User Experience Design revision notes

Introduction *week 1*

User experience design is an iterative process, centred explicitly on the user: their overall experience, their level of satisfaction, and how we can improve upon a device, system, or product's usability and accessibility.

Designers implement their conceptual model of how a system should work, while users create their mental model of how they understand the system to work through interacting with the system. The designer wants the user's mental model to match their conceptual model, otherwise the quality of the user's experience will be low.

Analysing a system, some questions to consider are: what is the system used for? Who is the expected user? What level of training/expertise is expected? What could go wrong? What steps could be taken to resolve any issues?

As time goes on, we are moving from expert systems to more widely available technologies. Technology is now necessary to participate in society, so our assumption of users, who is using the systems and the consequences if they can't use them, is changing.

Cognitive ergonomics is related to human factors, specifically the study of cognition. It aims to optimise human well-being and performance, taking cognitive limitations such as attention, memory, and workload into account.

Usability suddenly became huge as desktop computing and the need for interface designs that allow people to work well arose. Usability is based on cognitive psychology and understanding what people are capable of. Accessibility addresses the need to make systems available to everyone.

User experience puts the user at the centre. The field involves, among other things, requirements, design, prototyping, development, evaluation, cognitive abilities, subjective experience, narratives, and cultural impact. Dialogue is the key: constant, constructive dialogue between designers, users, and communities is very important.

Cognition *week 2*

Computational offloading is how external tools affect the amount of cognitive effort required to complete a task, e.g. how calculators make performing calculations easier.

There are many kinds of cognition, though user experience design mainly focuses on attention, memory, and learning. Other examples are reading, speaking and listening, problem-solving, planning, reasoning, and decision-making.

Don Norman's seven stages of action are:

1. perceiving the state of the world,
2. interpreting those perceptions,
3. evaluating those interpretations,
4. forming goals,
5. intention to act,
6. sequence of actions,
7. executing the action sequence.

Perception is out ability to make sense of the world around us and respond appropriately; it is how we sense the information around us. Sensory perception includes sight, sound, smell, taste, and touch.

Our image of the world is not given but constructed. What we can perceive is different to what we do perceive. The information that we extract depends on our motivations, our arousal, individual differences, and cultural differences. The way we perceive the world around us depends on how the brain interprets and constructs its meaning. Perception is active and constructive, not just receptive.

However, perception is not perfect; it can be fooled. Designers must help users construct the correct interpretation of the world (their system/product). We receive input and assimilate (take in and understand) it to construct our understanding, based on previous experiences.

Attention is the cognitive process of selectively concentrating on one aspect of the environment while ignoring other aspects; it is the allocation of the brain's processing resources.

Selective (a.k.a. focused) attention: focusing on one source of information while ignoring others (e.g. listening to one person speak in a crowd of others talking).

Divided attention: monitoring two or more tasks simultaneously, with attention being paid to both (e.g. talking while driving).

Sustained attention: focusing on a task over prolonged periods of time (e.g. reading a book).

We search for meaning, scanning input very quickly and looking for meaningful patterns, ignoring things that don't make sense or can't be decoded easily.

Our brains make assumptions and fill in missing details. Ambiguity causes us to 'see' different things. We also have built-in predispositions and expectations.

Designers must consider the relationship between their design and the user's attention, regarding both appropriate and inappropriate 'drawing' of attention. They must make sure that they do not bombard the user with every function at the same time, with the user's attention being drawn to appropriate functions at appropriate times. Fine grain details should only be accessible when they are needed.

Memory is how we store, manipulate, and retrieve information. Long-term memory is very large and associative; it needs time to retrieve information.

Short-term, or working, memory is where problem-solving and current processing take place. Here is where Miller's law (we can hold 7±2 things in our short-term memory), Gestalt psychology (the laws of proximity, closure, symmetry, and similarity), and chunking apply. Information is easily lost before it is transferred to long-term memory, due to disruption/interruption or distraction. Anxiety, frustration, and distraction can impede information processing, while familiarity aids processing and chunking.

Recognition is easy, while recall is difficult as recognition tasks provide memory cues that facilitate searching through memory, e.g. command lines vs GUI. When using a computer, people are already using a lot of short-term memory, so shouldn't need to think about the tool they are using.

To reduce memory load, information should be on-screen when it is needed, recognition should be used over recall, menus and paths should be shown, and screen components, menu structures, and commands should be consistent. In some cases, however, recall is more efficient than recognition. Whether it makes sense for a user to learn how to use an interface depends on the context of use (e.g. a cashier using recall on a till interface vs a customer using recognition on a self-checkout interface).

Learning is the act of acquiring new knowledge, behaviours, skills, values, preferences, or understanding. Feedback should be timely and specific – the system should be responsive – as people learn from experiences and consequences. Problems occur when feedback is not specific enough to allow us to infer cause-effect relationships, e.g. vague error messages.

It is easy to learn from structure and patterns as we need to order, categorise, and make sense of things. Learning is fastest when we can identify cause and effect, use prior knowledge to interpret, make connections, make things obvious.

Affordances are the perceived properties of an object that suggest how it can be used, e.g. pushing a button, flipping a switch, rotating a knob, labels (informative), metaphors (leveraging real-world functions), and patterns (leveraging previously learned behaviours).

When using a computer system, the user should have as little learning to do as possible. Time spent learning the system, rather than doing the task, is perceived as a waste of time. Designers should make use of existing knowledge and affordances. The system should be transparent, rather than constructive, and cause and effect should be obvious. Help should be available but well-timed. Don't obstruct the task with help messages and make it optional.

Usability and Accessibility *week 3*

Usability measures the quality of a user's experience: ease of learning, efficiency of use, memorability, error frequency and severity, and subjective satisfaction. It refers to how well users can learn and use a product and how satisfied they are with that process. Good usability allows users to complete their tasks quickly and easily, possibly considering cost-effectiveness and usefulness, too. User-centred design is key.

Usability can be defined as the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context.

Usability heuristics are rules-of-thumb to ensure products follow established usability principles. They are applied before real users use the product. The following are two sets of such heuristics.

Norman's design principles:

1. Visibility: make functional parts available and easily visible (not just by sight – by sound, touch, etc., too).
2. Affordances and constraints: use attributes people recognise.
3. Feedback: the user should be informed timely of their actions' outcomes.
4. Natural mapping: have a clear relationship between controls and their effect.
5. Good conceptual model: make sure the user's mental model is as close as can be to the designer's conceptual model.

Nielsen's 10 usability heuristics:

1. Visibility of system status: e.g. loading bars, error pop-ups.
2. Match between the system and the real world: e.g. relatable language.
3. User control and freedom: e.g. emergency exits (cancel, undo, quit).

4. Consistency and standards: words, situations, and actions always meaning the same thing as they do elsewhere.
5. Error prevention: e.g. provide clear messages about the effects of any irreversible actions.
6. Recognition rather than recall: the user should not have to remember information from one part of the system to another.
7. Flexibility and efficiency of use: provide shortcuts for experienced users.
8. Aesthetic and minimalist design: dialogues should not contain irrelevant or rarely needed information.
9. Help recognise, diagnose, and recover from errors: error messages should be expressed in plain language (no codes), indicate the problem, and suggest a solution.
10. Help and documentation: easy to search and user-focused help information.

Accessibility is the extent to which products, services, environments, etc. are accessible to as many diverse users as possible, in as many diverse contexts as possible (while usability refers to specifics). It becomes relevant when a user's environment makes performing a task difficult.

The Equality Act makes it a legal requirement to make products accessible.

Present information in multiple ways, for example, colour coded items should also be labelled and/or have a unique colour-dependent pattern.

Government guideline posters for designing for accessibility.

Designing for people from a different culture is part of accessibility as they may have different understandings of colours or symbols, for example.

Situational impairment is a (temporary) difficulty accessing a system due to the context or situation one is in, e.g. an injury, intoxication, being in a lecture, driving.

Accessible design is focused on diverse users to maximise the number of potential users who can readily use a system in diverse contexts. This can be achieved by designing systems that are ready to use by most users without modifications, making adaptable systems, and having standardised interfaces to be compatible with assistive products.

Assistive technology is any product used to increase, maintain, or improve the functional capabilities of a person with a disability or impairment.

There are tools online to analyse how well-suited to colour-blindness a webpage is and some that can recolour systems to be more accessible.

Quantitative Data *week 9*

Quantitative data looks at the magnitude, size, or amount of something, e.g. average incomes, ages, or percentage of a population. It works with large user groups as its data analysis is relatively quick but requires statistical knowledge. It allows us to gather structured feedback from people and is focused on a specific topic.

Closed questions are easy to answer, have easy-to-analyse responses, limit the number of possible responses, need to have all answers anticipated, and usually have a higher response rate, e.g. binary choice, multiple choice, ranking, semantic differential scales, combination of previous and a short answer.

Open questions can be difficult to answer, have costly-to-analyse responses, may give too many alternative answers, allows users to give any answer they want, and usually have a lower response rate. The common approach is to rely mostly on closed questions, with a few open questions as well.

Three main rules of question design: the participate should be able to comprehend, capable of answering, and willing to answer the question. Questions should be brief and relevant.

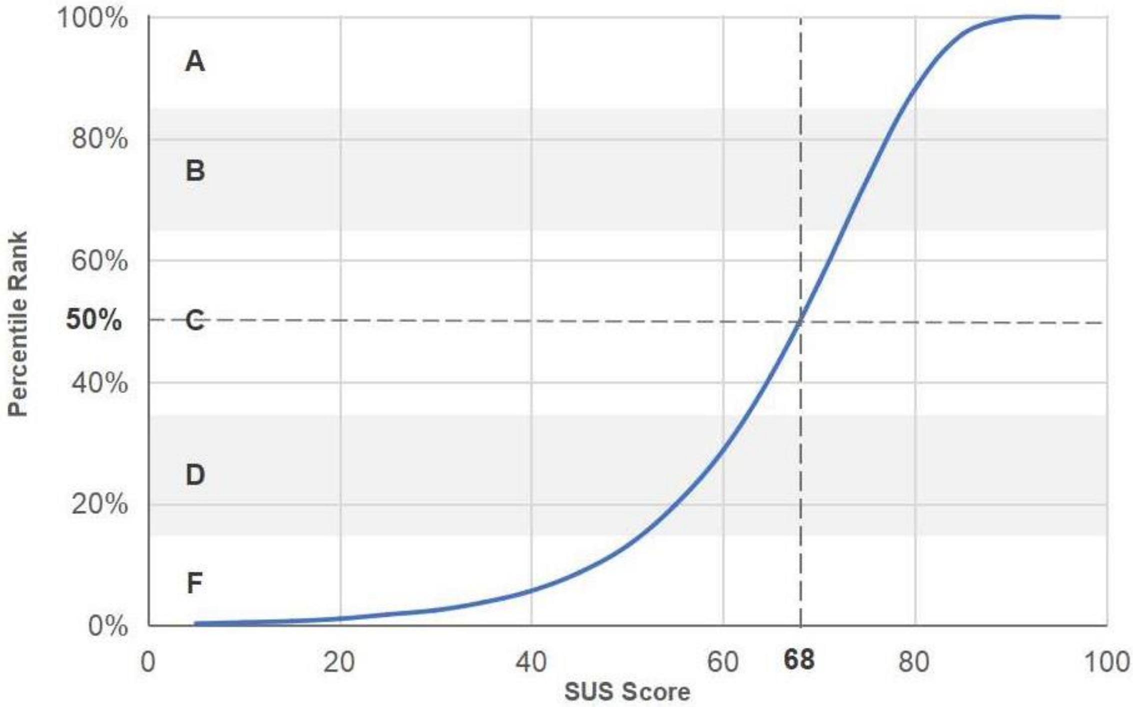
Common questionnaire errors: ranges should not overlap, don't asks double-barrelled questions, make sure scales are ordinal, give frame of reference - don't use relative terms, anticipate all possible answers (e.g. 'other'), avoid making assumptions about the participant, don't ask the participants to agree or disagree with someone.

The NASA-TLX questionnaire captures six subscales: mental demand, physical demand, temporal demand, performance, effort, and frustration.

The System Usability Scale has ten questions with a five-point Likert scale.

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.

One is taken from odd-numbered questions' answers while even-numbered questions' answers are taken from five. The sum of these new scores is multiplied by two and a half, giving the final score.



The Player Experience Inventory (PXI) captures twelve subscales: enjoyment, mastery, curiosity, meaning, immersion, autonomy, feedback, challenge, audio-visual appeal, ease of control, and clarity of goals.

Observing users is another way of measuring the quality of their experience, e.g. monitoring their facial expressions, verbal comments, how they perform a gesture on a tablet, or how they interact with a motion-based interface. Observing a person while they interact takes less time but requires you to know what you're looking for in advance. Watching a recording on a user takes more time but allows you to explore recordings first.

Automated observation can be used to improve the scale and reliability of detection, e.g. natural language processing, computer vision, eye-tracking, and accelerometers.

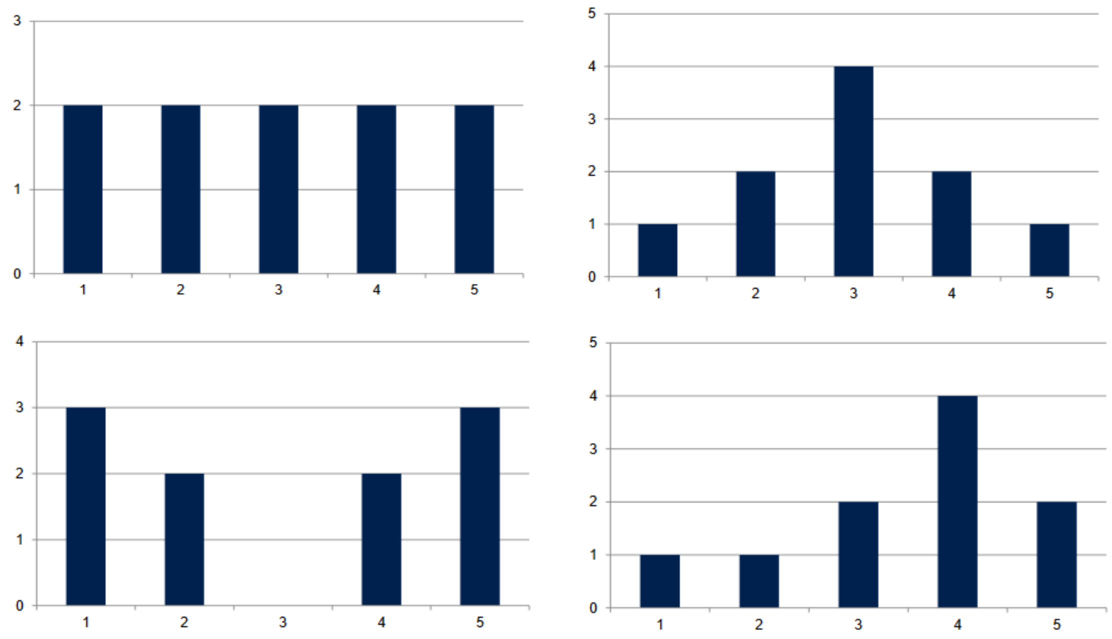
Physiological data can also be used to get very accurate and object measurements of a user's physical data, e.g. GSR (galvanic skin response), heartbeat, brain activity, and body temperature.

Performance metrics offer objective insights into how users interact with a system, e.g. time taken to complete tasks, number of errors made. They are simple to present and analyse, but do not provide a good enough insight if they are used as a sole measure. They are, however, a good way of backing up other results.

Data Analysis 1 *week 10*

System hypothesis experiment: collect data, visualise data, describe data, analyse data.

Visualising data includes looking at the distributions of answers. Distributions include uniform (top left), normal/Gaussian (top right), bimodal (bottom left), and skewed (bottom right).



Data distributions can give insights into which mathematical operations and statistical tests you can apply. Many tests require normal distribution of data, even things like the arithmetic mean, e.g. the normal distribution tells us that some people rated high, some rated low, but most rated average; the bimodal distribution shows us that half of the people rated high, while half rated low; however, the mean answer for both distributions is 3.

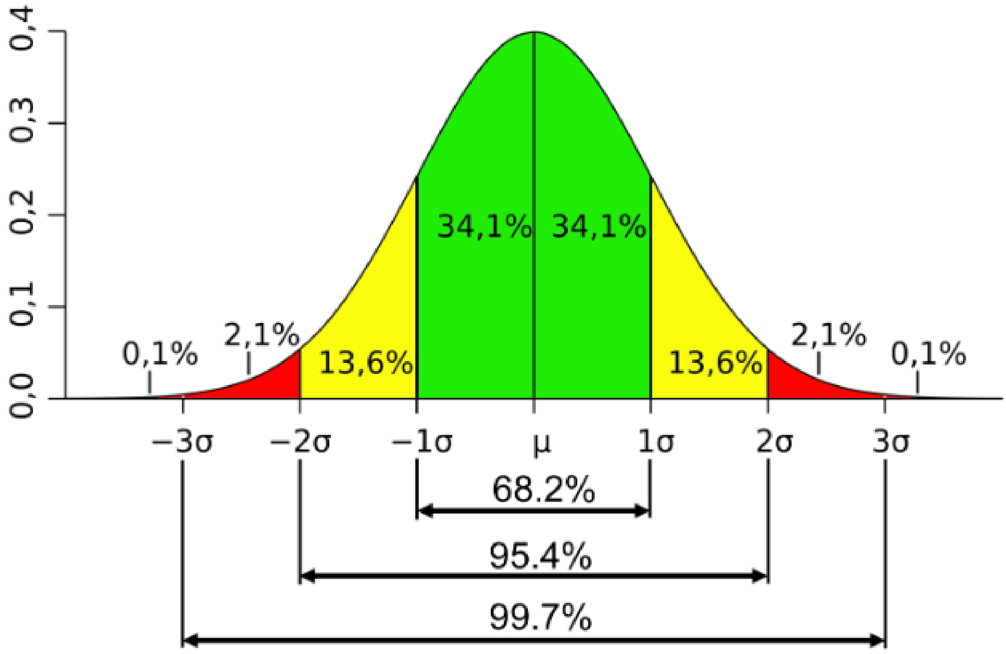
An average is a value that describes an entire distribution. The mean is the sum of all values divided by the amount of values. The mode is the most frequent value. The median is a value that splits the dataset at 50%. Depending on distribution, the chosen average leads to appropriate results. Means are used with normal distributions, medians with skewed, and modes for non-ordinal data, e.g. months.

Spread includes the range and deviation of a dataset. The range is the highest value and the lowest value of a dataset, e.g. 10 to 30 years old. The deviation tells us if the average model is a good representation of the data. Variance takes the sum of the squared (to account for direction differences) differences between the data and the average and divides it by N (whole population) or N-1 (population sample).

$$\frac{\sum(x_i - \bar{x})^2}{N-1}$$

Standard deviation takes the square root of variance to give a more realistic, smaller answer.

$$\sqrt{\frac{\sum(x_i - \overline{x})^2}{N-1}}$$



The above graph shows the percentage of answers within different amounts of the standard deviation (σ) from the mean (μ) in a normal distribution, i.e. 95.4% of answers are within two standard deviations either side of the mean.

Remember that <u>correlation</u> does not always mean <u>causation</u>. The two datasets could be affected by the same thing, e.g. ice cream sales and sunglasses sales increase during Summer, not because one causes the other, but because they are both influenced by the effects of more daily sun (hotter weather and brighter days).

Using <u>mixed methods</u> is good. <u>Quantitative</u> analysis helps explore, find patterns in, and generate high-level descriptions of data. <u>Qualitative</u> analysis helps interpret results and explains why the data is the way it is.

Data Analysis 2 *week 11*

<u>Simple random sampling</u> is selecting people at random from a known population. <u>Convenience sampling</u> is sampling people because they are convenient, e.g. physically close or part of an easily accessed group.

When comparing two sets of data, we want to know how large the effect is. What is the size of the difference between he two datasets? What impact is this likely to have?

Simple difference is the absolute value of A's mean minus B's mean.

$$diff = |\bar{A} - \bar{B}|$$

Cohen's d:

$$d = \frac{\bar{A} - \bar{B}}{s_{pooled}} \qquad s_{pooled} = \sqrt{\frac{s_A^2 + s_B^2}{2}}$$

Confidence intervals describe the level of uncertainty in a sample parameter (estimate of the population parameter), e.g. the mean of a sample. For example, if the mean of a sample is 50, you could say that you are confident that the population's mean is between 40 and 60.

A null hypothesis is the hypothesis that there is no significant difference between conditions or populations. They are used when a hypothesis cannot be proved. For example, if you were trying to prove that A was better than B, your hypothesis (H_1) "we hypothesise that A is better than B". Your null hypothesis (H_0) would be "there is no real difference between A and B". If your study found that A was better than B, you could say "we reject the null hypothesis H_0 and accept H_1".

The p-value tells us if there is a significant difference between two conditions; it is the chance that null hypothesis is true. If $p < 0.05$, that means that the hypothesis H_1 is likely true and that the null hypothesis is likely false.

A study's design is between participants if one group one task while another group completed another task, and within participants if all participants complete both tasks.

Independent variables are controlled inputs. In the previous "A is better than B" study, we only had one independent variable (A or B), while if we also wanted to see how, for example, gender impacts the results of the study, we would have two independent variables (A or B and participant gender). As we are comparing two things, A and B, we say we have two levels of independent variable.

Parametric significance tests assume the data is normally distributed. Continuous data is typically parametric. Non-parametric significance tests do not rely on any distribution.

The t-test assesses whether the means of two datasets are significantly different. It uses the two datasets' means, standard deviations, and numbers of participants. The t-test outputs the p-value.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Typically, you should aim to use a parametric test, unless the assumptions are specifically violated. First look at parametric tests, then non-parametric, as parametric tests have more statistical power.

The p-value says nothing about the magnitude of the effect. A smaller p-value does not imply a stronger effect than a larger p-value. P-values are not reliable indicators of replicability. Confidence intervals are a way of providing better information about replication.

Regarding correlations, an r value indicates the amount of variance in a set of results (a higher r value means a more significant correlation).