

5

UX Evaluation Methods and Techniques

Syllabus

Empirical UX evaluation : UX goals, metrics and Targets, Analytic UX evaluation : Data collection methods and Techniques.

Contents

- 5.1 UX Evaluation Methods and Techniques
- 5.2 Empirical UX Evaluation : UX Goals, Metrics, and Targets
- 5.3 Analytic UX Evaluation : Data Collection Methods and Techniques
- 5.4 Review Questions

5.1 UX Evaluation Methods and Techniques

5.1.1 Introduction

- Evaluate UX lifecycle activity stage of the overall UX design lifecycle template is shown in Fig. 5.1.1.

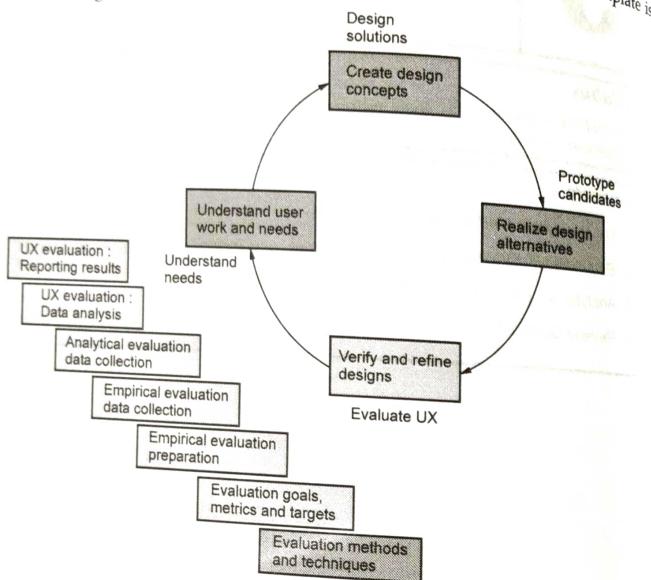


Fig. 5.1.1 Evaluate UX lifecycle activity

5.1.1.1 Methods versus Techniques

- The notions of procedures and techniques were previously defined. We will go over them and interpret them in the context of UX evaluation.
- In our work, there is no clear explanation of the difference between a method and a technique, with the key distinction being level. A technique is usually a lower-level manner of executing individual steps inside a process, whereas an evaluation method is a high-level overall approach of doing UX evaluation.
- A method of evaluation is, for example, lab-based empirical testing with users. Within that strategy, critical event identification is one of the techniques used to collect data on UX difficulties.

5.1.1.2 Types of UX Evaluation Data

- Data from UX evaluations can be objective or subjective, quantitative or qualitative. Because the two dimensions are orthogonal in practice, both objective and subjective data might be quantitative or qualitative. For example, questionnaire results are usually both subjective and quantitative.

Quantitative versus qualitative data

- Quantitative data is a type of numeric data that is used to judge a level of achievement. It is typically derived via measurements. Objective user performance data evaluated with benchmark tasks and subjective user opinion data measured with surveys are the two most prevalent types of quantitative data obtained in formative evaluation.
- Quantitative data form the foundation of the informal summative evaluation component, allowing the team to analyse UX achievements and track progress toward UX targets, usually in relation to the levels indicated in the UX targets.
- Non-numerical descriptive data is utilized to identify and solve UX issues. Qualitative data from a UX evaluation is usually a description of UX issues or difficulties that were noticed or experienced while using the product. Critical incident identification, the think-aloud approach, and UX inspections methods are commonly used to collect qualitative data, which is essential for detecting UX problems and their causes.

Objective versus subjective data

- Data collected directly from users is referred to as objective UX data. Observations by either the UX evaluator or the participant provide objective data. Empirical approaches are always related with objective data. Opinions, judgments, and other feedback are all examples of subjective UX data.
- Subjective data are judgments on the user experience and pleasure with the design expressed by either a UX evaluator or a participant. Methods of analytic UX evaluation produce only qualitative subjective data. Questionnaires provide quantitative and subjective information.

5.1.1.3 Formative Evaluation versus Summative Evaluation

- A long-standing dichotomy underpins the distinction between formative and summative evaluation :
 - Formative UX assessment is a type of diagnostic UX evaluation that collects qualitative data with the goal of forming a design, that is, detecting and fixing UX issues and refining the design.

- UX evaluation with the goal of summarizing or assessing the success of a UX design is known as summative UX evaluation.
- A apt way of looking at the distinction : When the cook tastes the soup, that's formative; when the visitors taste the soup, that's summative.
- The words formative and summative evaluation are used in the context of education and curriculum evaluation, which is the first usage we have. Formative assessment is primarily diagnostic in nature, with the goal of finding and correcting user experience problems and their sources in the design. Summative evaluation is primarily concerned with rating or scoring; it is concerned with gathering quantitative data in order to assess a design's level of quality.

Formal summative evaluation

- Both formal and informal methods are used in summative UX evaluation. An empirical technique that provides statistically significant results is a formal summative UX evaluation method. Because the technique is statistically rigorous, the term "formal" is utilized.
- To identify the "truth" in responses to science and research problems, formal summative investigations, inferential statistics, and statistically significant results are important in science. However, the majority of the work we perform in UX evaluation is more engineering than science, and getting to "truth" is a more practical and less precise process. Engineering is, in many respects, about making decisions based on hunches and intuition, which are founded on expertise and experience.
- Formal summative evaluation is based on an experimental design for controlled comparative hypothesis testing utilizing a m by n factorial design with y independent variables, with statistical tests for significance applied to the results. This requires specialized knowledge and abilities, so do not offer summative evaluation if you cannot follow through. Furthermore, doing a proper summative evaluation is both costly and time - consuming. To summarize, formal summative evaluation is a crucial research skill in HCI, but it is not part of UX practise in our opinion.
- Consider a particular button label as an example of a design modification that is probably not measurably better in terms of usability but is arguably better. If everyone on the team thinks that the old button label was unclear and confusing and the new button label is clear and easy to understand, the team should probably go ahead and make the design change.

- The rest of this book does not cover formal summative evaluation.

Informal summative evaluation

- A quantitative summative UX evaluation approach that is not statistically rigorous and does not provide statistically significant results is known as an informal summative UX evaluation method. As an engineering approach, informal summative evaluation is performed in support of formative evaluation to determine how effectively you are attaining good usability and UX.
- Informal summative evaluations are conducted without experimental controls, with fewer user participants, and merely summary descriptive statistics. The informal summative evaluation at the end of each iteration for a product version can be used as a kind of acceptance test to compare with our UX targets and ensure that the product design meets our UX and business goals.
- The contrasts between formal and informal summative UX evaluation approaches are highlighted in Table 5.1.1.

Formal Summative UX Evaluation	Informal Summative UX Evaluation
Science	Engineering
Randomly chosen subjects / participants.	To obtain the most formative information, participants were chosen in a nonrandom manner.
Concerned with having large enough sample size.	Deliberately uses relatively small number of participants.
Uses rigorous and powerful statistical techniques.	Simple, low-power statistical approaches are used on purpose.
Results can be used to make claims about "truth" in a scientific sense.	The findings cannot be utilized to make claims, but they can be used to make technical decisions.
Relatively expensive and time consuming to perform.	Relatively inexpensive and rapid to perform.
Rigorous constraints on methods and procedures.	Methods and procedures open to innovation and adaptation.
Tends to yield "truth" about very specific scientific questions (A vs. B).	Can provide insight into a larger variety of issues concerning the levels of UX that have been reached and the need for additional improvement.
Not used within a UX design process.	It is meant to be used as part of a UX design process to supplement formative methodologies.

Table 5.1.1 Formal vs informal summative UX evaluation methods

Engineering UX evaluation : Formative plus informal summative

- UX evaluation as an engineering process might comprise formative evaluation as well as an optional informal summative component (Fig. 5.1.2). The summative section cannot be used to prove anything, but it is an excellent resource for learning about the UX design process. Design reviews, heuristic approaches, and other UX methods are examples of evaluation methods.

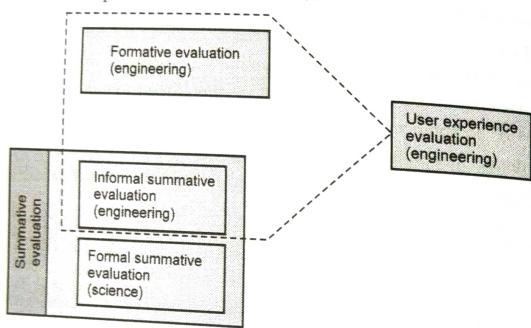


Fig. 5.1.2 Engineering user experience evaluation

- Procedures of inspection are excellent instances of completely formative evaluation methods.
- Empirical methods, such as user testing, might also be limited to a formative evaluation component, especially in the early stages of development, when we are still defining and refining the design and are not interested in performance indicators.

5.1.2 UX Evaluation Methods

- You can use either empirical or analytic UX evaluation approaches for formative UX evaluation.

5.1.2.1 Empirical UX Evaluation Methods

- Empirical approaches, by definition, are based on data collected from real-world user participants and data collected directly from them. These data contain crucial incident data from empirical evaluations as well as user remarks made while "thinking aloud" and / or in questionnaire responses.

- Empirical research can be carried out in a UX lab, at a conference table, or in the field. Empirical testing, such as user task performance, can generate both quantitative and qualitative data from the same measuring equipment.
- The UX lab is a more or less controlled environment which is beneficial in terms of avoiding distractions, but testing in a field setting can be more useful in terms of guaranteeing actual task conditions for ecological validity.

5.1.2.2 Analytic UX Evaluation Methods

- Analytic techniques are based on looking at the design's fundamental characteristics rather than observing it in action. Analytic approaches produce qualitative subjective data, with the exception of numerical ratings and related data.
- Analytic UX evaluation approaches were created as faster and less expensive methods to produce approximations to or predictors of empirical outcomes.
- Design reviews, design walkthroughs, and inspection methods such as heuristic evaluation are examples of analytical methods.

5.1.2.3 Comparison

- Because they are focused on how a design or design change pays off in real observable usage, empirical approaches are also referred to as "payoff methods". Because they are focused on examining intrinsic design aspects, analytical methods are also referred to as "intrinsic methods".
- In practice, several strategies combine analytic and empirical approaches. Expert UX inspection, for example, can include "simulated empirical" features in which the expert assumes the position of the user while executing tasks and "observing" UX issues.
- If you want to evaluate a tool, say an axe, you could look at the design of the bit, the weight distribution, the steel alloy used, the grade of hickory in the handle, and so on, or you could just look at the kind and speed of the cuts it produces in the hands of a good axeman.
- This example allows us to make a point about the need of clearly defining UX goals before developing evaluation criteria. Giving the axe example a UX viewpoint, we find that user performance monitoring in payout evaluation does not always necessitate a good axeman.
- The intended user classes of important job roles, as well as the expected type of usage, determine UX goals. An axe design that performs optimally in the hands of

an expert, for example, may be too deadly for a rookie user. For the novice user, safety may be a UX goal that extends beyond firewood production, necessitating a safer design at the expense of efficiency.

5.1.2.4 Some Specific Empirical UX Evaluation Methods

Lab-based evaluation

- An empirical UX evaluation method based on observing user participants executing activities in a UX laboratory context is known as lab-based UX evaluation. It collects qualitative and quantitative data using critical incident identification and think-aloud approaches.
- User participants are observed completing representative tasks in a lab-based empirical UX evaluation. When qualitative data is obtained, it may be used to indicate UX issues that need to be addressed, and quantitative data can be used to analyse how well people perform with a given design.

RITE

- One of the greatest rapid empirical assessment approaches is Rapid Iterative Test and Evaluation (RITE), a fast user-based UX evaluation strategy designed to harvest the low-hanging fruit at a minimal cost.
- The key to quick iteration with RITE is to address issues as soon as they arise. RITE is one of the most agile empirical approaches because of the quick turnaround of qualitative evaluation data and problem solving.
- While the team is still present, problems are reported, and they are already informed and involved in the process.

5.1.2.5 Weaknesses of UX Evaluation Methods

Measurability of user experience

- Quantitative evaluation of a feature like usability or User eXperience (UX) necessitates some form of measurement. Can you, however, quantify usability or user experience? It may come as a shock, but neither usability nor user experience can be measured directly. In fact, most fascinating phenomena, such as teaching and learning, face the same challenges.
- As a result, we revert to measuring what we can measure and using those measurements as markers of our more abstract and less measurable concepts. For example, we can measure observable user performance-based indicators like time to

finish a job and counts of errors encountered by users during task performance to better understand usability effects like productivity and ease of use.

- Questionnaires can also provide indices of user happiness by eliciting responses to questions that we believe are closely related to their perceptions of performance and satisfaction. Emotional impact variables like satisfaction and joy of usage, on the other hand, cannot be quantified directly and must rely on indirect signs.

Reliability of UX evaluation methods

- In simple terms, a UX evaluation method's reliability refers to its repeatability, which is a concern with both empirical and analytic methodologies.
- It means that you won't obtain the same list of UX problems each time you employ the same formative evaluation approach with multiple different user participants or several different UX inspectors. In fact, the differences can be rather significant. We have to make do with unreliable reliability.
- The good news is that even low-reliability UX evaluation methods and procedures can be quite effective—that is, the methodologies can still uncover UX problems that need to be fixed, and they often find the most critical problems. Low reliability is not usually a severe disadvantage; in UX practise, much of each iteration of formative evaluation is about learning as much about the design as possible for the least amount of money and then moving on.
- So, while complete reliability would be ideal, the procedure nevertheless works in practise if the method is relatively successful. You will get a list of UX issues every time you complete formative evaluation. Your list will be more complete and accurate if you apply additional rigour to the evaluation technique. You will obtain a new list if you cure all of those issues and run formative evaluation again. Most UX issues, especially the major ones, can be found and fixed eventually. Engineering is all about achieving a close approximation to the ideal.

5.1.2.6 Some Specific Analytic UX Evaluation Methods

Early design reviews and design walkthroughs

- Early design reviews and design walkthroughs are UX team demos of the design to elicit early reactions and input from team members and other stakeholders, such as users and representatives from the client organization. Because they are focused on explanations of how the design works rather than actual user usage, we describe these as analytic methods. Scenarios and storyboards for evaluating the ecological

view, or conceptual design and screen sketches for task-level evaluation—nothing interactive in the early presentations.

- These materials will quickly transform into clickable wireframe prototypes. It is up to you, the UX team, to conduct the "driving" to show interaction and navigation; it's too early for anyone else in a user position to do so. The group is led through important workflow patterns that the design is meant to enable by the leader. This will include an overview, a flow model, and a conceptual design in the early funnel stage of the lifecycle process.
- In the late funnel, the focus will be on the interaction design of a single feature at a time. The leader presents tales about users and usage, user goals and behaviours, and expected outcomes while the team follows the scenarios, looking systematically at aspects of the design and analyzing the benefits and potential concerns.

Expert UX inspection

- Expert UX inspection is a way of rapid analytic review. During an in-depth evaluation of the design, expert UX inspectors use their professional experience and knowledge of UX design rules to discover UX flaws. They also frequently imitate real-world usage by acting as users and completing crucial tasks in the search for issues.
- Although the UX inspector may be an expert in UX, he or she may not be an expert in the system in question or the job domain in which it is related. In these circumstances, the UX inspector can use their unfamiliarity to identify difficulties for new users, or they can collaborate with a subject matter expert.

Heuristic Evaluation (HE)

- The most well-known and often used inspection method is the Heuristic Evaluation (HE) method. Inspectors using the HE approach are directed by a list of roughly 20 experimentally derived "heuristics" or rules that govern successful UX design.
- The team's UX experts conduct an expert UX inspection, determining how effectively each of these guidelines is adhered to in the design. The HE technique is particularly beneficial for use early in the UX process because it is affordable, intuitive, and easy to convince practitioners to use.

5.1.3 Rigor vs Rapidness in UX Evaluation Methods and Techniques

- The connection between technique rigour and speed is complex:
- There is a compromise between the rigour with which any procedure is applied and the speed with which it can be accomplished.

TECHNICAL PUBLICATIONS

- All methods can be used to a variety of levels of rigour.
- Excessive rigidity isn't usually the purpose.
- Some strategies were devised to prioritize speed above accuracy.

5.1.3.1 Tradeoff between Rapidness and Achievable Rigor

- In general, implementing a more rigorous evaluation approach (or any method) can result in more complete and accurate results, but it will take longer and cost more money. Similarly, you can usually speed up and lower the cost of practically any UX evaluation approach by employing shortcuts, but at the sacrifice of rigour.

5.1.3.2 Range of Rigor and Speed

- Each UX review method has a different level of rigour. For example, by avoiding shortcuts and retaining all data, you can undertake a lab-based empirical approach in a highly rigorous manner that will increase effectiveness and decrease the possibility of errors.
- When high rigour is not necessary, lab-based empirical evaluation can be done quickly and with a variety of shortcuts. You can enhance efficiency by filtering and abstracting the evaluation data down to the most significant points.
- Similarly, an analytic procedure can be carried out quickly and with little discipline, or it can be carried out with great rigour, paying close regard to the sources of data, completeness, and purity.

5.1.3.3 Methods to Favor Rigor over Rigor

- There are decisions to be made to match your need for rigour because not all procedures cover the same spectrum of potential rigour. While not perfect, empirical UX evaluation procedures that are carried out with a high level of rigour have long been considered the gold standard in terms of approach effectiveness.
- Other UX evaluation methods, such as analytic methods, were created particularly to be quicker and less expensive alternatives to completely rigorous empirical methods. Analytic approaches are intended to be quick methods for estimating what truly matters, such as UX issues that can be discovered empirically.
- So, there are two ways you can view the rigor of a UX evaluation method:
- The rigor with which any given method is applied.
- The range of rigor inherent in the method itself.
- The range of rigor inherent in the method itself.
- Design reviews, walkthroughs, and UX inspections are popular choices for late-

funnel task-level evaluation since they can be completed quickly. RITE and quasiempirical approaches are two empirical methods that are expressly meant to be quick.

5.1.4 UX Evaluation Data Collection Techniques

5.1.4.1 Quantitative Data Collection Techniques

Objective data : User performance measures

- User performance measures taken during empirical UX testing with user participants are used in several quantitative data collection methodologies. Users complete benchmark tasks, and UX evaluators track objective metrics like task completion time.

Subjective data : User questionnaires

- Questionnaires or user surveys are used in other quantitative data gathering procedures to get subjective data about how people perceive the design.
- Questionnaires can be used as a stand-alone evaluation tool or to enhance your objective UX evaluation data with subjective data from the user. Questionnaires are easy to use for both the analyst and the subject, and they can be utilized in conjunction with or without a lab.
- Questionnaires are effective for assessing specific aspects of the user experience, such as perceived usability, utility, and emotional impact.

Warning : Modifying a questionnaire can damage its validity

- A questionnaire's validity is a statistical property that is particularly important in summative investigations. Ready-made questionnaires are frequently designed and statistical validity checked thoroughly. There are a variety of surveys for evaluating usability, usefulness, and emotional impact that have already been designed and verified.
- However, if you want or need to adapt an existing questionnaire to meet special demands, do not be concerned that doing so may compromise its validity; questionnaire validity is rarely a practical concern in UX practise.
- The greater the number of changes, the greater the risk. The methodologies for questionnaire validation, as well as the difficulties surrounding questionnaire validation are beyond the scope of this work.
- Homemade questionnaires and unvalidated questionnaire revisions are not allowed in summative evaluation because of this risk to validity but they are frequently used

in formative evaluation. This is not an invitation to be sloppy; rather, we are allowing ourselves to avoid having to go through validation for responsible changes.

5.1.4.2 Qualitative Data Collection Techniques

- To acquire data for UX problem diagnosis, qualitative data collecting techniques are utilized. Among the most prominent qualitative data collecting approaches are critical incident identification, think aloud and codiscovery.

Critical incident identification

- Critical incident identification is a qualitative UX data gathering technique in which the UX team observes user participants executing activities and detects "critical incidents," or instances in which users experience UX issues. Problems are then tracked back to their sources in the UX design and placed on a to - do list for future iterations.

User think - aloud techniques

- The think - aloud technique is frequently used in conjunction with critical event identification as a secondary method of detecting UX issues during user task execution. Users are encouraged to voice their ideas verbally while performing tasks and putting the UX design to the test, exposing otherwise concealed qualitative data concerning UX issues.
- The think - aloud technique is likely the most beneficial of all UX evaluation tools since it captures the user's mental state at the moment of use.

5.1.5 Specialized UX Evaluation Methods

- There are a variety of specific methods and strategies in addition to the "normal" UX evaluation methods and techniques.

5.1.5.1 Alpha and Beta Testing and Field Surveys

- Post-deployment evaluation approaches like as alpha and beta testing are useful. Manufacturers of software applications sometimes send out alpha and beta versions of the application software to select users, experts, customers, and professional reviewers as a preview after practically all development is completed.
- Users check it out in exchange for an early glimpse and provide comments on the experience. Beyond survey questions like "tell us what you think is good and

terrible and what needs to be fixed, what extra features you would like to see, etc., there is often little or no advice provided for the review process.

- An alpha version of a product is a less polished, early version that is usually tested with a smaller, more trusted "audience." The beta version is as near to the final product as they can get it and it is distributed to a bigger audience. Most businesses create a beta trial email list of early adopters and expert users who are renowned for being friendly to the company and its products and helpful in their feedback.
- Alpha and beta testing are quick and low - cost ways to receive high - level feedback based on real - world usage. However, alpha and beta testing fall short of being considered formative evaluation because :
 - A traditional formative evaluation process does not provide the kind of granular UX problem data that a UX problem analysis does.
 - When flaws are discovered, it is frequently too late to make any meaningful changes to the design.
- Alpha and beta testing are highly customized to a certain development team and environment. We are unable to provide detailed instructions on how to conduct alpha and beta testing.
- User field survey information, like alpha and beta testing, is retrospective and while it can be useful for determining user happiness, it does not capture the details of use inside the usage experience.
- Anything is better than nothing, but don't let these after - the - fact procedures be the only formative evaluation methods used in your organization's product lifecycle.

5.1.5.2 Remote UX Evaluation

- Methods for evaluating systems after they have been deployed in the field are useful. The following are some examples of methods :
 - Using the internet as a long extension chord to the user to simulate lab - based UX testing.
 - After - the - fact feedback can be obtained through online surveys.
 - Clickstream and usage event data are recorded using software.
 - Plug - ins for software that allow users to self-report UX concerns.

- The latter technique relies on users self - reporting UX issues as they arise during normal use, allowing you to get at the perishable elements of the user experience, particularly in real - world everyday work usage.
- As is always the case, the best input for design enhancement comes from inquiry within experience, or formative data collected while the product is being used rather than retroactive recall.

5.1.5.3 Automatic UX Evaluation

- Labor - intensive lab - based and UXinspection approaches are thus limited in scope. High and complicated systems with a large number of users, on the other hand, have the potential to generate a massive amount of usage data. Consider "observing" a hundred thousand Microsoft Word users.
- Automatic ways for collecting and analyzing usage data have been designed to take advantage of this limitless source of data, eliminating the need for UX specialists to deal with each individual action.
- Multiple versions of a product are sometimes distributed to various groups of users, and the data collected is compared to see which version is better. A - B testing is a term used to describe this form of evaluation, where A and B represent two design variations. As a result, there is a lot of information about keystrokes, clickstreams, and pause / idle times. However, all of the data is at the level of user activities, with no information on tasks, user intentions, cognitive processes and so on. In the midst of that stream of user activity data, there are no direct signals of whether the user is facing a UX difficulty.
- Within a large software application, basing redesign on click counts and low-level user navigation could lead to low-level optimization of a system with a suboptimal high - level design.

5.1.6 Adapting and Appropriating UX Evaluation Methods and Techniques

- It is not enough to select "standard" UX evaluation methodologies and procedures that are appropriate for your objectives and restrictions. It will come naturally to you to adjust and tune the approaches and processes to the specifics of your team and project until you have adopted them as your own for each new design circumstance.

- Our motto for UX evaluation, as well as maybe any UX work, is improvise, adapt, and overcome! Be adaptable and tailor your approaches and strategies to match your evaluation goals and requirements. This includes removing steps, introducing new steps and modifying the details of a step in any procedure.

5.2 Empirical UX Evaluation : UX Goals, Metrics, and Targets

5.2.1 Introduction

- Empirical UX evaluation stage of the overall UX design lifecycle template is shown in Fig. 5.2.1. Goals, metrics, and targets for UX help create a framework for evaluation planning that will successfully disclose issues with user performance and emotional satisfaction. If used, UX goals, metrics, and targets are established early in the assessment process as part of the preparation process, and can help to steer much of the process from analysis through evaluation.

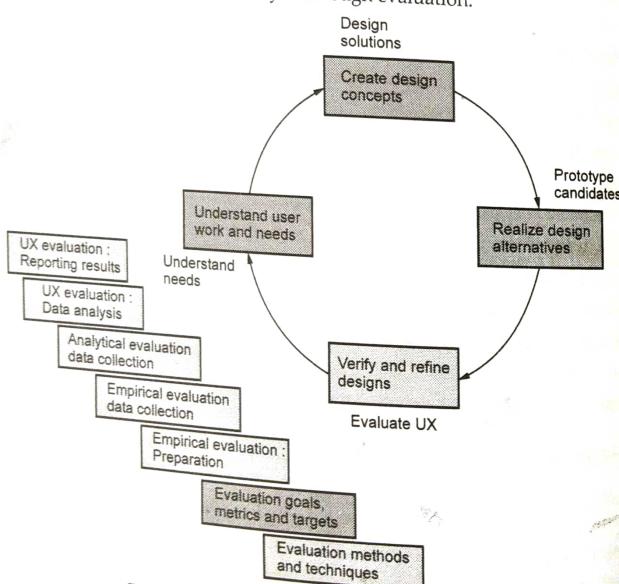


Fig. 5.2.1 UX evaluation goals, metrics and targets

5.2.1.1 Project Context for UX Metrics and Targets

- In the early stages, the focus of evaluation is mainly on qualitative data in order to identify UX issues. The lack of quantitative data prevents the adoption of UX metrics and targets in these early reviews. If you plan to utilize them in future evaluations, you can start establishing them at any time.
- However, you may want to avoid using UX measurements and KPIs entirely. Specifying UX measurements and targets, as well as the matching rigorous evaluation, may be prohibitively costly in most practical situations. Only a few firms with large established UX resources can achieve this level of completeness. In many projects, there is only one round of evaluation. Also, by looking at the results of the first round of evaluation, we can see which aspects of the design require further examination as designers. In such instances, quantitative UX metrics and targets may be ineffective, although benchmark tasks can still be used to evaluate driving performance.
- In any case, the UX sector is shifting away from quantitative user performance indicators and toward speedy qualitative evaluations of usability, user happiness, and enjoyment.
- In any case, we have found that defining UX goals, measurements, and targets is frequently missed, either due to a lack of understanding or time constraints. This is sometimes undesirable since it limits the scope of what you can achieve with the resources you will devote to user experience evaluation.
- Fortunately, after some practice, generating UX measurements and targets does not take long. Rather than merely waiting to see what happens when users interact with your UX design, you will have explicit quantified UX goals against which to test. The results can help you define where to focus on redesign most profitably because UX measurements and targets provide attainable objectives for formative UX evaluation activities.
- Finally, by setting a tangible end to what would otherwise feel like unending iteration, UX goals, metrics, and targets can assist manage the lifecycle. Designers and managers may run out of time, money or patience before achieving their UX goals - sometimes after only one round of evaluation-but at least they will know where they stand.

5.2.2 UX Target Tables

- The concept of a UX goal table is improved, as illustrated in Table 5.2.1, through years of working with real-world UX specialists and doing our own user experience evaluations.

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							

Table 5.2.1 UX target table

- One row in the table is designated as a "UX target" for ease of use. The work role and corresponding user class to which this UX target relates are listed in the first column. The linked UX goal and corresponding UX measure are listed in the following two columns.
- Because each UX measure supports a UX goal and is described with respect to a work position and user class combination, they all go together. Following that, you will see where you receive the data for these three columns.
- We will gradually create some UX targets for the new Ticket Kiosk System as a running example of how to use each field in the UX target table.

5.2.3 Work Role and User Classes

- We categorize each UX target by work job because UX targets are aimed at certain work roles. Remember that in the user models, various work positions do different task sets.
- As a result, key task sets for each work role will have associated usage scenarios or other task sequence representations, which will inform the benchmark task descriptions we create as measuring instruments for UX targets. Different user classes will be expected to perform to different standards, or at different target levels, within a given work role.

Example : A Work Role and User Class for the Ticket Kiosk System

- Let's start with the user work role of the ticket buyer in the Ticket Kiosk System. User class definitions for a job role can be based on a variety of factors, including level of knowledge, disabilities and limits, and other demographics, as we saw earlier.

- Choosing work role and user class for a UX target is shown in Table 5.2.2.

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							

Table 5.2.2 Choosing work role and user class

5.2.4 UX Goals

- User experience goals are high - level goals for a UX design that are expressed in terms of user experience objectives. UX goals can be influenced by business objectives and reflect actual product usage, as well as determine what is vital to a company, its customers and its users.
- They are expressed as desirable impacts that users of design features should experience when using them and they are translated into a set of UX measures that can be evaluated.
- User concerns collected in work activity notes, the flow model, social models and work objectives can all be used to extract UX goals, some of which will be market driven and reflect competitive imperatives for the product. Goals for user experience might be articulated in terms of all users, a certain job role or user class, or specific types of tasks.
- Ease of use for all users, ease of remembering for intermittent users, power performance for experts, avoiding errors for safety - critical systems, high customer happiness, walk-up-and-use learnability for new users, and so on are some examples of user experience goals.

Example : User Experience Goals for the Ticket Kiosk System

- We can outline the major high-level UX goals for the ticket buyer based on our usage research results, which include :
 - Quick and simple walk-up-and-use user experience with no user training required.
 - Fast learning ensures that the performance of new users (with little experience) is comparable to that of experienced users.
 - Customer satisfaction is strong, which leads to a high rate of repeat business.

- Some other possibilities :
 - High learnability for more advanced tasks.
 - Draw, engagement, attraction.
 - Low error rate for completing transactions properly, particularly in the interaction for payment.

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							

Table 5.2.3 Choosing UX goal

- It is simple to turn the aim of "quick and easy walk-up and use user experience" into a UX target table entry for the UX goal. This goal relates to a normal infrequent user's ability to complete at least the most fundamental activities on the first try, without any training or manuals. Table 5.2.3 shows the beginnings of a UX objective.

5.2.5 UX Measures

- The UX measure is the general user experience characteristic to be measured with respect to the use of your UX design inside a UX target. The sort of measuring instrument and UX metrics acceptable for a given UX measure is implied by the choice of UX measure.
- Quantitative data - both objective data, such as observable user performance, and subjective data, such as user opinion and satisfaction-is used to set UX goals.
- The following are some examples of common UX measurements that can be combined with quantitative metrics :
 - Objective UX measures :
 - Initial performance.
 - Long-term performance.
 - Learnability.
 - Retainability.
 - Advanced feature usage.

- Subjective UX measures :
 - First impression.
 - Long-term user satisfaction.
 - Emotional impact.
 - Meaningfulness to user.
- The performance of a user at the first use is referred to as initial performance. Because every user of a system must use it for the first time at some point, initial performance is a critical UX metric.
- Long-term performance usually refers to performance over a longer length of time with more consistent use. Long - term usage usually indicates that the user has reached a steady - state learning plateau; the user has become comfortable with the system and is no longer in a learning mode.
- Learnability and retainability refer to how quickly and readily people can learn to utilize a system, as well as how well they remember what they have learned over time.
- Advanced feature utilization is a UX metric that assesses the user experience of a system's more complex features. A first impression UX metric can capture the user's initial perception of the system, whereas long-term user satisfaction relates to the user's opinion after using the system for a longer amount of time, after some learning time.
- For almost any UX design, first performance and first impression are relevant UX measurements. Other UX measurements are frequently used as a supplement to fulfil more specialized UX needs. Conflicts between UX measures are not uncommon. You may, for example, require both strong learning ability and expert performance. Those needs may conflict with one another in the design. This, on the other hand, is merely an example of a common design tradeoff. User performance requirements based on the two UX measurements indicate user performance tugging in two distinct ways, compelling designers to stretch the design and face the choice honestly.

Example : UX measures for the ticket kiosk system

- Let's start with just two UX measures for our casual new user's walk - up ease - of - use goal : Initial performance and first impression. In the UX goal table, each UX measure will have its own UX target, with the work role and user class repeated, as shown in Table 5.2.4.

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 ease of use for new user	Initial user performance				
Ticket buyer : casual new user, for occasional personal use	Initial customer satisfaction	First impression					

Table 5.2.4 Choosing UX measures

5.2.6 Measuring Instruments : Benchmark Tasks

- A measuring instrument is the mechanism of generating values for a specific UX measure inside a UX target; it is the vehicle through which values for the UX measure are measured.
- Although you can be creative with your measuring tools, objective metrics are frequently used in conjunction with a benchmark task - subjective measures are frequently associated with a user questionnaire-for example, the average user rating - scale scores for a specific set of questions-and objective measures are commonly associated with a user questionnaire-for example, a time-on-task measure as timed on a stopwatch or an error rate measure as counted user errors.
- For example, in the UX goal Table 5.2.3 for the Ticket Kiosk System, the objective "first user performance" UX measure is linked to a benchmark test, whereas the "first impression" UX measure is linked to a questionnaire. For developing and assessing user experience resulting from a design, both subjective and objective measures and statistics might be useful.

5.2.6.1 What is a Benchmark Task ?

- A benchmark task is a representative task that you will have user participants attempt to complete in assessment while you watch their performance and behaviour as a measuring instrument for an objective UX metric.
- As a result, a benchmark task is a "standardised" activity that can be used to compare performance across users and design versions.

5.2.6.2 Selecting Benchmark Tasks

- Here are some suggestions for selecting benchmark projects.

Address designer questions with benchmark tasks and UX targets

- As designers work on UX designs, they are continually bombarded with inquiries. When the design team is unable to make a decision for itself, they defer to UX testing. Or perhaps you agree on the design of a feature but are wondering about how it will be used by real people.
- If you kept a list of such design questions as they arose during design activities, they may now be used to set benchmark assignments for gathering user input.

Create benchmark tasks for a representative spectrum of user tasks

- Choose actual tasks that will be used across the system by each user class of a work role. Your choices should represent a cross section of genuine jobs in terms of frequency of performance and criticality to the aims of users of the envisioned product to get the most out of your evaluation expenditure.
- New features, "edge cases," and business - critical or mission - critical jobs are all evaluated using benchmark tasks. While some of these duties aren't carried out on a regular basis, getting them wrong can have major implications.

Include some navigation where appropriate

- Necessity for navigation should be included in initial benchmark tasks since users frequently have to navigate to get to where they will execute the operations unique to executing a task in real-world usage. It assesses their understanding of why they need to go somewhere else, where they need to go and how they will get there.

Match the benchmark task to the UX measure

- Obviously, if the UX measure is "initial user performance," the task should be among those a first - time user realistically would face. If the UX measure is about advanced feature usage, then, of course, the task should involve use of that feature to match this requirement.

- If the UX metric is "first user performance," the task should clearly be one that a first-time user would encounter. If the UX metric is advanced feature utilization, the task should, of course, include the use of that feature to meet this criteria.

Use tasks in realistic combinations to evaluate task flow

- Use combinations of tasks and activities that will occur together frequently to measure user performance related to task flow. In these circumstances, you should create UX targets for such combinations since user experience issues that arise during the performance of the combined tasks may differ from those that arise when the identical actions are completed individually.
- You could want to monitor user performance on the task thread of searching for an event and then purchasing tickets for that event in the Ticket Kiosk System, for example.

Evaluate with your power users

- When it comes to product testing, the user experience for power users is frequently overlooked. Do your product business and UX objectives incorporate the use of power by a trained user base?
- Do they require assistance with tasks that require quick repetition or that are complex and maybe quite long? Does their quest for efficiency necessitate the use of shortcuts and direct commands rather than interactive handholding?
- If any of these statements are correct, you must include benchmark tasks that correspond to this level of skilled and demanding power utilization. These benchmark tasks must, of course, be utilized as the measurement instrument in UX targets that correspond to the respective user classes and UX goals.

To evaluate error recovery, a benchmark task can begin in an error state

- Error recovery is a type of "feature" that designers and assessors sometimes overlook. However, no UX design can ensure error-free use, and most users are familiar with and can connect to the process of recovering from faults.
- Users will be able to recover from errors more easily if the design is "forgiving." This is undoubtedly a feature of your design that should be assessed using one or more benchmark tasks.

Example : Benchmark tasks as measuring instruments for the Ticket Kiosk system

- The first UX target in Table 5.2.4 for the Ticket Kiosk System includes an objective UX measure for "Initial user performance." A benchmark job is a natural choice for

the corresponding measuring instrument. In a walk-up ease-of-use situation, we require a basic and commonly utilized job that can be completed in a short amount of time by a casual new user.

- Purchasing event tickets might be an acceptable benchmark task. Here is an example of a possible description for the user participant:
- BT1 : Purchase three tickets for the Monster Truck Pull on February 28 at 7 p.m. at the Ticket Kiosk System. As close to the front as feasible, group three seats together. Use a major credit card to pay.**
- This is added to Table 5.2.5 as the measurement instrument for the first UX target.
- Let's imagine we want to add another UX objective to the "initial performance" UX measure, but this time we want to spice things up by using a different benchmark task as the measuring instrument--namely, the task of purchasing a movie ticket. We've put this benchmark task in the second UX objective in Table 5.2.6, lowering the "first impression" UX target by one.

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					

Table 5.2.5 Choosing benchmark task as measuring instrument

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				

Agile Development and UI / UX Design				5 - 26	UX Evaluation Methods and Techniques
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			

Table 5.2.6 Choosing another benchmark task as measuring instrument

5.2.6.3 Crafting Benchmark Task Contents

Remove ambiguities

- We must make the directions in benchmark task descriptions explicit and not confusing unless we want users to resolve ambiguity as part of the work. For consistent findings, unambiguous benchmark tasks are required; we want discrepancies in user performance to be due to changes in users or in designs, not to diverse interpretations of the same benchmark work.
- Consider this "add appointment" benchmark assignment for an interdepartmental event scheduling system's "initial performance" UX measure: "Schedule a meeting with Dr. Arvind for a month from today at 10 a.m. in GCC Hall concerning the HCI research project."
- The term "a month from now" can be confusing for certain people. Why? It may be on the same date next month, or it could be four weeks from now, on the same day of the week. If the discrepancy in meaning affects the user's work performance, the wording should be more specific to the intended meaning.
- You should also make your benchmark activities focused so that participants aren't distracted during testing by extraneous elements. If a "find event" benchmark task is simply stated as "Find an entertainment event for sometime next week," some participants may make it a long, involved task, searching for the "best" combination of event type and date, whereas others may do the bare minimum and take the first event that appears on the screen. Add detailed information regarding event selection criteria to mitigate such disparities.

Agile Development and UI / UX Design

5 - 27

UX Evaluation Methods and Techniques

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

- This recommendation is crucial; it will determine the success of your task-based evaluation. In early evaluation activities, we occasionally see students giving users task instructions that lay out a sequence of actions to follow. They should not be surprised if the results of the evaluation session are dull.
- The users are simply repeating the procedures verbatim as they read them from the benchmark task description. If you want to see if your UX design helps consumers figure out how to execute a task on their own, you should avoid offering them any instructions. Just tell them what to do and leave it up to them to figure out how to execute it.

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

- You must avoid using any words that occur in menu headlines, menu options, button labels, icon pop-ups, or anywhere else in the UX design in your benchmark work descriptions. For example, if there is a button in the UX design labeled "Find", don't say "Find the first event (that has such and such a characteristic)". Instead, use phrases like "search for..." or "locate..."
- When your users are told to "Find" something, it is highly handy for them to use a button labeled "Find." It does not require them to think and hence does not assess whether the design would have assisted them in locating the correct button on their own during actual use.

Have clear start and end points for timing

- Make sure that each benchmark task has clearly observable and recognizable start and end points in your mind, and that the benchmark task description is written so that these end points may be used successfully. These will, for example, assure your ability to precisely measure time spent on a task.
- Not only must the assessors know when the work is completed with certainty, but the participant must also know when the task is completed. The task cannot be considered finished for evaluation purposes until the user has experienced closure.
- The evaluator must also be aware of when the user realizes the task is finished. Even if you expressly ask for it in the benchmark task description or user instructions, do not rely on the user to tell you when the work is finished. As a result, rather than concluding task execution with a mental or sensory state, it is preferable to include a user action verifying the task's completion.

Keep some mystery in it for the user

- Do not get too explicit about what users will see or the parameters they will face all of the time. Remember that true first-time users will come to your app with no prior knowledge of how it works. Try to employ benchmark jobs that provide estimated values for some parameters to seek for, leaving the rest to the user's discretion. If you wish to avoid different users in the evaluation ending up in different states in the system, you can still develop a prototype in which there is only one possible "solution" to this task.

Annotate situations where evaluators must ensure preconditions for running benchmark tasks

- You must attach a note in the form of rubrics to this benchmark task to that effect-a note that will be read and followed in the evaluation activity.
- The prototype calendar must start on the same "current" date every time a user does this action during assessment, and it must contain an existing appointment at some future date in the calendar so that each user may discover and cancel it. To that end, you must attach a note in the form of rubrics to this benchmark task-a remark that will be read and followed during the evaluation activity.

Use "title" for special instructions to evaluators

- Add a "title" portion to your benchmark task descriptions when needed or beneficial as special instructions to evaluators, but not to participants in evaluation sessions. Use these rubrics to express anything that needs to be done or set up ahead of time to establish task preconditions, such as an existing event in the kiosk system, work environment for ecological validity, or a specific beginning state for a task.
- Titles are very well suited to benchmark activities that answer designer questions. You can notify evaluators to check for user performance or behavior that can give light on these specific designer questions in a note that comes with your benchmark work.

5.2.7 Measuring Instrument : User Satisfaction Questionnaires

- A questionnaire connected to various user UX design characteristics can be used to gauge a user's happiness with the UX design as a measuring instrument for a subjective UX measure. The satisfaction of a user gives a subjective but yet measurable, UX indicator for the linked UX measure. It is worth noting, though, that objective and subjective metrics aren't always orthogonal. For example, low user satisfaction, can have a long-term negative impact on user performance. The QUIS questionnaire is used in the following examples, however there are other good options, such as the System Usability Scale or SUS.

Example : Questionnaire as Measuring Instrument for the Ticket Kiosk system

- If you believe that the first two benchmark tasks (buying tickets) provide a good foundation for evaluating the "first-impression" UX measure, you can specify that a specific user satisfaction questionnaire or a subset thereof be administered after those two initial tasks, specifying it as the measuring instrument in the Table 5.2.7 third UX target of the growing UX target table, as we did in the previous example.

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	Questionnaire of user interface satisfaction				

Table 5.2.7 Choosing another benchmark task as measuring instrument

5.2.8 UX Metrics

- A UX metric is a term that indicates the type of value that may be obtained from a UX metre. It specifies the unit of measurement. For every given measure, there might be several metrics. Software complexity is a measure in the software engineering world, and one metric for the software complexity measure is "counting lines of code."
- The most popular UX metrics are objective, performance-based, and collected while the participant is performing a benchmark job. Other UX indicators can be subjective, based on a rating or score derived from survey responses.
- Time to perform task and the amount of errors made by the user are two common objective UX metrics. Others are the number of commands, mouse clicks, or other user actions required to complete a task; time spent in errors and recovery; number of repetitions of failed commands; and the number of commands, mouse clicks, or other user actions required to complete a task.
- If you are feeling brave, you can utilize the number of times the user expresses annoyance or satisfaction during his or her first session as an indicator of his or her initial assessment of the UX design. Because the amount of remarks is proportional to the length of the session, design your levels accordingly, or set them as a count per unit time, such as comments per minute, to account for time disparities.
- This metric is admittedly participant-dependent, as it depends on how demonstrative a participant feels during a session, whether a person is a complainer in general, and so on, but it can yield some intriguing results.
- Subjective UX metrics typically convey the type of quantitative result you are looking for from a survey, and are usually based on simple arithmetic statistical measurements like the numeric average. Keep in mind that you are simply looking for a technical indicator of user experience, not statistical significance.
- In circumstances when there are performance tradeoffs, do not forget to use a combination of measures. If you provide your UX measure as a function of two other performance-related metrics, such as time on task and error rate, you are indicating that you're ready to sacrifice some performance in one area in exchange for better performance in the other.
- We hope you will look at a variety of other UX metrics, in addition to the ones we've given here, such as:

- The percentage of a task that is accomplished in a particular amount of time.
- The success - to - failure ratio.
- Spending time moving the cursor.
- "Eye tracking can be used to assess visibility and other difficulties, such as screen fixations, cognitive load as indicated by a correlation to pupil diameter, and so on."
- Finally, ensure sure your UX measures, measuring tools, and metrics all make sense in the context of a UX goal. For example, do not term a UX metric "initial performance" if you plan to utilise a questionnaire in a UX goal. A questionnaire does not assess performance; rather, it assesses user happiness and opinion.

Example : UX metrics for the Ticket Kiosk System

- The amount of time it takes to buy a special event ticket is a good value to measure for the initial performance UX measure in the first UX target. We express this by include average "time on job" as a metric in the first UX target of Table 5.2.8.
- You may use the average number of errors a user makes while purchasing a movie ticket as a different objective performance measure. This was chosen as the value to measure in second UX target of Table 5.2.8. During a participant's single performance of the same single job, you will often wish to assess both of these variables.
- A participant does not, for example, need to complete one "buy ticket" activity while timing performance and then another "purchase ticket" task while counting errors.
- Finally, let us utilize the simple average of the numeric ratings supplied across all users and over all questions for which ratings were given for the UX metric in the third UX target of Table 5.2.8, the subjective UX target for the 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	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	Questionnaire of User Interface Satisfaction	Average rating across users and questions

Table 5.2.8 Choosing UX metrics for UX measures

5.2.9 Baseline Level

- The baseline level is the benchmark level of UX metric; it serves as a "talking point" against which other levels are measured. It is frequently the level that has been measured for the current system version.

5.2.10 Target Level

- The target-level specification for a UX metric is a numerical expression of an aimed-at or hoped-for value. The goal level is a criterion for the expected user experience's success that is operationally defined.
- The minimal value showing user experience success is the target level for a UX statistic. Designers might use the target levels that were not met in the evaluation as a focal point for improvement.

5.2.11 Setting Levels

- The UX target table's baseline and target levels are critical for assessing user experience metrics. However, determining a baseline and target level might be difficult at times. To find the solution, you must first figure out what level of user performance and experience the system will support.

- Although level values are frequently "best guesses," UX professionals can become quite adept at generating realistic and credible target levels and setting suitable values with practise.
- The following are some of the yardsticks you can use to determine your baseline and target levels :
 - A earlier version of the new system being designed or an existing system.
 - System competitors, such as those having a substantial market share or a well-received user experience.
- The baseline and target values presented in a UX target table refer to the mean over all participants of the associated measure, even if it is not always explicitly stated.
- That is, not every participant in every session must achieve the levels stated. As an example, if we set a target level of four errors as the worst acceptable level of performance for benchmark task BT 2 in the second UX target of Table 5.2.9, there must be no more than an average of four errors among all participants who complete the "purchase movie ticket" task.

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	Three minutes		
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		
Ticket buyer : casual new user, for occasional personal use	Initial customer satisfaction	First impression	Questionnaire of user interface satisfaction	Average rating across users and questions	75 / 10		

Table 5.2.9 Setting baseline levels for UX measures

5.2.11.1 Setting the Baseline Level**Example : Baseline level values for the Ticket Kiosk system**

- We can have someone complete the benchmark activities for buying a ticket for a special event and a movie using the MUTTS ticket counter to establish the values for the first two UX target baseline levels for the Ticket Kiosk System.
- That may be quite different from what you anticipate people will be able to accomplish with our new system but it serves as a point of reference. Measuring a baseline level ensures that the UX metric is measurable.
- Assume that purchasing a ticket for a particular event takes around three minutes. If that's the case, 3 minutes seems like a reasonable baseline for the first UX target in Table 5.2.9. Because most individuals have dealt with ticket offices before, this value isn't actually for the first performance, but it gives an indication of what that value might be.
- To set a baseline value for the second UX target, purchasing a movie ticket, consider that nearly no one will make any mistakes at a ticket counter, therefore set the baseline level to less than 1, as shown in Table 5.2.9. We might administer the questionnaire to some MUTTS users to obtain a baseline value for the first impression UX measure in the third UX target. Let's pretend we did that and received a 7.5 out of 10 for the first impression UX assessment.

5.2.11.2 Setting the Target Level

- Because "passing" the user experience test entails meeting all of your target levels at the same time, you must guarantee that the target levels for all UX metrics in the table are all attainable at the same time. That is, do not include compromises where achieving one target level makes achieving another related target level considerably more difficult.
- A goal level is usually established to be an improvement above the matching baseline level as a general rule of thumb. Improved user performance isn't the sole reason to construct a new system; additional functionality or just meeting user wants at a higher level in the design can also be compelling reasons. However, the emphasis here is on improving the user experience, which frequently translates to improved user performance and pleasure. You should define target levels for early performance metrics that allow new users adequate time to read menus and labels, ponder a little, and glance around each screen to get their bearings. As a result, avoid using levels for early performance indicators that assume users are familiar with the design.

Example : Target level values for the Ticket Kiosk system

- Set the target level for the first first performance UX measure in Table 5.2.10 to 2.5 minutes. In the absence of any other information, this seems like a plausible decision for our baseline level of 3 minutes. For the first UX target in the UX target table in Table 5.2.10, we enter this value in the "Target level" field.
- It might be tempting to set the target level at zero with a baseline level of fewer than one error for the "Buy movie ticket" task, but that would prevent anyone from ever making an error. As a result, we'll keep the present level of 1 as the error rate target level, as specified in second UX target of Table 5.2.10.
- Let's be conservative and set a target level of an average score of 8 out of 10 on the questionnaire for the first impression UX metric. In almost anyone's book or course, 80 percent is unquestionably passing. This belongs in third UX target of Table 5.2.10.

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	Three minutes	2.5 minutes	
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	
Ticket buyer : casual new user, for occasional personal use	Initial customer satisfaction	First impression	Questionnaire of user interface satisfaction	Average rating across users and across questions	7.5 / 10	8 / 10	

Table 5.2.10 Setting target levels for UX measures

5.2.12 Observed Results

- Final column of Table 5.2.10 is for observed results, which is a place where values measured while witnessing users executing the specified activities during formative evaluation sessions can be recorded. This column, which is part of the UX goal table, allows for direct comparisons between defined levels and testing outcomes.
- You can either record several values in a single observed results column or, if needed, add more columns for observed results and utilize this column for the average of the observed values because you will often have more than one user from which observed results are acquired.
- It is easier to manage observed data and findings later in UX evaluation analysis if you keep your UX goal tables in spreadsheets, as we recommend.

5.3 Analytic UX Evaluation : Data Collection Methods and Techniques**5.3.1 Introduction**

- Analytic UX evaluation phase of the overall UX design lifecycle template is shown in Fig. 5.3.1.

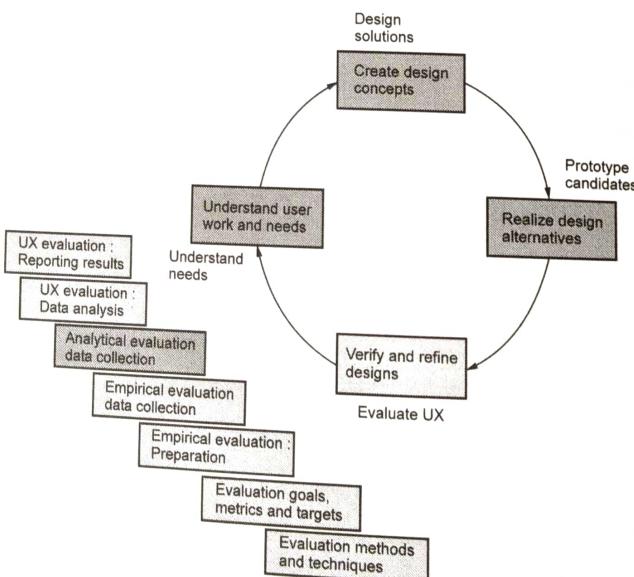


Fig. 5.3.1 Analytic UX evaluation data collection

5.3.1.1 Adding Analytic Methods to the Mix

- Some projects, particularly large domain - complex system projects, can considerably benefit from the high rigour that empirical UX evaluation can provide.
- Analytic UX evaluation approaches are an alternative for most other types of initiatives. Analytic methods were designed as speedier and less expensive approaches to provide approximations to or predictors of empirical outcomes, despite the fact that they can be executed with high rigour and hence less speed.
- As a result, analytic approaches are faster and less expensive in practise because you do not have to :
 - Identify and enlist the participation of users.
 - Participant sessions should be scheduled and participants should be brought in.
 - Long task - based evaluation sessions should be held.
- Analytic techniques rely on dismantling a UX design and studying its underlying characteristics rather than viewing the design in action, resulting in mostly qualitative, subjective data. Design reviews, design walk-throughs, and inspection methods, such as heuristic evaluation, are among the analytic approaches discussed in this chapter, which are particularly ideal for smaller fast-track projects, agile environments and product development.
- The following are some characteristics of analytic assessment methods in general :
 - They necessitate expertise in both UX and the subject-matter domain because they are dependent on expert judgments rather than empirical usage data.
 - They are nearly entirely focused on locating the most important qualitative data, as well as the most cost-effective solutions to UX issues.
 - Practical techniques are heavily relied upon.
 - They are usually less formal, with less convention and restrictions.
 - The process is significantly more variable, with practically every review "session" being unique and adapted to the current circumstances.
 - This flexibility in adapting to changing circumstances allows for more spontaneous innovation, which is something that experienced UX specialists excel at.
- Your prototype may not be well-developed enough for dealing with consumers or users in the early phases of a project. Design reviews, focus groups and walk-throughs can still be used for early design validation. You will normally use some form of the UX inspection methods if you have an interactive prototype, at least a click-through wireframe prototype, beyond these early approaches.

5.3.1.2 Criticism of Analytic Methods

- In the past, HCI literature has critiqued analytic UX evaluation methodologies, including most inspection methods. Analytic approaches have been chastised for being insufficiently thorough and unscientific and have been dubbed "discount methods" by some.
- Despite the fact that these two claims are correct, the swift approaches we explore in this chapter are the result of merely trading off the thoroughness of high rigour for low cost and quick implementation. The bread and butter of agile UX work is these strategies.

5.3.2 Design Walk - Throughs and Reviews**5.3.2.1 Design Walk - Throughs**

- Design reviews and walk - throughs are not as in-depth as the UX inspection methods, but they are still analytic because they are based on looking at design rather than empirical data from usage. A design walk - through is a non-formal method of gathering first feedback on design concepts. You normally only have scenarios, storyboards, screen sketches and / or basic wireframes at this phase. If it's too early for anybody else in a user position to engage in genuine interaction, you might have to perform the "driving."
- Walkthroughs are a great method to receive early input from the rest of the design team, as well as customers, future users, subject matter experts and other stakeholders.

5.3.2.2 Design Reviews

- Design reviews which are more advanced than early walk-throughs, are frequently conducted with click-through wireframe prototypes to demonstrate the workflow and navigation. In the late funnel's fast iteration, design reviews are frequently used as the primary means of evaluating task-level UX designs.
- The reviews are frequently a team - based UX examination in this type of use. Even though they are click-through prototypes, they are rarely interactive enough to let someone else to perform the clicking, thus you will almost always have to do the driving.
- Design reviews are less time - consuming and less expensive than participant-based testing, and their flexibility and scalability allow the effort to be scaled to meet the needs of the circumstance.

5.3.2.3 Prepare for a Design Review

- The purpose of a design walk-through or design review, like nearly any other type of analytic method is to investigate a design on behalf of users in order to imitate the user's perspective of moving through the design while seeing it through the eyes of an expert. The team is attempting to predict issues that users may encounter if they were to adopt the design.
- Do the following to prepare for a design review or walk-through :
 - Check for completeness, consistency, glitches, defects, inconsistencies, and breakdowns in your prototype.
 - Start with storyboards to illustrate flow if necessary.
 - Have a list of relevant users, work roles, and user classes on hand.
 - To guide the walk-through, practise with your design scenarios or user stories.
 - Schedule the session with the appropriate users and stakeholders :
 - Decide on a start and finish time.
 - Your UX studio is a good venue for this
 - Decide who will be the prototype's commander and who will run it.
 - Decide who will be the note-taker for the UX issues that are noticed and the modifications that are required.

5.3.2.4 Conduct a Design Review Session

- The following items can be included in an informal session protocol :
 - The leader gives an overview of the design, including its purpose and context.
 - In a guided tour of the design in action, the leader does the clicking and navigation :
 - Start with the most important workflow and navigational paths.
 - Next, go over edge situations, exceptions, and error and recovery scenarios.
 - The entire group is invited to comment and discuss.
 - The note-taker keeps track of any UX issues that are noticed.
 - The wireframe numbers should be referenced in each note.
 - Finish on time; keep to the schedule you have set.
- UX assessors examine early UX ideas through the lens of usage or design scenarios in order to be realistic and engaging. The group is led through important workflow patterns that the system is designed to enable by the leader.

- The leader presents tales about users and usage, user goals and behaviours, and expected outcomes while the team follows the scenarios, looking systematically at aspects of the design and analyzing the benefits and potential concerns. The leader describes what the user will accomplish, what the user may be thinking, and how the task fits into the work practise, workflow and environment.
- Others on the team consider how well all of this might work for users. Someone keeps track of potential UX issues and adds them to a list for future consideration.
- Compliance with design requirements and style guides, as well as concerns about emotional impact, such as aesthetics and fun, may be included in reviews. It is a good approach to communicate about the design and stay on the same page inside the project, beyond simply the nuances of UX and other design difficulties that may arise.

5.3.3 Focus Groups

- A focus group, consisting of a moderator and a half-dozen other participants, is an analytic method for (in our experience) early evaluation of UX ideas. According to Martin and Hanington (2012), "The group dynamic that focus groups produce is what gives them their power." As peers, group members share their thoughts and analyse the benefits and drawbacks. Members of the group will relate stories and utilise metaphors and analogies to convey their experiences and feelings.
- The moderator guides the group in writing a summary of the discussion at the end of each session.

5.3.4 Expert UX Inspection

5.3.4.1 What is UX Inspection ?

- A UX inspection is an analytic evaluation method in which you evaluate the design yourself as a UX expert by looking at and testing it out rather than having people do so while you observe. The evaluator serves as a surrogate for the participants as well as an observer.
- Inspectors pose inquiries to themselves about what can cause problems for consumers. As a result, the inspector's role in these approaches is to provide an opinion on potential UX issues.

5.3.4.2 Inspection is a Valuable Tool in the UX Toolbox

- For years, UX specialists have used UX inspection methodologies with tremendous success. In our own profession, we've adopted it as our primary way of evaluation,

- relying on empirical methodologies only when greater rigour is required.
- Inspection of the user experience is very beneficial :
 - When used in the early phases of design revisions.
 - When you are hired to analyse an existing system that has not been subjected to a prior UX assessment and iterative redesign.
 - When you can not afford or can not undertake scientific testing for any reason, but nevertheless want to evaluate something.
- When you do not have the time or other resources to conduct a thorough empirical review, UX inspection can still help. However, there is a tradeoff in that there will always be certain UX issues that are not visible in an inspection or design review in real-life user-based interaction.

5.3.4.3 How Many Inspectors are Needed ?

- You can improve the effectiveness of empirical UX testing by adding more participants until you reach diminishing returns. Similarly, you can increase the number of inspectors in UX inspection to boost efficacy. Because poor individual problem detection rates prevent one person from detecting enough problems, a team approach is advantageous, if not necessary.
- Different evaluators identify different problems, according to research and experience, and this diversity of skills is significant because the total number of problems detected by a group of inspectors is usually considerably larger than the total number of problems found by any single inspector. The majority of heuristic inspections are performed by a group of UX inspectors, usually two or three.
- But what is the best number to use ? It depends on the circumstances and a great deal on the system you are looking at.
- In practise, as with practically any type of evaluation, some is preferable to none. We sometimes have to make do with a single examination by one or two inspectors working jointly in the early stages of a project.

5.3.4.4 What Kind of Inspectors are Needed ?

- Nielsen discovered that the best inspection evaluators are UX professionals, which is why this type of assessment method is also known as "expert evaluation" or "expert inspection".
- It's sometimes preferable to gain a new perspective by bringing in an expert assessor who isn't part of the project team. It is even better if those UX experts are also experts in the subject matter area of the interface being assessed.

- These individuals are known as dual experts because they can evaluate from both a design standards and a job activity, workflow, and task perspective. By combining a UX expert with a work domain expert, the equivalent of having a dual expert can be approximated.

5.3.5 Heuristic Evaluation, a UX Inspection Method

- Heuristic UX evaluation is a method of analytic evaluation based on expert UX inspection in which the evaluator compares features of the design to a collection of heuristics, which are high-level UX design standards.

5.3.5.1 Introduction

- The Heuristic Evaluation (HE) method has the advantages of being inexpensive, intuitive, and easy to encourage practitioners to do, and it is effective for use early in the UX process. It is no wonder that the HE method is the most well-known and widely used of all the inspection methods.

The heuristics

- After the original heuristics were published, Nielsen improved them with a study based on factor analysis of a large number of real usability difficulties.
- The revised heuristics that resulted are as follows :

1. Visibility of system status

- The system should always keep users up to date on what is going on by providing suitable feedback in a timely manner.

2. Match between system and the real world

- Instead of using system-oriented jargon, the system should speak the users' language, using words, ideas, and concepts that are known to them. Follow real-world norms to present data in a logical and natural sequence.

3. User control and freedom

- Users frequently select system functions by accident, necessitating the presence of a clearly defined "emergency escape" that allows them to quit the unwanted state without having to go through a lengthy discussion. Undo and redo are encouraged.

4. Consistency and standards

- Users should not have to guess what different words, situations or actions signify. Stick to the platform's conventions.

5. Error prevention

- A smart design that prevents a problem from developing in the first place is even better than effective error messages. Either eliminate or check for error-prone circumstances and provide users with a confirmation option before committing to an action.

6. Recognition rather than recall

- Make objects, actions, and options visible to reduce the user's memory load. The user should not have to recall information from one dialogue segment to the next. When possible, instructions for using the system should be visible or easily accessible.

7. Flexibility and efficiency of use

- Unseen by the beginner user, accelerators can sometimes speed up the interaction for the expert user, allowing the system to cater to both inexperienced and experienced users. Allow users to customize their routine actions.

8. Aesthetic and minimalist design

- Information that is irrelevant or only occasionally required should not be included in dialogues. In a discourse, each additional unit of information competes with the relevant pieces of information, lowering their relative exposure.

9. Help users recognize, diagnose, and recover from errors

- Error messages should be written in simple English (no codes), state the problem clearly, and offer a helpful remedy.

10. Help and documentation

- Even though it is preferable if the system can be utilized without documentation, assistance and documentation may be required. Any such material should be simple to find, concentrate on the user's task, include a list of concrete steps to follow, and not be excessively big.

The procedure

- Despite the many variations in practise, we try to convey what we think of as a "simple" or "standard" form. These inspection sessions might last anywhere from a few hours to many days depending on the size of the system. Here's how to go about it :
 - A group of three to five evaluators is chosen by the project team or management.
 - The team chooses a small number of "heuristics," or generalized and simplified

- design rules in the form of inspection questions, such as "Does the interaction design employ natural language that the target user is familiar with?"
- Each inspector goes over each aspect of the interaction design independently, asking heuristic questions regarding that section and, for each heuristic question, a different set of heuristic questions.
 - Examines the compliance of each component of the design.
 - Assigns candidate usability concerns to sites where a heuristic is violated
 - Places where heuristics are supported are highlighted.
- All of the inspectors gather and work together to :
 - Combine their problem lists.
 - Choose the most critical issues to address.
 - Make a list of possible solutions.
 - Make design recommendations based on the most frequently viewed screens, screens with the most usability issues, guidelines that are most frequently broken, and resources available to make improvements.
 - Issue a group report.
- A heuristic evaluation report should include the following information :
 - Begin with a high-level overview of the system under consideration.
 - Give a high-level overview of the inspection procedure.
 - Make a list of the inspection questions based on the heuristics that were applied.
 - Report on any potential usability issues discovered during the inspection :
 - Give examples of design breaches and ways in which the design supports the heuristic for each heuristic.
 - Or, for each component of the design, offer concrete examples of heuristics that were violated and / or supported.
 - Include as many informative screen shots or other visual examples as you can.
- The team then presents the design improvements that they agreed on, using language that will encourage others to want to make these changes as well. They highlight a realistic list of the "Top 3" (or 4 or 5) proposals for changes and rank them in order to provide the most usability improvement for the least expense .

Documenting UX problems

- We have found that it best keeping HE issue documentation simple. Long forms with several fields might gather more data, but they can be onerous for UX specialists dealing with a huge number of issues.

- Include complexity and depth. Be particular and perceptive. "The system does not have good colour options because it does not use colour," is a cliche that is not particularly helpful. Furthermore, if you are evaluating a prototype, stating that some functions are not implemented is both clear and unhelpful.

Variations abound

- The HE technique, like most other analytic approaches, has one "constant": the variety with which it is applied in practise. Almost every team that adopts these methods adapts and customizes them, frequently in undocumented and unpublished ways.
- Expert UX inspections that are task - based or heuristic-based can be done with only one evaluator or two or more evaluators, each acting independently or all working together. Other types of expert UX inspections include scenario-based, persona - based, checklist-based, and "Can you break it ?" tests.
- Participatory heuristic evaluation, for example, extends the HE method with additional heuristics to address broader issues of task and workflow, beyond just the design of user interface artifacts, to "consider how the system can contribute to human goals and human experience," as described in the literature. The participation of users, or work domain experts, to the inspection team makes a significant difference in participatory HE.
- Heuristic walkthroughs were developed as an extension of the HE approach. User tasks, inspection heuristics, and "thought-focusing questions" are among the lists generated and distributed to each practitioner conducting the inspection. Each inspector conducts two inspections, one guided by the tasks and the other by the thought-focusing questions.
- The second type of inspection, which employs heuristics, is the more traditional type. According to their findings, "heuristic walkthroughs found more faults than cognitive walkthroughs and produced fewer false positives than heuristic evaluations."
- Another published version on the HE method is perspective - based usability examination. Perspective - based usability inspection allows inspectors to focus on a subset of usability concerns in each inspection session because a huge system can have a scope that is too broad for any given inspection session. Within that restricted perspective, the resultant focus of attention allowed for a higher rate of problem detection.
- Novice use, expert use, and error handling are examples of viewpoints that can be used to guide usability inspections. It is discovered that using a perspective-based

technique improved the detection of usability issues in a web - based application significantly. Persona - based UX inspection differs from perspective-based inspection in that it takes into account the context of use through the needs of personas.

- An extended problem - reporting format is created which improves heuristic inspection methods by identifying and reducing many of the false positives common in usability inspection approaches. Their Discovery and Analysis Resource (DARe) paradigm enables analysts to use different discovery and analysis resources to separate and assess both false positives and false negatives.

Limitations

- While heuristics can be useful for inexperienced practitioners, we have found that they often get in the way of professionals. To be fair to the heuristic method, it was designed as a form of "scaffolding" to assist inexperienced practitioners in conducting usability inspections, therefore it should not be compared to expert usability inspection methods in the first place.
- It is discovered that experts are the ones that detect difficulties with inspection, not experts using heuristics. The inspection procedures do not inspire analysts to take a rich or comprehensive picture of interaction. While this is true for heuristic procedures, it is not true for all inspection techniques.
- Any inspection approach, including the HE method, has the risk of new practitioners becoming too accustomed to it and believing that the heuristics are sufficient for any evaluation situation. There are few cues in its application that tell a new practitioner when it isn't working and when a different way should be tried.
- Furthermore, the HE technique, like any UX inspection methods, might result in a large number of false negatives, or scenarios in which inspectors identify "problems" that turn out to be neither real problems or minor UX issues. Finally, the HE technique, like most other analytic UX evaluation methodologies, is ineffective at uncovering usability issues beneath the surface, such as sequencing and process issues.

5.3.6 Practical Approach to UX Inspection

- We combined existing UX inspection approaches to create a reasonably basic and clear method that, unlike the heuristic method, is intended for UX specialists rather than novices. As a type of apprentice training, we occasionally have novices sit in and observe the process but they do not undertake these inspections on their own.

5.3.6.1 Guided by Insight and Experience

- This type of UX analysis is not driven by a list of criteria, either explicitly or implicitly. We have not found specific heuristics to be as effective as we would want in our own industry and consultancy expertise.
- By concentrating on tasks and work activities, we drive our inspection process with usage in context. However, we use our professional understanding of UX design guidelines to determine which issues are true difficulties and to comprehend the underlying nature of the issues and possible remedies.
- A usage - based approach appeals to us because it allows the practitioner to more fully assume the role of user. We can see and even foresee, UX problems using this technique and our UX intuition honed over the years. Many of them would not have been shown under a purely heuristic spotlight.

5.3.6.2 Use a Codiscovery or Team Approach in UX Inspection

- Inspectors who are experts in user experience play the job of "UX detectives." To help with the investigative job, two practitioners can collaborate as mutual sounding boards in a give - and - take exchange, amplifying each other's efforts to keep the juices flowing, generate a steady flow of think-aloud remarks from the inspectors, and keep a barrage of problem notes flying.
- Customers, users, designers, and other persons familiar with the overall system can often assist compensate for any lack of system knowledge on your part, especially if you have not been with the team for the entire project. Work domain specialists can supplement your user - surrogate job by providing additional subject - matter expertise.

5.3.6.3 Explore Systematically with a Rich and Comprehensive Usage - Oriented View

- As an inspector, you should search for more than just minor issues related to certain activities or functions. To see the big picture, use all of your experience and expertise. Maintain a keen eye on the high-level view of workflow, overall functionality integration, and emotional effect aspects that extend beyond usability.
- Usage and design scenarios are good places to start looking for essential user work roles and tasks that need to be addressed in the design.

5.3.6.4 Inspection is Driven by Tasks and by the Design Itself

- User tasks that are representative of the users allow us to put ourselves in their shoes. We can imagine what real users could encounter in their usage by

experiencing the activities ourselves and collecting our own think - aloud data. A hierarchical task inventory is useful for gaining a thorough grasp of the task structure and ensuring that the entire range of tasks is covered.

- Using the interaction design to drive the inspection entails performing all conceivable actions on all user interface artifacts, as well as all user interface elements such as buttons, icons, and menus. It also entails being opportunistic when it comes to following leads and hunches prompted by design elements.
- *How much time do you have ?*
 - The amount of time and effort necessary for a thorough inspection is roughly proportional to the system's size. Inspection time and effort can be impacted much more by system complexity.
- *Expertise is required*
 - The detective's "eagle eye" for strange or suspicious incidences or phenomena is the most important talent for detecting UX flaws as you inspect the design. The knowledge requirement focuses on design concepts and guidelines, as well as your mental inventory of common interface design problems you have noticed. You will need to know the design standards like the back of your hand, and your mental database of problem examples will help you anticipate and recognize fresh recurrence of the same types of issues.

5.3.6.5 Analytic UX Evaluation in the Layers of the Needs Pyramid

- There are several basic questions to ask to assist evaluate UX in the tiers of the requirements pyramid, regardless of which analytic UX evaluation method you choose :
 - Ecological layer.
 - Interaction layer.
 - Emotional layer.

5.3.6.6 Ecological - Layer Inspection

- Is there any logic to the system ecology ? Is the conceptual design suitable for the work that will be done in that domain ? If you are unsure about your capacity to check the conceptual design, focus groups and other similar approaches can be used to assess the ecological level and efficacy of the conceptual design in the early funnel.

5.3.6.7 Interaction - Layer Inspection

- The interaction layer is the most typical target of analytic UX assessment methodologies, which is usually in the context of detecting UX issues in task - level evaluation inside the late funnel stages.

5.3.6.8 Emotional - Layer Inspection

- In the past, usability inspections at the interaction level were nearly solely used to evaluate UX designs. However, by considering questions of emotional effect, this type of review may easily be expanded to a more comprehensive UX inspection. The procedure is largely the same, but you must consider the total user experience in addition to the task. Inquire about something else.
- The following are some of the emotional impact questions to consider during a UX audit :
 - Is it enjoyable to use ?
 - Is the visual design appealing and imaginative ?
 - Will the user be aesthetically, aurally, or tactilely delighted by the design ?
 - If the goal is to sell a product :
 - Is the product presentation and packaging appealing ?
 - Is it thrilling to have an out-of-the-box experience ?
 - Is the product sturdy and pleasant to hold ?
 - Is it possible for the product to boost the user's self-esteem ?
 - Is the product environmentally friendly and sustainable ?
 - Is the product consistent with the organization's branding ?
 - Is the logo associated with progressive, social, or civic values ?
 - Is there a way to boost emotional effect in any of the aforementioned areas ?
 - Focus groups and / or codiscovery can also be used to assess the emotional layer.
 - The majority of the items in an emotional impact questionnaire can likewise be used as inspection questions in this situation. Using attributes from AttrakDiff as an example :
 - Is the system or product appealing to you ?
 - Is it thrilling ?
 - Is it creative ?
 - Is it interesting ?
 - Is it energizing ?
 - Is it a good idea ?

5.4 Review Questions

1. Explain types of UX evaluation data (**Refer section 5.1.1.2**)
2. Differentiate between formative evaluation versus summative evaluation.
(Refer section 5.1.1.3)
3. Explain UX evaluation methods (**Refer section 5.1.2**)
4. Explain rigor versus rapidness in UX evaluation methods and techniques.
(Refer section 5.1.3)
5. Explain UX evaluation data collection techniques (**Refer section 5.1.4**)
6. Explain specialized UX evaluation methods (**Refer section 5.1.5**)
7. Explain UX target tables (**Refer section 5.2.2**)
8. Explain UX goals. (**Refer section 5.2.4**)
9. Explain UX measures. (**Refer section 5.2.5**)
10. What is benchmark task ? Explain guidelines for choosing benchmark tasks.
(Refer sections 5.2.6.1 and 5.2.6.2)
11. Explain UX metrics. (**Refer section 5.2.8**)
12. Explain design walk-throughs and reviews. (**Refer section 5.3.2**)
13. Explain expert UX inspection. (**Refer section 5.3.4**)
14. Explain heuristic evaluation-A UX inspection method. (**Refer section 5.3.5**)
15. Explain practical approach to UX inspection. (**Refer section 5.3.6**)

