Chapter 8

Evaluation studies: From controlled to natural settings

The aims:

- Explain how to do usability testing
- Outline the basics of experimental design
- Describe how to do field studies

Usability testing

- Involves recording performance of typical users doing typical tasks.
- Controlled settings.
- Users are observed and timed.
- Data is recorded on video & key presses are logged.
- The data is used to calculate performance times, and to identify & explain errors.
- User satisfaction is evaluated using questionnaires & interviews.
- Field observations may be used to provide contextual understanding.

Why What Where and When

- Why
  - users look for much more than just a usable system, they look for a pleasing and engaging experience
  - Merits of Doing Evaluation:
    - Designers get feedback about their early design ideas,
    - Major problems are fixed before the product goes on sale;
    - Designers focus on real problems rather than debating what each.
- What
  - developers of a new web browser may want to know whether users find items faster with their product
  - Government authorities may ask if a computerized system for controlling traffic lights results in fewer accidents.

Why What Where and When

- What
  - A new company may want to assess market reaction to its new homepage design.
  - Aesthetic, emotional, engaging, and motivating qualities
  - Time, Accuracy, Time to finish task
- Where
  - Controlled environment (Laboratory)
  - Field environment (Visiting client usage)

Why What Where and When

- When
  - If product is new, then considerable time is usually invested in market research and establishing user requirements
  - Prototypes-> if the designers have interpreted the users' requirements correctly and embodied them in their designs appropriately.
  - Upgraded-> No need for major changes
  - *summative evaluations.* Agencies such as the British Standards Institute (BSI), the National Institute of Standards and Technology (NIST) in the USA and the International Standards Organization (ISO) set standards by which products may be evaluated
  - Formative evaluations cover a broad span of design, from the development of early sketches and prototypes through to tweaking and perfecting an almost finished design, and then maintaining the product, which may involve several upgrades.

Language of Evaluation

Analytical evaluation: an approach to evaluation that does not involve end-users. Heuristic evaluation, walkthroughs, and modeling are forms of analytical evaluation.

Controlled experiment: a study that is performed in a laboratory, which is controlled by the evaluator. Aspects controlled typically include the task that participants are asked to perform, the environment in which the study occurs, and the amount of time available to complete the study.

Field study: a study that is done in a natural environment such as at home, as opposed to a study in a controlled setting such as a laboratory.

Heuristic evaluation: an approach to evaluation in which knowledge of typical users is applied, often guided by heuristics, to identify usability problems.

Predictive evaluation: an approach to evaluation in which theoretically based models are used to predict user performance.

Usability laboratory: a laboratory that is designed for usability testing.

User studies: any evaluation that invokes users directly, either in their natural environments, or in the laboratory.

Usability study: an evaluation that is performed to examine the usability of a design or system.

Usability testing: an approach to evaluation that involves measuring users' performance and testing their satisfaction with the system in question on certain tasks in a laboratory setting.

User testing: an evaluation approach where users are asked to perform certain tasks using a system or prototype in an informal or laboratory setting.

Evaluation Paradigms

- Usability Testing
- Field Testing
- Analytical Evaluation

Usability testing

- Involves measuring typical users' performance on typical tasks.
- Noting the number and kinds of errors that the users make and recording the time that it takes them to complete the task.
  - As the users perform these tasks, they are watched and recorded on video and their interactions with the software are recorded, usually by logging input to and output from the system.
  - User satisfaction questionnaires and interviews are also used to elicit users' opinions.
- Typically, tests take place in a laboratory or in laboratory-like conditions where the user is isolated from the normal day-to-day interruptions. Visitors are not allowed and telephone calls are stopped, and there is no possibility of talking to colleagues, checking email, or doing any of the other tasks that most of us rapidly switch among in our normal lives
- Findings from a usability test are summarized in a usability specification

Field studies

- What people do naturally and how products have an affect on their activities
  - (1) help identify opportunities for new technology
  - *(2)* establish the requirements for design;
  - (3) facilitate the introduction of technology, or how to deploy existing technology in new contexts;
  - (4) evaluate technology.
- Activities include interviews and observation
- The data takes the form of events and conversations that are recorded as notes, or by audio or video recording, and later analysed using a variety of methods.
- Artefacts are also collected and questionnaires may also be administered.

Analytical evaluation

- 2 types
  - Inspections, which include heuristic evaluation and walkthroughs
  - Theoretically based models, which are used to predict user performance.
- Knowledge of typical users is applied, often guided by heuristics, e.g. guidelines and standards, to identify usability problems. Walkthroughs, as the name suggests, involve experts in walking through scenarios with prototypes of the application.
- A key feature for analytical evaluations is that users need not be present
- Cognitive walkthroughs, simulating a user's problem-solving process at each step in the human-computer dialog, and checking to see how users progress from step to step in these interactions. A key feature of cognitive walkthroughs is that they focus on evaluating designs for ease of learning.

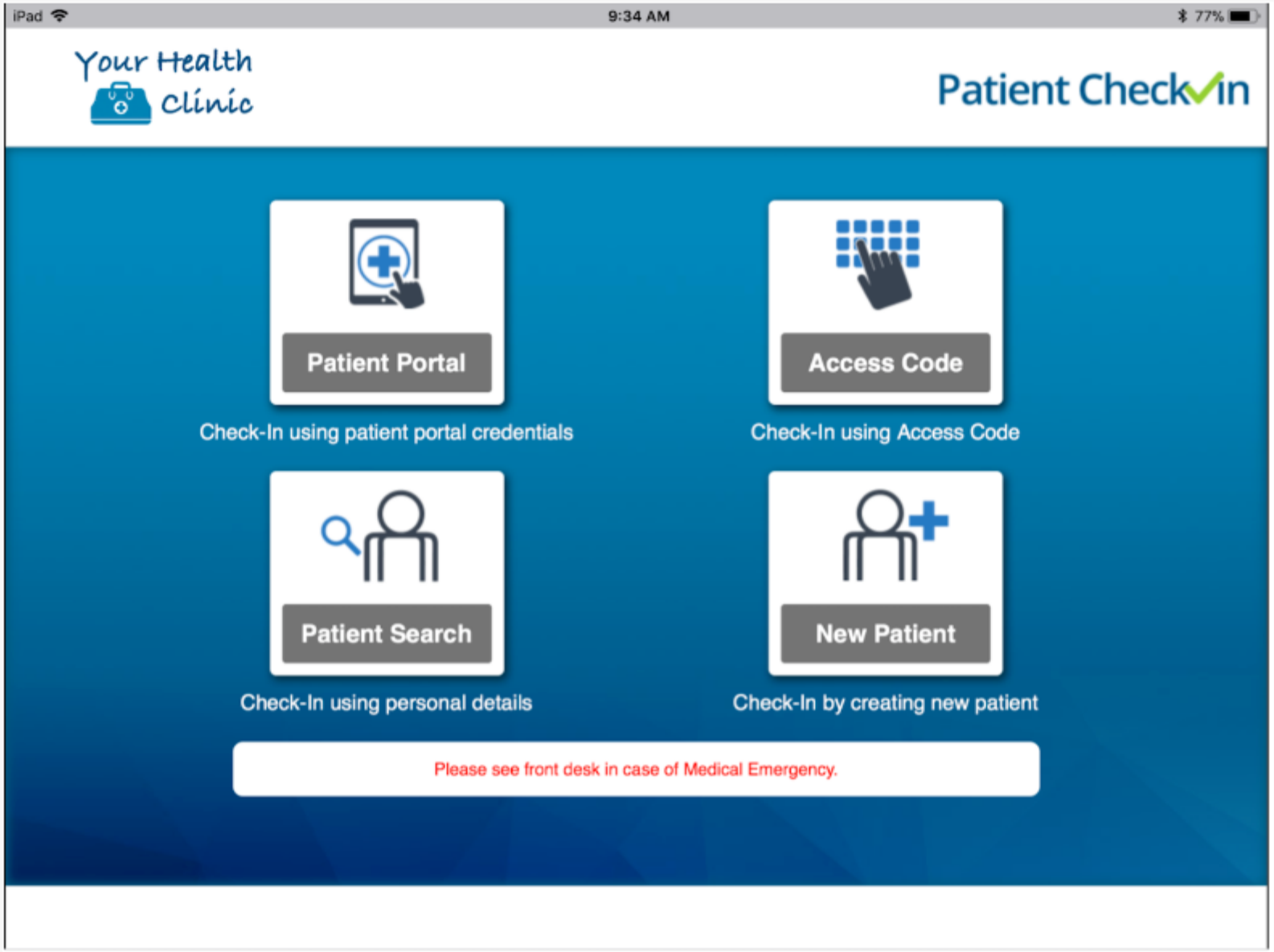
Cognitive Walkthroughs

1. **Will users try to achieve the right result?** In other words, do users understand that the action (step) at hand is needed to reach their larger goal?
2. **Will users notice that the correct action is available?** In other words, is the interactive element that achieves the step visible or easily findable?
3. **Will users associate the correct action with the result they're trying to achieve?** Perhaps the right button is visible, but will users understand the label and will they know to engage with it?
4. **After the action is performed, will users see that progress is made toward the goal?** Based on what occurs after the action is taken, will users know that this action was correct and helped them make progress toward their larger goal?

Case Study https://www.nngroup.com/articles/cognitive-walkthroughs/
The key user tasks that should be evaluated using this methodology would include:

•**Checkin:** A patient new to the clinic (of a predefined persona) arrives for an appointment and is asked by the receptionist to check in using the provided tablet application.

•**Record update:** A returning patient (of a predefined persona) arrives for an appointment and is asked by the receptionist to review and update patient information and health history using the provided tablet application.



Walkthrough evaluators would assess the first step in the patient-checkin flow for learnability. The correct action at this point is to tap the New Patient *square.*

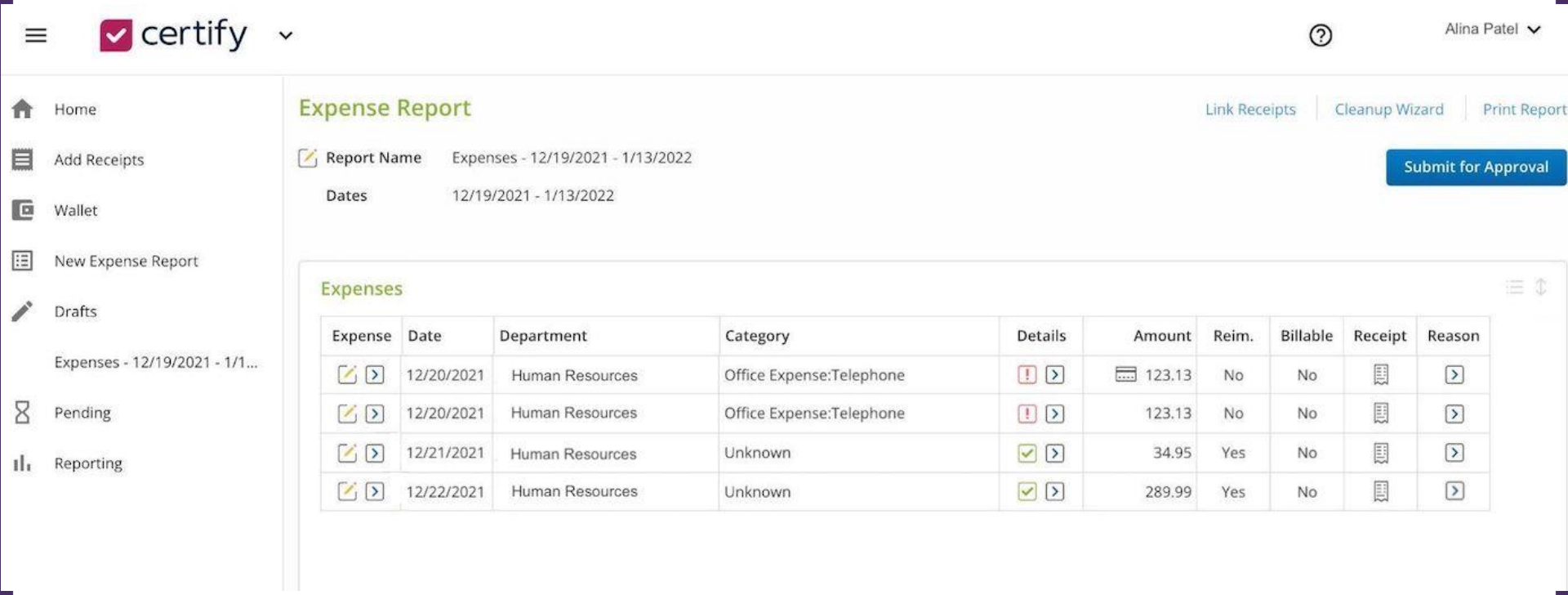
Analysis Question	Group Determination
1. Will users try to achieve the right result?	**Yes**: patients will be directed by a receptionist upon entry to check in for their appointment, and the application includes the phrase *Patient Check in* in the header. **Note:** Group discusses that there may be instances where the receptionist is away from the desk. Although the phrase *Patient Check in,* is shown in the app, its placement in the top right corner could be perceived as branding, causing it to be overlooked. They agree to further look for design solutions for this situation.
2. **Will users notice that the correct action is available?**	**Yes**: all action buttons are positioned within the body of the page using a highly salient visual styling that effectively communicates tapability.

3. **Will users associate the correct action with the result they're trying to achieve?**	**No:** the group discusses that selecting from the four options provided on the screen requires a lot of cognitive effort for new patients, because they must assess and eliminate the incorrect options before determining the correct one, *New Patient*. Some patients may assume they have a patient record because they have an appointment. Others may simply see the *Patient Search* option first and take action before assessing the *New Patient* option. The group agrees to further seek ways to simplify the design by first asking whether the patient is a new or existing patient and then providing returning visitors various record-lookup options.
4. **After the action is performed, will users see that progress is made toward the goal?**	**Yes:** the page changes and a form with the heading *Enter your personal information* is displayed.

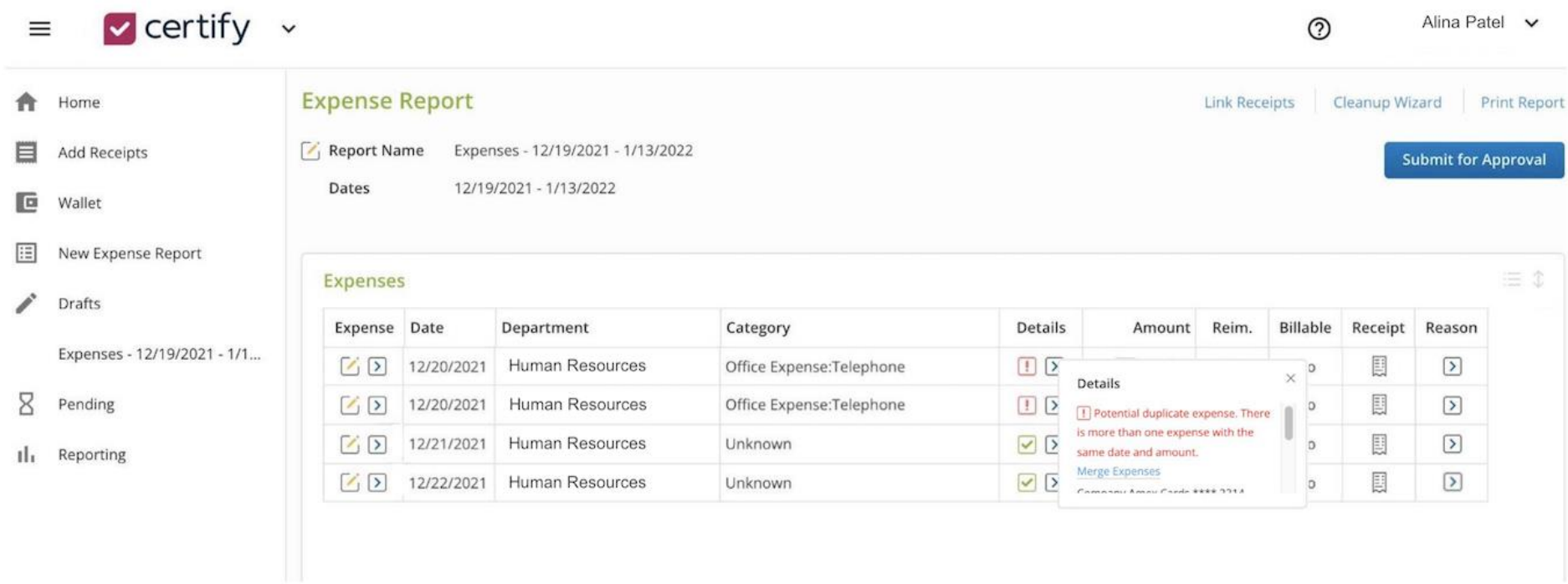
Cognitive walkthroughs Example 2

- **Persona:** Pamela, a new employee at the company.
- **Task:** Pamela needs to resolve a duplicate-expense error in her expense report.
- **Action sequence:** We'll assume Pamela already has already created her first expense report.

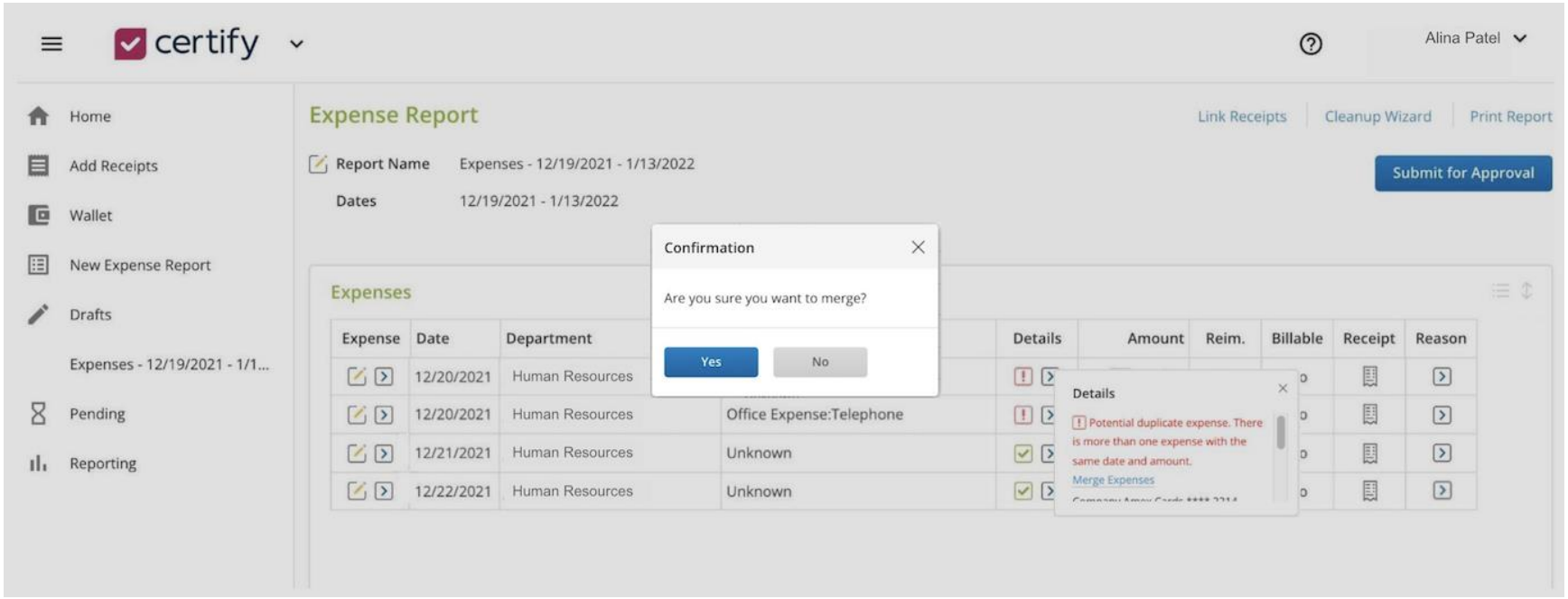
Step 1. Select the Play *icon next to the exclamation-point icon on one of the expenses in error.*



Step 2. Select the Merge *Expenses link.*



Step 3. Approve the merge in the Confirmation *modal.*



Exercise Template to fill

Task Action step				Action success	Action failure
Will the user try to achieve the right result?	yes ☐	from experience	the system tells them to	no ☐	
Will the user notice that the correct action is available?	yes ☐	from experience	they would see a call-to-action	no ☐	
Will the user associate the correct action with the effect they're trying to achieve?	yes ☐	from experience	a prompt/label matches action	no ☐	
After the action is performed, will the user see that progress is being made toward the goal?	yes ☐	from experience	there's a connection between the system response and user goal	no ☐	

Example template for recording success or failure at each step of the cognitive walkthrough.

Example of filled form

Task Pamela needs to resolve a duplicate-expense error in her expense report.

Action step Select the *Play* icon next to the error icon on one of the expenses in error.

	Action success	✓ Action failure
Will the user try to achieve the right result?	yes ☑ from experience the system tells them to The user will see the error icon next to the expenses in error. If the user does not discover and resolve these errors before attempting to submit the report, the system prompts them to do so.	no ☐
Will the user notice that the correct action is available?	yes ☑ from experience they would see a call-to-action The user will see the error icon next to the expenses in error.	no ☐
Will the user associate the correct action with the effect they're trying to achieve?	yes ☐ from experience a prompt/label matches action	no ☑ The user will likely attempt to click the error icon to resolve the error. They will not know that they must select the *Play* icon next to the error icon to fix the error.
After the action is performed, will the user see that progress is being made toward the goal?	yes ☑ from experience there's a connection between the system response and user goal The system displays a popover dialog which explains the error and offers an action to resolve the error.	no ☐

The first action in the sequence failed the third analysis question; as a result, the action was recorded as a failure in the tip right corner.

Task Pamela needs to resolve a duplicate-expense error in her expense report.

Action step Select the *Merge Expenses* link.

	✓ Action success	Action failure
Will the user try to achieve the right result?	yes ☑ from experience the system tells them to The system uses red error text to communicate that there is a potential duplicate expense. She will know she must evaluate and resolve the duplicate, so the expense is only recorded once.	no ☐
Will the user notice that the correct action is available?	yes ☑ from experience they would see a call-to-action A standard blue link reading *Merge Expenses* is offered on close proximity, directly following the error text.	no ☐
Will the user associate the correct action with the effect they're trying to achieve?	yes ☑ from experience a prompt/label matches action The label *Merge Expenses* is a straightforward resolution for resolving duplicates by combining the two into one record.	no ☐
After the action is performed, will the user see that progress is being made toward the goal?	yes ☑ from experience there's a connection between the system response and user goal One of the records is removed, and the error icon disappears from the remaining record.	no ☐

The second action in the sequence passed all analysis questions and was recorded as a success in the top right corner.

Recommendations

Action	Determination
Select the *Play* icon next to the exclamation-point icon on one of the expenses in error.	Fail
Select the *Merge Expenses* link.	Success
Approve the merge in the *Confirmation* modal.	Success

Theoretical Models

- **Keystroke-Level Model**
- Goals Operators Methods and selection models (GOMS)
- FITTS Law

GOMS

- 1983, Goals – what the user wants to achieve eg. find a website.
- Operators - the cognitive processes & physical actions needed to attain goals, eg. decide which search engine to use.
- Methods - the procedures to accomplish the goals, eg. drag mouse over field, type in keywords, press the go button.
- Selection rules - decide which method to select when there is more than one.

GOMS example

- Goal: Find a website on interaction design
- Operation: decide which search engine to use
- Methods: Type keywords in google search and press search.
- Selection rules: Either press enter or click search

GOAL COPY-AND-PASTE-TEXT

- **GOAL** COPY-TEXT
  - **GOAL** HIGHLIGH-TEXT
    - **Operator** MOVE-CURSOR-TO-BEGINNING
    - **Operator** CLICK-MOUSE-BUTTON
    - **Operator** MOVE-CURSOR-TO-END
    - **Operator** SHIFT-CLICK-MOUSE-BUTTON
    - **Operator** VERIFY-HIGHLIGHT
  - **GOAL** ISSUE-COPY-COMMAND
    - **Select***:
      - **GOAL** USE-MOUSE
        - **Operator** MOVE-CURSOR-TO-EDIT-MENU
        - **Operator** PRESS-MOUSE-BUTTON
        - **Operator** MOVE-CURSOR-TO-COPY-ITEM
        - **Operator** VERIFY-HIGHLIGHT
        - **Operator** RELEASE-MOUSE-BUTTON
      - **GOAL** USE-KEYBOARD
        - **Operator** PRESS-KEY-STRG
        - **Operator** PRESS-KEY-C
        - **Operator** RELEASE-KEYS
- **GOAL** PASTE-TEXT[...]

*Selection rule for **GOAL** ISSUE-COPY-COMMAND

if HANDS-ARE-ON-KEYBOARD **then**

select **GOAL** USE-KEYBOARD

else

select **GOAL** USE-MOUSE

Keystroke level model

- GOMS has also been developed to provide a quantitative model - the keystroke level model.
- The keystroke model allows predictions to be made about how long it takes an expert user to perform a task.

Response times for keystroke level operators (Card et al., 1983)

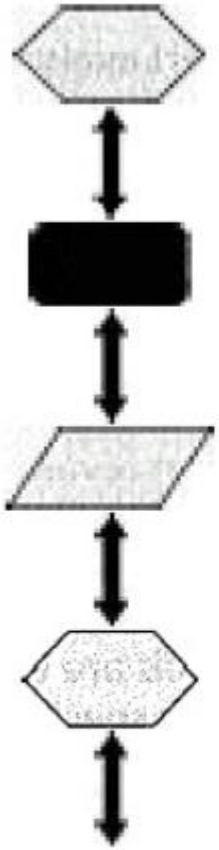
Operator	Description	Time (sec)
K	Pressing a single key or button	
	Average skilled typist (55 wpm)	0.22
	Average non-skilled typist (40 wpm)	0.28
	Pressing shift or control key	0.08
	Typist unfamiliar with the keyboard	1.20
P	Pointing with a mouse or other device on a display to select an object. This value is derived from Fitts' Law which is discussed below.	0.40
P1	Clicking the mouse or similar device	0.20
H	Bring 'home' hands on the keyboard or other device	0.40
M	Mentally prepare/respond	1.35
R(t)	The response time is counted only if it causes the user to wait.	t

$$T_{execute} = T_K + T_P + T_H + T_D + T_M + T_R$$

Fitts' Law (Fitts, 1954)

- Fitts' Law predicts that the time to point at an object using a device is a function of the distance from the target object & the object's size.
- The further away & the smaller the object, the longer the time to locate it & point to it.
- Fitts' Law is useful for evaluating systems for which the time to locate an object is important, e.g., a cell phone, a handheld devices.

Combining Paradigms



Field study to evaluate initial design ideas and get early feedback

Make some design changes

Usability test to check specific design features

field **study** to see what happens `when` used in natural environment

Make some final design changes

Example of the way usability testing and field studies can complement each other

	Evaluation Paradigms		
Characteristic	**Usability Testing**	**Field Studies**	**Analytical Evaluation**
Role of users	To carry out set tasks	Natural behaviour.	Users generally not involved
Who controls	Evaluators strongly in control	Evaluators try to develop relationships with users	Expert evaluators
Location	Laboratory	Natural environment	Laboratory-oriented but often happens on customer's premises
When used	With a prototype or product.	Most often used early in design to check that users' needs are being met or to assess problems or design opportunities	Expert reviews (often done by consultants) with a prototype, but can occur at any time. Models are used to assess specific aspects of a potential design

	Evaluation Paradigms		
Characteristic	**Usability Testing**	**Field Studies**	**Analytical Evaluation**
Type of data	Quantitative. Sometimes statistically validated. Users' opinions collected by questionnaire or interview.	Qualitative descriptions often accompanied with sketches, scenarios, quotes and other artefacts.	List of problems from expert reviews Quantitative figures from model, e.g., how long it takes to perform a task using two designs.
Fed back into design by ...	Report of performance measures, errors etc. Findings provide a benchmark for future versions.	Descriptions that include quotes, sketches, anecdotes, and sometimes time logs.	Reviewers provide a list of problems, often with suggested solutions. Times calculated from models are given to designers.
Philosophy	Applied approach based on experimentation, i.e., usability engineering.	May be objective observation or ethnographic.	Practical heuristics and practitioner expertise underpin expert reviews. Theory underpins models.

Methods for each evaluation

	Evaluation Paradigms		
Methods	**Usability Testing**	**Field Studies**	**Analytical Evaluation**
Observing users	Video and interaction logging, which can be analysed to identify errors, investigate routes through the software, or calculate performance time.	Observation is the central part of any field study. In ethnographic studies evaluators immerse themselves in the environment. In other types of studies the evaluator looks on objectively.	NIA
Asking users	User satisfaction questionnaires are administered to collect users' opinions. Interviews may also be used to get more details.	The evaluator may interview or discuss what she sees with participants. Ethnographic interviews are used in ethnographic studies.	N/A

Methods for each evaluation

	Evaluation Paradigms		
Methods	**Usability Testing**	**Field Studies**	**Analytical Evaluation**
Asking experts	N/A	NIA	Experts use heuristics early in design to predict the efficacy of an interface.
User testing	Testing typical users on typical tasks in a controlled laboratory-like setting is the cornerstone of usability testing.	N/A	N/A
Modeling users' task performance	N/A	N/A	Models are used to predict the efficacy of an interface or compare performance times between versions.

Evaluation General Guidelines

- **Basic evaluation step-by-step**

  1. Establish the aims of the evaluation, the intended users and context of use for the software; obtain or construct scenarios illustrating how the application will be used.
  2. Select evaluation methods – should be a combination of expert review and end-user testing
  3. Carry out expert review
  4. Plan user testing; use the results of the expert review to help focus this
  5. Recruit users and organise testing venue and equipment
  6. Carry out user testing

  7. Analyse results, write up and report back to designers

Experiments & usability testing

- Experiments test hypotheses to discover new knowledge by investigating the relationship between two or more things – i.e., variables.
- Usability testing is applied experimentation.
- Developers check that the system is usable by the intended user population for their tasks.
- Experiments may also be done in usability testing.

Usability testing & research

Usability testing

- Improve products
- Few participants
- Results inform design
- Usually not completely replicable
- Conditions controlled as much as possible
- Procedure planned
- Results reported to developers

Experiments for research

- Discover knowledge
- Many participants
- Results validated statistically
- Must be replicable
- Strongly controlled conditions
- Experimental design
- Scientific report to scientific community

Usability testing

- Goals & questions focus on how well users perform tasks with the product.
- Comparison of products or prototypes common.
- Focus is on time to complete task & number & type of errors.
- Data collected by video & interaction logging.
- Testing is central.
- User satisfaction questionnaires & interviews provide data about users' opinions.

Usability lab with observers watching a user & assistant

Portable equipment for use in the field

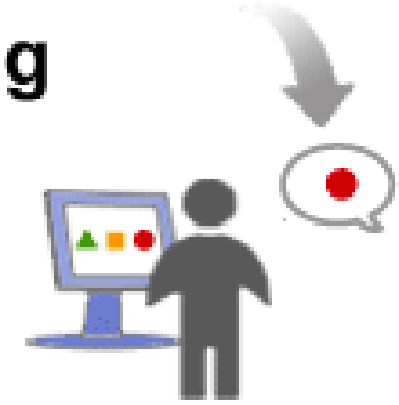
Remote Usability Testing

A selected group of panelists are invited to participate

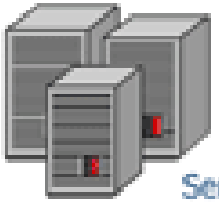
...They are asked to evaluate the web from their natural context, using Internet Explorer

...A robot (UZ Bar) guides the users and monitors their behavior

The users are asked to complete certain tasks and answer questions



The UZ Platform gatheres and saves the data in real-time



The data is analysed and a final report is prepared

Mobile head-mounted eye tracker

Picture courtesy of SensoMotoric Instruments (SMI), copyright 2010

Testing conditions

- Usability lab or other controlled space.
- Emphasis on:
  - selecting representative users;
  - developing representative tasks.
- 5-10 users typically selected.
- Tasks usually last no more than 30 minutes.
- The test conditions should be the same for every participant.
- Informed consent form explains procedures and deals with ethical issues.

Some type of data (Stop)

- Time to complete a task.
- Time to complete a task after a specified. time away from the product.
- Number and type of errors per task.
- Number of errors per unit of time.
- Number of navigations to online help or manuals.
- Number of users making a particular error.
- Number of users completing task successfully.

Usability engineering orientation

- Aim is improvement with each version.
- Current level of performance.
- Minimum acceptable level of performance.
- Target level of performance.

How many participants is enough for user testing?

- The number is a practical issue.
- Depends on:
  - schedule for testing;
  - availability of participants;
  - cost of running tests.
- Typically 5-10 participants.
- Some experts argue that testing should continue until no new insights are gained.

Name 3 features for each that can be tested by usability testing (Stop)

iPad

Experiments

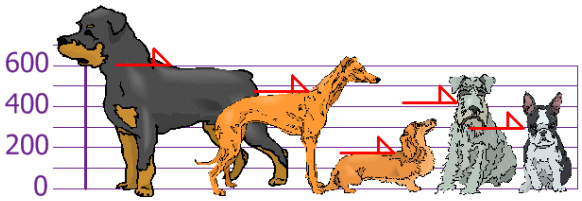
- Predict the relationship between two or more variables.
- Independent variable is manipulated by the researcher.
- Dependent variable depends on the independent variable.
- Typical experimental designs have one or two independent variable.
- Validated statistically & replicable.

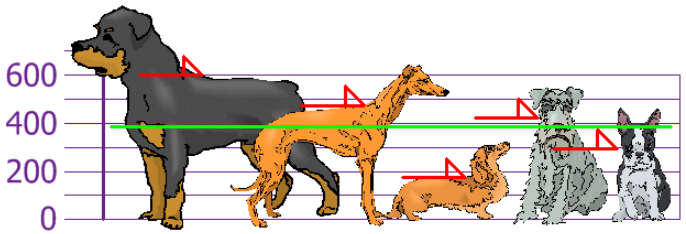
Dependent and independent variables

- Independent variable: manipulated variable. No of hours burnt by Uber Driver determine money earnt.
- Dependent Variable: measured value, no of calories burnt doing exercises

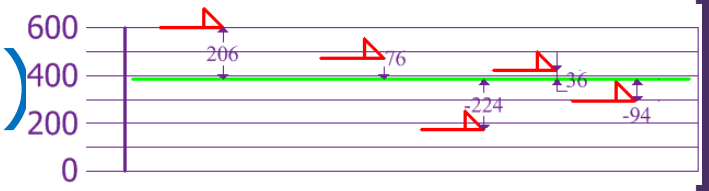
Statistical measurements

- Average

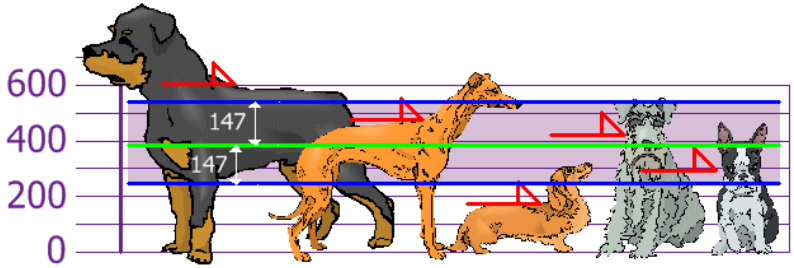




- Variance (Sigma Square)



- Standard Deviation (Square Root)
- ANOVA

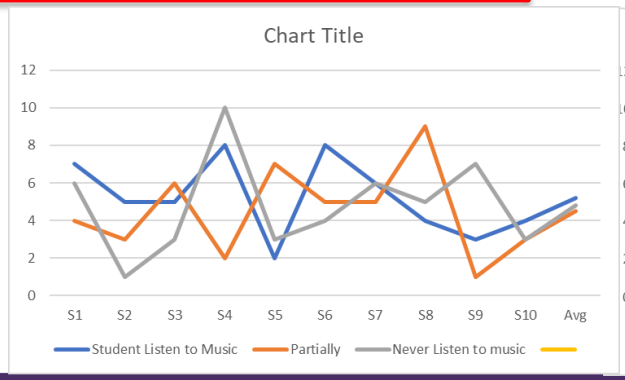


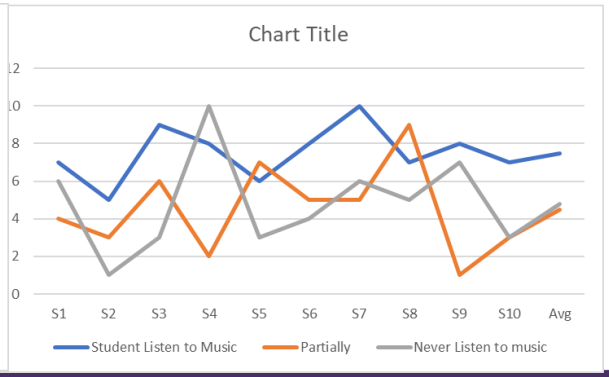
Case Study (Study and music)

- Scores of Students listen to music in exams

	Student Lister	Partially	Never Listen to music
S1	7	4	6
S2	5	3	1
S3	5	6	3
S4	8	2	10
S5	2	7	3
S6	8	5	4
S7	6	5	6
S8	4	9	5
S9	3	1	7
S10	4	3	3
Avg	5.2	4.5	4.8

	Student Li	Partially	Never Listen to music
S1	7	4	6
S2	5	3	1
S3	9	6	3
S4	8	2	10
S5	6	7	3
S6	8	5	4
S7	10	5	6
S8	7	9	5
S9	8	1	7
S10	7	3	3
Avg	7.5	4.5	4.8





Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Student Li	10	52	5.2	4.177778
Partially	10	45	4.5	5.833333
Mever List	10	48	4.8	6.622222

ANOVA

ce of Varic	SS	df	MS	F	P-value	F crit
Between (	2.466667	2	1.233333	0.22 2445	0.802012	3.35 4131
Within Gr	149.7	27	5.544444			
Total	152.1667	29				

0.05

P is greater than 0.5 this means no significant differnece

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Student Li	11	82.5	7.5	1.85
Partially	11	49.5	4.5	5.25
Mever List	11	52.8	4.8	5.96

ANOVA

ce of Varic	SS	df	MS	F	P-value	F crit
Between (	60.06	2	30.03	6.898162	0.00343	3.1 1583
Within Gr	130.6	30	4.353333			
Total	190.66	32				

Significance differnce as p<0.5

https://www.goskills.com/Lean-Six-Sigma/Resources/Use-anova-in-Excel

Experimental designs

- Different participants - single group of participants is allocated randomly to the experimental conditions.
- Same participants - all participants appear in both conditions.
- Matched participants - participants are matched in pairs, e.g., based on expertise, gender, etc.

Different, same, matched participant design

Design	Advantages	Disadvantages
Different	No order effects	Many subjects & individual differences a problem
Same	Few individuals, no individual differences	Counter-balancing needed because of ordering effects
Matched	Same as different participants but individual differences reduced	Cannot be sure of perfect matching on all differences

Field studies

- Field studies are done in natural settings.
- “in the wild” is a term for prototypes being used freely in natural settings.
- Aim to understand what users do naturally and how technology impacts them.
- Field studies are used in product design to:
  - identify opportunities for new technology;
  - determine design requirements;
  - decide how best to introduce new technology;
  - evaluate technology in use.

Data collection & analysis

- Observation & interviews
  - Notes, pictures, recordings
  - Video
  - Logging
- Analyzes
  - Categorized
  - Categories can be provided by theory
    - Grounded theory
    - Activity theory

Cooperative Usability evaluation

The Cooperative Usability Evaluation is a usability testing technique developed by Monk as a means of maximising the data from a simple testing session. The technique is 'cooperative' because users are not passive subjects but work as co-evaluators. It has proved a reliable but economical technique in diverse applications.

Step	Notes
1. Using scenarios prepared earlier, write a draft list of tasks.	Tasks must be realistic, do-able with the software, and explore the system thoroughly.
2. Try out the tasks and estimate how long they will take a user to complete.	Allow 50% longer than the total task time for each user test session.
3. Prepare a task sheet for the users.	Be specific and explain the tasks so a novice user can understand.
4. Get ready for the test session.	Have the prototype ready in a suitable environment and list of prompt questions, notebook and pens ready. An audio recorder would be very useful here.
5. Tell the users that it is the system that is under test, not them; explain the procedure and introduce the tasks.	Users should work individually – you will not be able to monitor more than one user at once. Start recording if equipment is available.
6. Users start the tasks. Have them give you a running commentary on what they are doing, why they are doing it and difficulties or uncertainties they encounter.	Take notes of where users find problems, do something unexpected, and their comments. Do this even if you are recording the session. You may need to help if users are stuck or have them move to the next task.
7. Encourage users to keep talking.	Some useful prompt questions are provided below.
8. When the users have finished, interview them briefly about the usability of the prototype and the session itself. Remember to thank the users.	Some useful questions are provided below. If you have a large number of users, a simple questionnaire may be helpful.
9. Write up your notes as soon as possible and incorporate into a usability report.	

Questions during interview

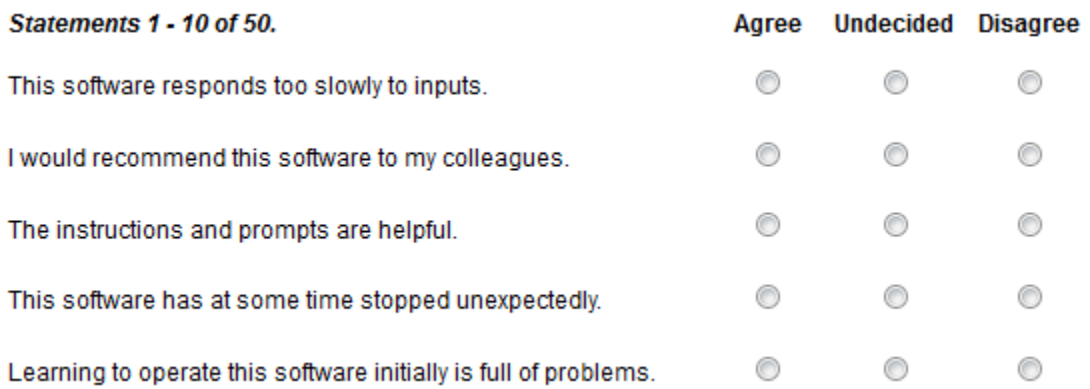
Sample questions during the test session	Sample questions after the session
What do you want to do?	The best/worst thing about the prototype.
What were you expecting to happen?	What most needs changing?
What is the system telling you?	How easy were the tasks?
Why has the system done that?	How realistic were the tasks?
What are you doing now?	Did giving a commentary distract you?

UX Research Standardise Usability Questionnaire

- Benefits
  - **Quantification**: Standardized measurements allow practitioners to report results in finer detail than they could by using only personal judgment.
  - **Scientific** generalization: Standardization is key to generalizing a finding from a sample to the greater population.
  - **Communication**: It is easier for researchers to communicate findings when referring to standardized metrics.
  - **Quick** Comparison: By using standardized questionnaires, it's easy to compare different design iterations throughout the development process.

Software Usability Measurement Inventory (SUMI)

- The SUMI is a 50-item questionnaire
- Measures users' perception of the Efficiency, Affect, Helpfulness, Control and Learnability of a system
- Using the SUMI requires purchasing a license that is approximately $700 a month.



Statements 1 - 10 of 50.	Agree	Undecided	Disagree
This software responds too slowly to inputs.	○	○	○
I would recommend this software to my colleagues.	○	○	○
The instructions and prompts are helpful.	○	○	○
This software has at some time stopped unexpectedly.	○	○	○
Learning to operate this software initially is full of problems.	○	○	○

https://www.digital-management.at/sumi-software-usability-measurement-inventory/

Post-Study Usability Questionnaire (PSSUQ)

- The PSSUQ is a 16-item survey that measures users' perceived satisfaction with a product or system.
- Obtaining an overall satisfaction score is done by averaging the four sub-scales of
  - Over all
  - System Quality (the average of items 1-6)
  - Information Quality (the average of items 7-12),
  - Interface Quality (the average of items 13-16).
- The PSSUQ is highly reliable (.94) and is entirely free.
- https://cdn.uiuxtrend.com/wp-content/uploads/PSSUQ-Questionnaire-PDF-Template.pdf

- Overa
- It was
- I was
- I felt
- It was
- I belie
- The s
- When
- The ir
  provic
- It was
- The ir
- The o
- The ir
- I liked
- This s
- Overa

	The Post-Study Usability Questionnaire Version 3	Strongly agree							Strongly disagree	
			1	2	3	4	5	6	7	NA
1	Overall, I am satisfied with how easy it is to use this system.		O	O	O	O	O	O	O	O
2	It was simple to use this system.		O	O	O	O	O	O	O	O
3	I was able to complete the tasks and scenarios quickly using this system.		O	O	O	O	O	O	O	O
4	I felt comfortable using this system.		O	O	O	O	O	O	O	O
5	It was easy to learn to use this system.		O	O	O	O	O	O	O	O
6	I believe I could become productive quickly using this system.		O	O	O	O	O	O	O	O
7	The system gave error messages that clearly told me how to fix problems.		O	O	O	O	O	O	O	O
8	Whenever I made a mistake using the system, I could recover easily and quickly.		O	O	O	O	O	O	O	O
9	The information (such as online help, on-screen messages and other documentation) provided with this system was clear.		O	O	O	O	O	O	O	O
10	It was easy to find the information I needed.		O	O	O	O	O	O	O	O
11	The information was effective in helping me complete the tasks and scenarios.		O	O	O	O	O	O	O	O
12	The organization of information on the system screens was clear.		O	O	O	O	O	O	O	O
13	The interface* of this system was pleasant.		O	O	O	O	O	O	O	O
14	I liked using the interface of this system.		O	O	O	O	O	O	O	O
15	This system has all the functions and capabilities I expect it to have.		O	O	O	O	O	O	O	O
16	Overall, I am satisfied with this system.		O	O	O	O	O	O	O	O

The System Usability Scale (SUS)

- The SUS is perhaps the most popular standardized usability questionnaire, accounting for approximately 43% of unpublished usability studies.
- The SUS is highly reliable (.91) and is entirely free
- To score the SUS, subtract the scale position from 1 on all oddly numbered items, and subtract 5 from the scale position on all evenly numbered items, then multiply the sum of all items by 2.5 to get an overall SUS score that ranges from 0-100.

1. I think that I would like to use this system frequently. +
2. I found the system unnecessarily complex.-
3. I thought the system was easy to use. +
4. I think that I would need the support of a technical person to be able to use this system. -
5. I found the various functions in this system were well integrated.+
6. I thought there was too much inconsistency in this system.-
7. I would imagine that most people would learn to use this system very quickly.+
8. I found the system very cumbersome to use.-
9. I felt very confident using the system.+
10. I needed to learn a lot of things before I could get going with this system.-

System Usability Scale

	Strongly disagree				Strongly agree	
1. I think that I would like to use this system frequently					√	4
	1	2	3	4	5	
2. I found the system unnecessarily complex				√		1
	1	2	3	4	5	
3. I thought the system was easy to use		√				1
	1	2	3	4	5	
4. I think that I would need the support of a technical person to be able to use this system	√					4
	1	2	3	4	5	
5. I found the various functions in this system were well integrated		√				1
	1	2	3	4	5	
6. I thought there was too much inconsistency in this system			√			2
	1	2	3	4	5	
7. I would imagine that most people would learn to use this system very quickly		√				1
	1	2	3	4	5	
8. I found the system very cumbersome to use				√		1
	1	2	3	4	5	
9. I felt very confident using the system					√	4
	1	2	3	4	5	
10. I needed to learn a lot of things before I could get going with this system		√				3
	1	2	3	4	5	

Total score = 22

SUS Score = 22 *2.5 = 55

68 to pass

Which Questionnaire to use

- Budget use the SUMI
- Use the PSSUQ if measuring users' satisfaction is important to the project. (all +)
- Use the SUS if measuring the users' perceived usability is important to the project. (Mix + and -)

Data presentation (Stop)

- The aim is to show how the products are being appropriated and integrated into their surroundings.
- Typical presentation forms include: vignettes, excerpts, critical incidents, patterns, and narratives.

Writing Usability report

Quantitative Data	Qualitative Data

• Enter the data in a spreadsheet to record data or make calculations such as:
  ◦ Success rates
  ◦ Task time
  ◦ Error rates
  ◦ Satisfaction questionnaire ratings
• You may want to add participant's demographic data so that you can sort by demographics to see if any of the data differ by the demographic variables.
• Make sure you identify the task scenarios for each of the metrics.
	• Record data related to:
	  ◦ Observations about pathways participants took
	  ◦ Problems experienced
	  ◦ Comments/recommendations
	  ◦ Answers to open-ended questions
	• Make sure your problem statements are exact and concise. For example:
	  ◦ Good problem statement: Clicked on link to Research instead of Clinical Trials.
	  ◦ Poor problem statement: Clicked on wrong link.
	  ◦ Poor problem statement: Was confused about links.

https://www.usability.gov/how-to-and-tools/methods/reporting-usability-test-results.html

Template for Test Report

- **Background Summary**: Include a brief summary including what you tested (website or web application), where and when the test was held, equipment information, what you did during the test (include all testing materials as an appendix), the testing team, and a brief description of the problems encountered as well as what worked well.

Methodology

- Include the test methodology so that others can recreate the test. Explain how you conducted the test by describing the test sessions, the type of interface tested, metrics collected, and an overview of task scenarios. Describe the participants and provide summary tables of the background/demographic questionnaire responses (e.g., age, professions, internet usage, site visited, etc.). Provide brief summaries of the demographic data, but do not include the full names of the participants

- **Test Results**: Include an analysis of what the facilitator and data loggers recorded. Describe the tasks that had the highest and lowest completion rates. Provide a summary of the successful task completion rates by participant, task, and average success rate by task and show the data in a table.

Results- Important to show

- Number and percent of participants who completed each scenario, and all scenarios (a bar chart often works well for this)
- Average time taken to complete each scenario for those who completed the scenario
- Satisfaction results
- Participant comments can be included if they are illustrative.

Findings and Recommendations

- List your findings and recommendations using all your data (quantitative and qualitative, notes and spreadsheets).
- Each finding should have a basis in data—in what you actually saw and heard.
- Keep in mind:
  - Although most usability test reports focus on problems, it is also useful to report positive findings.
  - What is working well must be maintained through further development.
  - An entirely negative report can be disheartening; it helps the team to know when there is a lot about the Web site that is going well.
  - Each finding should include as specific a statement of the situation as possible.
  - Each finding (or group of related findings) should include recommendations on what to do.

https://uxqb.org/public/documents/CPUX-UT_EN_Usability-Test-Report-Example.pdf

UbiFit Garden: An in the wild study

Key points

- Usability testing is done in controlled conditions.
- Usability testing is an adapted form of experimentation.
- Experiments aim to test hypotheses by manipulating certain variables while keeping others constant.
- The experimenter controls the independent variable(s) but not the dependent variable(s).
- There are three types of experimental design: different-participants, same- participants, & matched participants.
- Field studies are done in natural environments.
- "In the wild" is a recent term for studies in which a prototype is freely used in a natural setting.
- Typically observation and interviews are used to collect field studies data.
- Data is usually presented as anecdotes, excerpts, critical incidents, patterns and narratives.