ease of Use (USE) Questionnaire: A questionnaire that gauges the usefulness, satisfaction, and ease of use of a product. The Usefulness, Satisfaction, and Ease of Use (USE) Questionnaire

Usefulness	It helps me be more effective. It helps me be more productive. It is useful. It gives me more control over the activities in my life. It makes the things I want to accomplish easier to get done. It saves me time when I use it. It meets my needs. It does everything I would expect it to do.
Ease of use	It is easy to use. It is simple to use. It is user-friendly. It requires the fewest steps possible to accomplish what I want to do with it. It is flexible. Using it is effortless. I can use it without written instructions. I do not notice any inconsistencies as I use it. Both occasional and regular users would like it. I can recover from mistakes quickly and easily. I can use it successfully every time.
Ease of learning	I learned to use it quickly. I easily remember how to use it. It is easy to learn to use it. I quickly became skillful with it.
Satisfaction	I am satisfied with it. I would recommend it to a friend. It is fun to use. It works the way I want it to work. It is wonderful. I feel I need to have it. It is pleasant to use.

Questions in USE questionnaire

Data Collection Techniques Especially for Evaluating Emotional Impact

Data Collection Techniques Especially for Evaluating Emotional Impact

- Self-reported indicators of emotional impact
  Questionnaires as a verbal self-reporting technique for collecting emotional impact data
- Observing physiological responses as indicators of emotional impact
- Bio-metrics to detect physiological responses to emotional impact

“Emotion is a multifaceted phenomenon which people deliver through feeling states, verbal and non-verbal languages, facial expressions, behaviors, and so on.” Therefore, these are the things to “measure” or at least observe or ask about.

To access these emotional reactions, we must tap into the user’s subjective feelings; one effective way to do that is to have the user or participant do the reporting. Thus, verbal participant self-reporting techniques are a primary way that we collect emotional impact indicators.

(The paragraph highlights the importance of evaluating the emotional impact of a product or system. Emotions are complex and can be expressed through various means such as feelings, language, facial expressions, and behaviors. To assess these emotional reactions, it's crucial to tap into the

user's subjective feelings. This is typically done by having users report on their emotions and experiences. By allowing users to verbally express their feelings, you can collect valuable insights into the emotional aspects of their interaction with the product or system. This user self-reporting technique helps uncover how users feel about their experience and the emotional impact it has on them.)

Data Collection Techniques to Evaluate Phenomenological Aspects of Interaction

phenomenological usage focuses on understanding human activities and how people integrate systems or products into their daily lives over time. It's not just about tasks but about the overall user experience in a real-life context.

To collect data in phenomenological studies, researchers often need to engage with users for extended periods, similar to contextual inquiry or traditional ethnography. The goal is to observe and identify various aspects of long-term product usage, such as:

1. Typical usage patterns and behaviors.
2. Moments of joy and satisfaction in the user experience, providing insights into what aspects of the design contribute to a positive experience.
3. Problems and difficulties users encounter during usage, which can hinder a high-quality experience.
4. Unmet user needs or desires that the product does not address.
5. How the primary way of using the product changes or evolves over time.
6. How users adapt the product to their needs, potentially uncovering new and unexpected usage scenarios.

These studies aim to gain a deeper understanding of how users interact with products or systems in their daily lives and how these interactions evolve over time. This information can help improve the design and user experience of the product.

- Long-term studies required for phenomenological evaluation
- Goals of phenomenological data collection techniques
- Diaries in situated longitudinal studies
- Evaluator triggered reporting for more representative data
- Periodic questionnaires over time
- Direct observation and interviews in simulated real usage situations

VARIATIONS IN FORMATIVE EVALUATION RESULTS

Formative evaluation results are the findings and feedback obtained during the development phase to improve a project or product. They guide design enhancements.

- vague goals (varying evaluation focus)
- vague evaluation procedures (the methods do not pin down the procedures so each application is a variation and an adaptation)
- vague problem criteria (it is not clear how to decide when an issue represents a real problem)

Variations in formative evaluation results can occur due to several factors, including:

1. Vague Goals: When the goals of the evaluation are not well-defined or vary from one evaluation to another, the focus of the

assessment may differ. This can lead to variations in the results, as the objectives are not consistent.

2. Vague Evaluation Procedures: If the evaluation methods and procedures are not clearly specified, practitioners may adapt or interpret them differently. These variations in the application of methods can lead to inconsistent results.
3. Vague Problem Criteria: In some cases, it may not be clear how to determine whether an issue identified during the evaluation represents a genuine problem. When the criteria for problem identification are ambiguous, different evaluators may reach different conclusions, resulting in varying results.

To ensure more consistent and reliable formative evaluation outcomes, it's essential to have clear and well-defined goals, evaluation procedures, and problem criteria in place. This helps in standardizing the evaluation process and reducing the potential for variations in results.

FEEDBACK TO PROCESS

- Now that you have been through an iteration of the UX process lifecycle, it is time to reflect not just on the design itself, but also on how well your process worked.
- If you have any suspicions after doing the testing that the quantitative criteria were not quite right, you might ask if your UX targets worked well.
- For example, if all target levels were met or exceeded on the very first round of evaluation, it will almost certainly be the case that your UX targets were too lenient.
- Even in later iterations, if all UX targets are met but observations during evaluation sessions indicate that participants were frustrated and performed tasks poorly, your intuition will probably tell you that the design is nevertheless not acceptable in terms of its quality of user experience.

- Then, obviously, the UX team should revisit and adjust the UX targets or add more considerations to your criteria for evaluation success.
- Next, ask yourself whether the benchmark tasks supported the evaluation process in the most effective way.
- Should they have been simpler or more complex, narrower or broader? Should any benchmark task description be reworded for clarification or to give less information about how to do a task?
- Finally, assess how well the overall process worked for the team. You will never be in a better position to sit down, discuss it, and document possible improvements for the next time.

importance of reflecting on your user experience (UX) design process after completing a round of evaluations. Here's a detailed and simpler breakdown:

1. **Reflect on the Process**: After going through a cycle of designing and testing a user experience, it's crucial to not just look at the design itself but also at how well your overall process worked.
2. **Evaluate Your UX Targets**: If, during testing, you suspect that the goals you set for the user experience weren't quite right, consider whether your UX targets, which are the specific goals you set, were effective.
3. **Consider Leniency of UX Targets**: For instance, if you easily met or exceeded all your UX targets in the first round of evaluation, it's a sign that your targets might have been too easy or lenient. Meeting all goals too quickly suggests there was room for improvement.
4. **User Frustration**: Even in later iterations of evaluation, if you manage to meet all your UX targets but still observe users getting frustrated and struggling during testing, it indicates that

there are issues with the design that your targets didn't account for.

5. **Revisit and Adjust**: In response to these observations, it's important for the UX team to review and possibly revise the UX targets. You might need to set higher or more challenging goals and consider additional criteria for evaluating the success of your design.
6. **Benchmark Task Evaluation**: Assess whether the tasks you used for benchmarking (comparing your design to a standard) were appropriate. Were they too easy or too complex, too specific or too broad? Consider if task descriptions need clarification or if they revealed too much information about how to complete the tasks.
7. **Team Evaluation**: Finally, evaluate how well the entire design and testing process worked for your team. This is a good opportunity for your team to sit down, discuss what worked and what didn't, and document ways to improve the process for the next round.

In essence, it's about learning from your design and testing experiences, adjusting your goals when needed, and continually improving your UX design process.

INFORMAL SUMMATIVE (QUANTITATIVE) DATA ANALYSIS

Work Role: User Class	UX Goal	UX Measure	Measuring Instrument	UX Metric	Baseline Level	Target Level	Observed Results	Meet Target?
Ticket buyer: Casual new user, for occasional personal use	Walk-up ease of use	Initial user performance	BT1: Buy special event ticket	Average time on task	3 min as measured at the kiosk	2.5 min	3.5 min	No
Ticket buyer: Casual new user, for occasional personal use	Walk-up ease of use for new user	Initial user performance	BT2: Buy movie ticket	Average number of errors	<1	<1	2	No
Ticket buyer: Casual new user, for occasional personal use	Initial customer satisfaction	First impression	Questions Q1–Q10 in questionnaire XYZ	Average rating across users and across questions	7.5/10	8/10	7.5	No

1. **Using Descriptive Statistics**: Designers rely on simple statistics like the mean (average) and standard deviation to evaluate whether the design has reached its UX target levels. These statistics help make an engineering judgment about the design's success.
2. **Informal Analysis**: If the design falls short of the target levels, qualitative analysis is used to understand how to improve the design and work towards meeting those goals in the next rounds of formative evaluation. It's a way to learn from the process and refine the design iteratively.
3. **Iterative Process**: Iteration, or going through multiple rounds of testing and refinement, might seem repetitive, but it's a crucial part of improving the user experience. It ensures that the design keeps getting better.

4. **Quantitative Data Analysis**: To analyze quantitative data, you first calculate basic statistics like averages for things like time taken, error counts, or user ratings, as specified in the UX targets. However, relying solely on averages can be misleading because outliers (unusually high or low data points) can skew the results.
5. **Caution with Mean Values**: Mean values can be influenced by outliers, so they don't always tell the full story. In the context of formative evaluation (where you're still refining the design), having a mean value that meets a target doesn't necessarily mean there are no significant UX issues.
6. **Comparing Results**: By comparing the observed results with the UX goals you've set, you can quickly see which targets have been met and which haven't during this round of formative evaluation.

FORMATIVE (QUALITATIVE) DATA ANALYSIS