UNIT - 04
UX Design Process

1.UX Goals

"UX goals are high-level objectives for an interaction design, stated in terms of anticipated user experience"

- UX goals can be driven by business goals and reflect real use of a product and identify what is important to an organization, its customers, and its users.
- They are expressed as desired effects to be experienced in usage by users of features in the design and they translate into a set of UX measures.
- A UX measure is a usage attribute to be assessed in evaluating a UX goal.

Example: User Experience Goals for Ticket Kiosk System

We can define the primary high-level UX goals for the ticket buyer to include:

- Fast and easy walk-up-and-use user experience, with absolutely no user training
- Fast learning so new user performance (after limited experience) is on par with that of an experienced user [from AB-4-8]
- High customer satisfaction leading to high rate of repeat customers [from BC-6-16]

Some other possibilities:

- High learnability for more advanced tasks [from BB-1-5]
- Draw, engagement, attraction
- Low error rate for completing transactions correctly, especially in the interaction for payment [from CG-13-17]

Measuring Instrument

"A measuring instrument is the means for providing values for a particular UX measure; it is the vehicle through which values are generated and measured. A typical measuring instrument for generating objective UX data is a benchmark task—for example, user performance of a task gives time and error data—while a typical measuring instrument for generating subjective UX data is a questionnaire"

UX Target Tables

- 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.

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

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					

Benchmark Task

"A benchmark task is a description of a task performed by a participant in formative evaluation so that UX measures such as time-on-task and error rates can be obtained and compared to a baseline value across the performances of multiple participants"

Measuring Instruments

Benchmark Tasks

- Address designer questions with benchmark tasks and UX targets
- Selecting benchmark tasks
  - Create benchmark tasks for a representative spectrum of user tasks.
  - Start with short and easy tasks and then increase difficulty progressively.
  - Include some navigation where appropriate.
  - Avoid large amounts of typing (unless typing skill is being evaluated).
  - Match the benchmark task to the UX measure.
  - Adapt scenarios already developed for design.
  - Use tasks in realistic combinations to evaluate task flow.

- Do not forget to evaluate with your power users.
- To evaluate error recovery, a benchmark task can begin in an error state.
- Consider tasks to evaluate performance in “degraded modes” due to partial equipment failure.
- Do not try to make a benchmark task for everything.

- Constructing benchmark task content

  Remove any ambiguities with clear, precise, specific, and repeatable instructions.

  - Tell the user what task to do, but not how to do it.

    Do not use words in benchmark tasks that appear specifically in the interaction design.

- Use work context and usage-centered wording, not system-oriented wording.
- Have clear start and end points for timing.
- Keep some mystery in it for the user.
- Annotate situations where evaluators must ensure pre-conditions for running benchmark tasks.
- Use “rubrics” for special instructions to evaluators.
- Put each benchmark task on a separate sheet of paper.
- Write a “task script” for each benchmark task.

Choosing "buy special event ticket" benchmark task as measuring instrument for "initial performance" UX measure in first 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	Initial user performance	BT1: Buy special event ticket				
Ticket buyer: Casual new user, for occasional personal use	Initial customer satisfaction	First impression					

Choosing "buy movie ticket" benchmark task as measuring instrument for second initial performance UX measure

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	BT1: Buy special event ticket				
Ticket buyer: Casual new user, for occasional personal use	Walk-up ease of use for new user	Initial user performance	BT2: Buy movie ticket				
Ticket buyer: Casual new user, for occasional personal use	Initial customer satisfaction	First impression					

- How many benchmark tasks and UX targets do you need?

  Ensure ecological validity [Write your benchmark task descriptions, how can the setting be made more realistic?]

  - What are constraints in user or work context?
  - Does the task involve more than one person or role?
  - Does the task require a telephone or other physical props?
  - Does the task involve background noise?
  - Does the task involve interference or interruption?
  - Does the user have to deal with multiple simultaneous inputs, for example ,multiple audio feeds through headsets?

User Satisfaction Questionnaires

As a measuring instrument for a subjective UX measure, a questionnaire related to various user interaction design features can be used to determine a user's satisfaction with the interaction design.

Choosing questionnaire as measuring instrument for first-impression UX measure

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	BT1: Buy special event ticket				
Ticket buyer: Casual new user, for occasional personal use	Walk-up ease of use for new user	Initial user performance	BT2: Buy movie ticket				
Ticket buyer: Casual new user, for occasional personal use	Initial customer satisfaction	First impression	Questions Q1–Q10 in the QUIS questionnaire				

UX METRICS

A UX metric describes the kind of value to be obtained for a UX measure. It states what is being measured. There can be more than one metric for a given measure.

Most commonly, UX metrics are objective, performance-oriented, and taken while the participant is doing a benchmark task. Other UX metrics can be subjective, based on a rating or score computed from questionnaire results.

Typical objective UX metrics include time to complete task1 and number of errors made by the user. Others include frequency of help or documentation use; time spent in errors and recovery; number of repetitions of failed commands and the number of commands, mouse-clicks, or other user actions to perform task(s)

Typically, subjective UX metrics will represent the kind of numeric outcome you want from a questionnaire, usually based on simple arithmetic statistical measures such as the numeric average.

Table 10-8

Choosing UX metrics for 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	BT1: Buy special event ticket	Average time on task			
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			
Ticket buyer: Casual new user, for occasional personal use	Initial customer satisfaction	First impression	Questions Q1–Q10 in the QUIS questionnaire	Average rating across users and across questions			

BASELINE LEVEL

The baseline level is the benchmark level of the UX metric; it is the “talking point” level against which other levels are compared. It is often the level that has been measured for the current version of the system (automated or manual).

UX Evaluation Techniques

Formative Vs. Summative Evaluation

- When the cook tastes the soup, that's formative; when the guests taste the soup, that's summative" (Stake, 2004, p. 17).
  - **Formative evaluation** is primarily diagnostic; it is about collecting qualitative data to identify and fix UX problems and their causes in the design.
  - **Summative evaluation** is about collecting quantitative data for assessing a level of quality due to a design, especially for assessing improvement in the user experience due to formative evaluation.

- **Qualitative Data**
  - Qualitative data are non-numeric and descriptive data, usually describing a UX problem or issue observed or experienced during usage.

- **Quantitative Data**

  Quantitative data are numeric data, such as user performance metrics or opinion ratings.

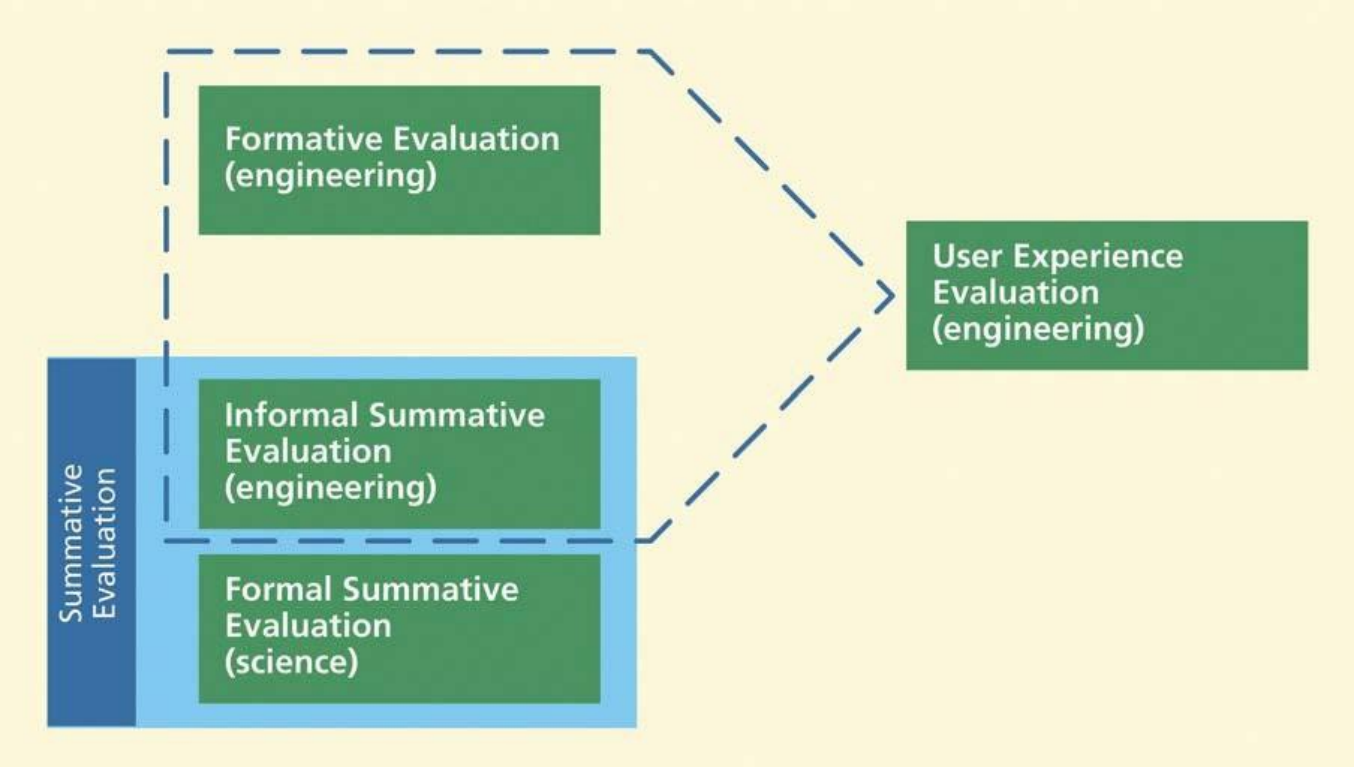
- **Formal summative evaluation**

  Formal summative evaluation is a kind of controlled hypothesis testing with a m by n factorial design with y independent variables, the results of which are subjected to statistical tests for significance. Formal summative evaluation is an important HCI skill.

- **Informal summative evaluation**

  is used, as a partner of formative evaluation, for quantitatively summing up or assessing UX levels using metrics for user performance (such as the time on task), for example, as indicators of progress in UX improvement, usually in comparison with pre-established UX target levels.

- **Engineering Evaluation of UX: Formative Plus Informal Summative**



UX evaluation is a combination of formative and informal summative evaluation.

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**

  Choose a rigorous empirical method such as lab-based testing when you need effectiveness and thoroughness, but expect it to be more expensive and time-consuming.

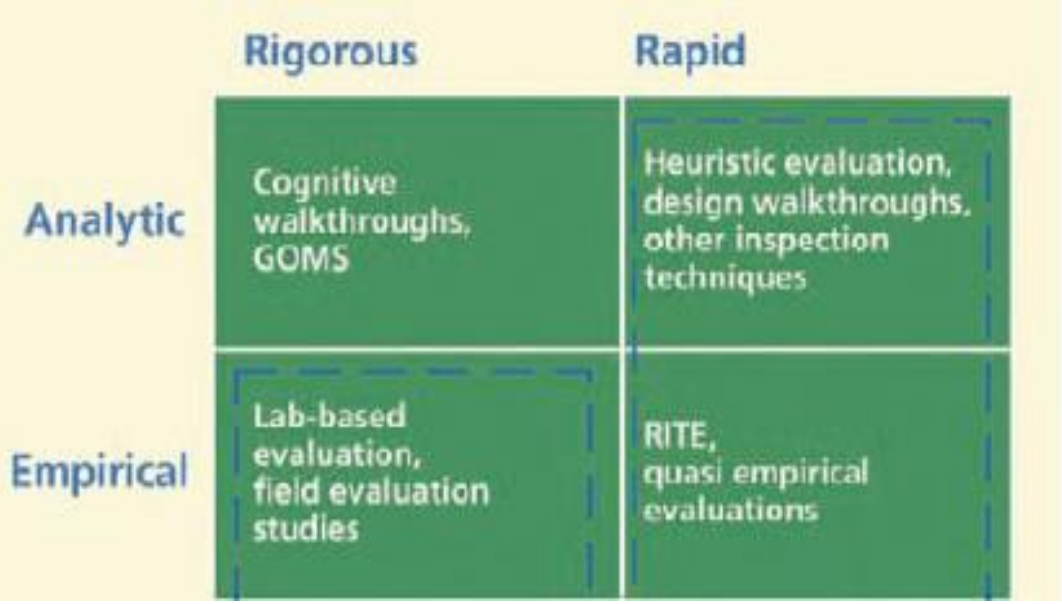
  Choose the lab-based method to assess quantitative UX measures and metrics, such as time-on-task and error rates, as indications of how well the user does in a performance- oriented context.

- Choose lab-based testing if you need a controlled environment to limit distractions.
- Choose empirical testing in the field if you need more realistic usage conditions for ecological validity than you can establish in a lab.
- However, UX evaluation methods can be faster and less expensive.
- Choose a rapid evaluation method for speed and cost savings, but expect it to be (or possibly acceptably) less effective.
- Choose a rapid UX evaluation method for early stages of progress, when things are changing a lot, anyway, and investing in detailed evaluation is not warranted.
- Choose a rapid method, such as a design walkthrough, an informal demonstration of design concepts, as a platform for getting initial reactions and early feedback from the rest of the design team, customers, and potential users.

- **Analytic Method vs. Empirical Method**

  - **Analytical** methods are based on looking at inherent attributes of the design
  - rather than seeing the design in use.
  - **Empirical** methods employ data observed in the performance of real user participants, usually data collected in lab-based testing.

- **Where the Dimensions Intersect**



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

6.3 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**

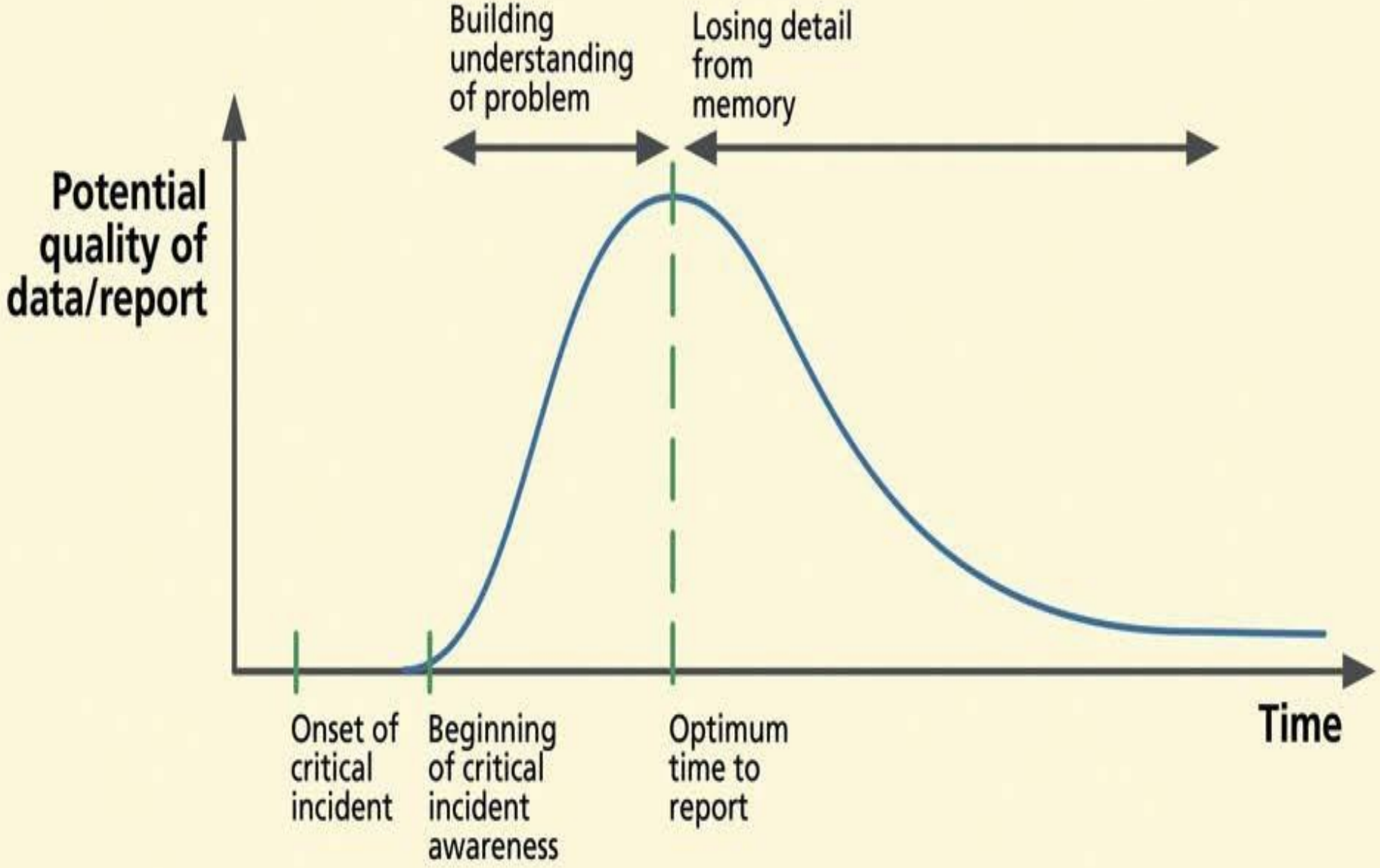
  Quantitative data are numeric data, such as data obtained by user performance metrics or opinion ratings. Quantitative data are the basis of a informal summative evaluation component and help the team assess U achievements and monitor convergence toward UX targets, usually in comparison with the specified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

6.4 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 descri[C]p[om]ti[p]o[ile]n[d] critical incident.

- **The Think-Aloud Technique**
- **Questionnaires**
  - Semantic differential scales
  - The Questionnaire for User Interface Satisfaction (QUIS)
  - The System Usability Scale (SUS)
  - The Usefulness, Satisfaction, and Ease of Use (USE) Questionnaire

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

An excerpt from QUIS, with permission

- **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

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

- **General-purpose usability questionnaires:**
  - Computer System Usability Questionnaire (CSUQ), developed by Jim Lewis (1995, 2002) at IBM, is well-regarded and available in the public domain.
  - Software Usability Measurement Inventory (SUMI) is "a rigorously tested and proven method of measuring software quality from the end user's point of view" (Human Factor Research Group, 1990). According to Usability Net, SUMI is "a mature questionnaire whose standardization base and manual have been regularly updated." It is applicable to a range of application types from desk-top applications to large domain-complex applications.
  - After Scenario Questionnaire (ASQ), developed by IBM, is available in the public domain (Bangor, Kortum, & Miller, 2008, p. 575).
  - Post-Study System Usability Questionnaire (PSSUQ), developed by IBM, is available in the public domain (Bangor, Kortum, & Miller, 2008, p. 575).

- **Web evaluation questionnaires:**
  - Website Analysis and Measurement Inventory (WAMMI) is “a short but very reliable questionnaire that tells you what your visitors think about your web site” (Human Factor Research Group, 1996b).
- **Multimedia system evaluation questionnaires:**
  - Measuring the Usability of Multi-Media Systems (MUMMS) is a questionnaire “designed for evaluating quality of use of multimedia software products” (Human Factor Research Group, 1996a).
- **Hedonic quality evaluation questionnaires:**
  - The Lavie and Tractinsky (2004) questionnaire
  - The Kim and Moon (1998) questionnaire with differential emotions scale

- **Modifying questionnaires for your evaluation**
  - choosing a subset of the questions
  - changing the wording in some of the questions
  - adding questions of your own to address specific areas of concern
  - using different scale values

- **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

Scale Item	Semantic Anchors	
Pragmatic Quality 1	Comprehensible	Incomprehensible
Pragmatic Quality 2	Supporting	Obstructing
Pragmatic Quality 3	Simple	Complex
Pragmatic Quality 4	Predictable	Unpredictable
Pragmatic Quality 5	Clear	Confusing
Pragmatic Quality 6	Trustworthy	Shady
Pragmatic Quality 7	Controllable	Uncontrollable
Hedonic Quality 1	Interesting	Boring
Hedonic Quality 2	Costly	Cheap
Hedonic Quality 3	Exciting	Dull
Hedonic Quality 4	Exclusive	Standard
Hedonic Quality 5	Impressive	Nondescript
Hedonic Quality 6	Original	Ordinary
Hedonic Quality 7	Innovative	Conservative
Appeal 1	Pleasant	Unpleasant
Appeal 2	Good	Bad
Appeal 3	Aesthetic	Unaesthetic
Appeal 4	Inviting	Rejecting
Appeal 5	Attractive	Unattractive
Appeal 6	Sympathetic	Unsympathetic
Appeal 7	Motivating	Discouraging
Appeal 8	Desirable	Undesirable

AttrakDiff emotional impact questionnaire as listed by Hassenzahl, Scho¨bel, and Trautman (2008), with permission

Scale	Item	English Anchor 1	English Anchor 2
Pragmatic quality	PQ1	People centric	Technical
Pragmatic quality	PQ2	Simple	Complex
Pragmatic quality	PQ3	Practical	Impractical
Pragmatic quality	PQ4	Cumbersome	Facile
Pragmatic quality	PQ5	Predictable	Unpredictable
Pragmatic quality	PQ6	Confusing	Clear
Pragmatic quality	PQ7	Unmanageable	Manageable
Hedonic – identity	HQI1	Isolates	Connects
Hedonic – identity	HQI2	Professional	Unprofessional
Hedonic – identity	HQI3	Stylish	Lacking style
Hedonic – identity	HQI4	Poor quality	High quality
Hedonic – identity	HQI5	Excludes	Draws you in
Hedonic – identity	HQI6	Brings me closer to people	Separates me from people
Hedonic – identity	HQI7	Not presentable	Presentable
Hedonic – stimulation	HQS1	Original	Conventional
Hedonic – stimulation	HQS2	Unimaginative	Creative
Hedonic – stimulation	HQS3	Bold	Cautious
Hedonic – stimulation	HQS4	Innovative	Conservative
Hedonic – stimulation	HQS5	Dull	Absorbing
Hedonic – stimulation	HQS6	Harmless	Challenging
Hedonic – stimulation	HQS7	Novel	Conventional
Attractiveness	ATT1	Pleasant	Unpleasant
Attractiveness	ATT2	Ugly	Pretty
Attractiveness	ATT3	Appealing	Unappealing
Attractiveness	ATT4	Rejecting	Inviting
Attractiveness	ATT5	Good	Bad
Attractiveness	ATT6	Repulsive	Pleasing
Attractiveness	ATT7	Motivating	Discouraging

A variation of the AttrakDiff emotional impact questionnaire, as listed , Held and Laugwitz (2006), reordered to group related items together, with permission

- **Data Collection Techniques to Evaluate Phenomenological Aspects of Interaction**
  - 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

6.5 Variations In Formative Evaluation Results

- **Reasons given by Hertzum and Jacobsen (2003) for the wide variation**
  - 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)

7. Informal Summative (Quantitative) Data Analysis

“ It’s a Rigorous UX evaluation – Data Analysis “

7.1 Lining up Your Quantitative Ducks

- The first step in analyzing quantitative data is to compute simple descriptive statistics (e.g., averages) for timing, error counts, questionnaire ratings, and so on, as stated in the UX targets.
- After you compute summary statistics of quantitative data, you add them to the “Observed Results” column at the end of the UX target table. As an example, partial results from a hypothetical evaluation of the Ticket Kiosk System are shown,

Example of partial informal quantitative testing results for the Ticket Kiosk System

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

7.2 The Big Decision: Can We Stop Iterating?

- Now it is time for a major project management decision: Should you continue to iterate? This decision should be a team affair and made at a global level, not just considering quantitative data. Here are some questions to consider:
  - Did you simultaneously meet all your target-level goals?
  - What is your general team feeling about the conceptual design, the overall interaction design, the metaphor, and the user experiences they have observed?

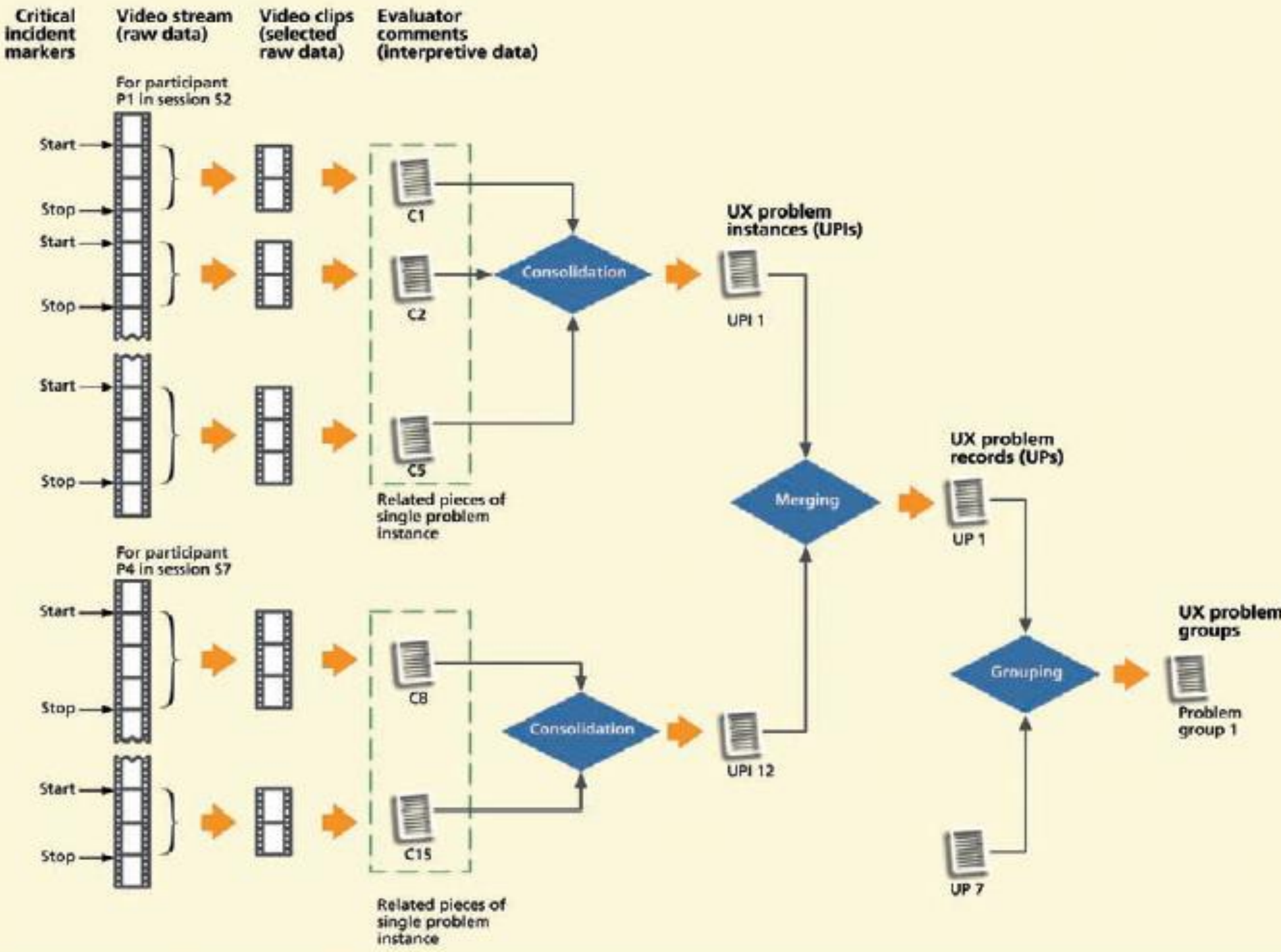
- If you can answer these questions positively, you can accept the design as is and stop iterating.
- Resource limitations also can force you to stop iterating and get on with pushing this version out in the hope of fixing known flaws in the next version.
- If and when you do decide to stop iterating, do not throw your qualitative data away, though; you paid to get it, so keep this round of problem data for next time.

8. Formative (Qualitative) Data Analysis

- Formative analysis of qualitative data is the bread and butter of UX evaluation.
- The goal of formative data analysis is to identify UX problems and causes (design flaws) so that they can be fixed, thereby improving product user experience.
- The process of determining how to convert collected data into scheduled design and implementation solutions is essentially one of negotiation in which, at various times, all members of the project team are involved.
- In the first part of qualitative analysis you should have all your qualitative data represented as a set of UX problem instances so that you can proceed with
- diagnosis and problem solutions.

  Figure illustrates the steps of qualitative data analysis: consolidating large sets of raw critical incident comments into UX problem instances, merging UX problem instances into UX problem records, and grouping of UX problem records so that we can fix related problems together.

- For practical purposes we have to separate our material into chapters.
- In practice, early analysis—especially for qualitative data—overlaps with the data collection process. Because evaluator comments are interpretive, we have already begun to overlap analysis of qualitative data with their capture.
- The earlier you think about UX problems and their causes, the better chance you have at getting all the information you will need for problem diagnosis.


Critical incident markers
Video stream (raw data)
Video clips (selected raw data)
Evaluator comments (interpretive data)
For participant P1 in session 52
Start
Stop
Start
Stop
Start
Stop
C1
C2
C5
Consolidation
Related pieces of single problem instance
UX problem instances (UPIs)
UPI 1
Merging
UX problem records (UPs)
UP 1
For participant P4 in session 57
Start
Stop
Start
Stop
C8
C15
Consolidation
Related pieces of single problem instance
UPI 12
Grouping
UP 7
UX problem groups
Problem group 1

8.1 Get an Early Jump on Problem Analysis

- Keep a participant around to help with early analysis
- Early UX problem data records
- Clean up your raw data before your memory fades
- Clarify and amplify your emotional impact data

8.2 Sources of Raw Qualitative Data

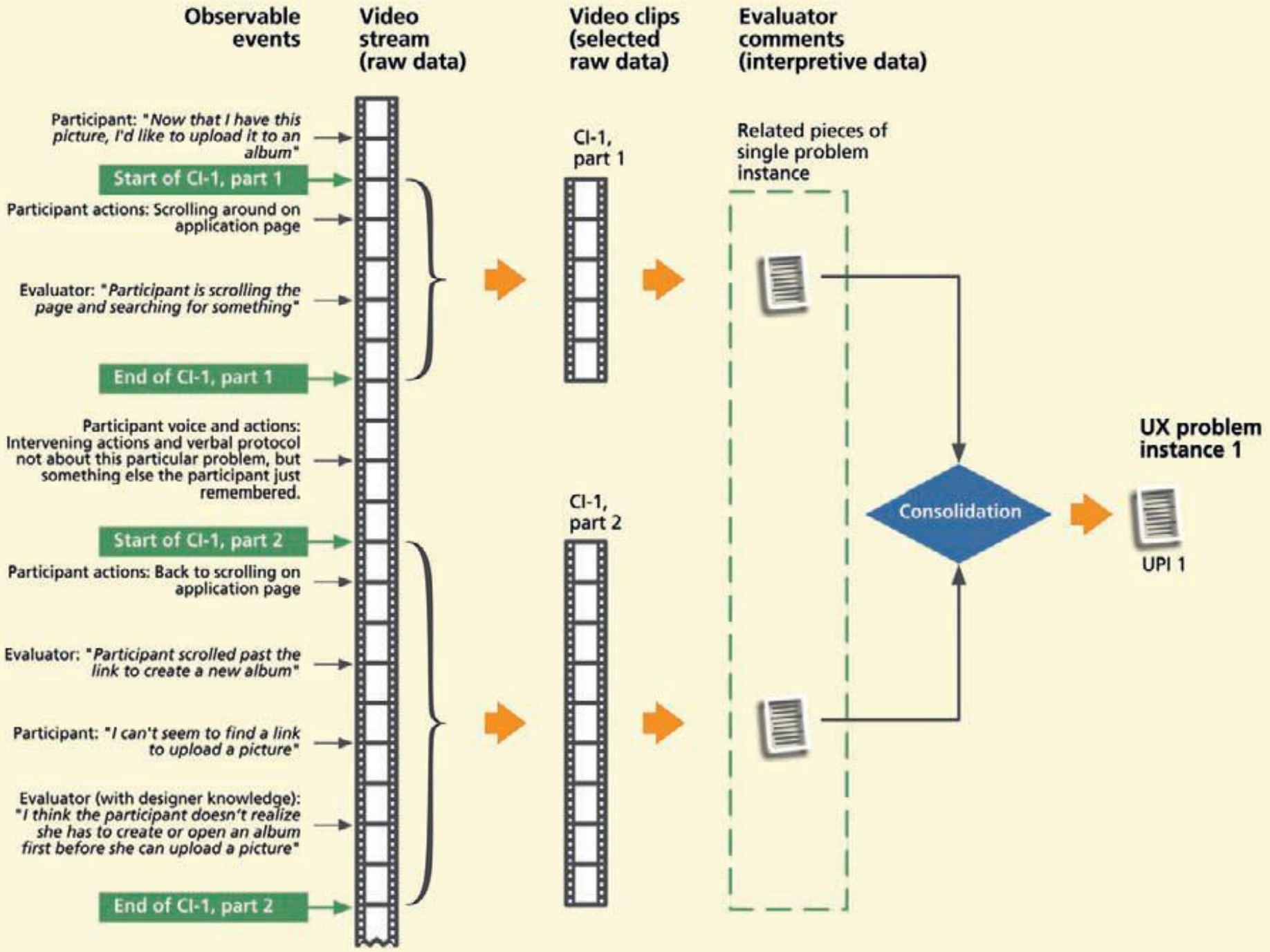
8.3 Isolate Individual Critical Incident Descriptions

8.4 Consolidating Raw Critical Incident Notes into UX Problem Instances

- The UX problem instance concept
- Gathering up parts of data for a critical incident

Example: Consolidating Critical Incident Parts of a Single UX Problem Instance

- Putting it into a UX problem instance


Observable events
Video stream (raw data)
Video clips (selected raw data)
Evaluator comments (interpretive data)
Participant: "Now that I have this picture, I'd like to upload it to an album"
Start of CI-1, part 1
Participant actions: Scrolling around on application page
Evaluator: "Participant is scrolling the page and searching for something"
End of CI-1, part 1
Participant voice and actions: Intervening actions and verbal protocol not about this particular problem, but something else the participant just remembered.
Start of CI-1, part 2
Participant actions: Back to scrolling on application page
Evaluator: "Participant scrolled past the link to create a new album"
Participant: "I can't seem to find a link to upload a picture"
Evaluator (with designer knowledge): "I think the participant doesn't realize she has to create or open an album first before she can upload a picture"
End of CI-1, part 2
CI-1, part 1
CI-1, part 2
Related pieces of single problem instance
Consolidation
UX problem instance 1
UPI 1

Problem Instance (UX)

"A UX problem instance is a single occurrence of an encounter with a given problem by a given user, inspector, or participant. When more than one participant experiences what is essentially the same problem, the encounters are counted as different instances so they are not reported as different problems."

Critical Incident

"A critical incident is a UX evaluation event that occurs during user task performance or other user interaction, observed by the facilitator or other observers or sometimes expressed by the user participant, that indicates a possible UX problem. Critical incident identification is arguably the single most important source of qualitative data."

8.5 UX Problem Instances

- UX problem instance content
  - Understand each problem
  - Glean insight into its causes and possible solutions
  - Be conscious of relationships among similar problems
- UX problem instance project context
  - Organization (e.g., company, department)
  - Project (e.g., product or system, project management, dates, budget, personnel)
  - Version (e.g., design/product release, version number, iteration number)
  - Evaluation session (e.g., date, participants, evaluators, associated UX target table)
  - Task run (e.g., which task, associated UX targets)

8.6 Merge Congruent UX Problem Instances into UX Problem Records

- Find and merge multiple UX problem instances representing the same problem
- Create UX problem records
- Group Records of Related UX Problems for Fixing Together
  - Problems may be in physical or logical proximity (e.g., may involve objects or actions within the same dialogue box).
  - Problems may involve objects or actions used in the same task.
  - Problems may be in the same category of issues or design features but scattered throughout the user interaction design.
  - Problems may have consistency issues that require similar treatments.
  - Observed problem instances are indirect symptoms of common, more deeply rooted,
  - UX problems. A telling indicator of such a deeply rooted problem is complexity and difficulty in its analysis.

8.7 UX Problem Data Management

Each UX problem record will eventually contain information about the problem: diagnosis by problem type and subtype, interaction design flaws as problem causes, cost/importance data estimating severity, management decisions to fix (or not) the problem, costs, implementation efforts, and downstream effectiveness.

8.8 Abridged Qualitative Data Analysis

- As an abridged approach formative (qualitative) data analysis:
  - Just take notes about UX problems in real time during the session.
  - Immediately after session, make UX problem records from the notes.
- As an alternative, if you have the necessary simple tools for creating UX problem records:

  Create UX problem records as you encounter each UX problem during the session.
  - Immediately after the session, expand and fill in missing information in the records.

  Analyze each problem, focusing on the real essence of the problem and noting causes (design flaws) and possible solutions.

9. 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.

10. Evaluation Report

10.1 Introduction

10.1.1 Importance of Quality Communication and Reporting

- All the effort and cost you invested thus far in UX evaluation can be wasted at the last minute if you do not follow up now to:
  - Inform the team and project management about the UX problems in the current design
  - Persuade them of the need to invest even more in fixing those problems.

10.1.2 Participant Anonymity

- Participant anonymity does not mean that you, as the evaluator or facilitator, do not know the names of participants. Somewhere along the line someone must have recruited and signed up and possibly even paid the participants.
- You should keep participant identification information in just one place—on a sheet of paper or in a database mapping the names to identification codes.
- Codes, never names, are used everywhere else in the evaluation process—on data collection forms, during data analysis, and in all reports.

10.2 Reporting Informal Summative Results

- the audience for your informal summative evaluation results should be strictly limited to your own project group.
- They are to be used only as an engineering tool within the project.
- You must be answerable to these questions:
  - What if You Are Required to Produce a Formative Evaluation Report for Consumption Beyond the Team?
  - What if You Need a Report to Convince the Team to Fix the Problems?

10.3 Reporting Qualitative Formative Results

- Common Industry Format (CIF) for Reporting Formal Summative UX Evaluation Results . The CIF standard calls out requirements for reports to include:
  - A description of the product
  - Goals of the testing
  - A description of the number and types of participants
  - Tasks used in evaluation

  - The experimental design of the test (very important for formal summative studies because of the need for eliminating any biases and to ensure the results do not suffer from external, internal, and other validity concerns)
  - Evaluation methods used
  - Usability measures and data collection methods employed
  - Numerical results, including graphical methods of presentation
- The American National Standards Institute (ANSI) approved this standard for summative reporting as ANSI-NCITS 354-2001 in December 2001 and it became an international standard, ISO/IEC 25062: Software Engineering—Software Product Quality Requirements and Evaluation (SQuaRE), in May 2005.
- Common Industry Format (CIF) for Reporting Qualitative Formative Results.

10.4 Formative Reporting Content

- Individual Problem Reporting Content

The basic information needed includes:

  - the problem description
  - a best judgment of the causes of the problem in the design
  - an estimate of its severity or impact
  - suggested solutions

- Include Video Clips Where Appropriate
- Pay Special Attention to Reporting on Emotional Impact Problems
- Including Cost-Importance Data

10.5 Formative Reporting Tone

- Respect feelings: Bridle your acrimony
- Accentuate the positive and avoid blaming

10.6 Formative Reporting over Time

10.7 Problem Report Effectiveness: The Need to Convince and Get Action with Formative Results

- Problem severity: more severe problems are more salient and carry more weight with designers
- Problem frequency: more frequently occurring problems are more likely to be perceived as "real"
- Perceived relevance of problems: designers disagreeing with usability practitioners on the relevance (similar to "realness") of problems did not fix problems that practitioners recommended be fixed