

UNIT 2: EVALUATION METHODS

Table of contents

1.0	Introduction
2.0	Objectives
3.0	Main Content
3.0	MAIN CONTENT
3.1	Evaluation Techniques
3.1.1	Cognitive Walkthrough
3.1.2	Heuristic Evaluation
3.1.3	Review-based evaluation
3.1.4	Evaluating through user Participation
3.2	Evaluating Implementations
3.2.1	Experimental evaluation
3.2.2	Analysis of data
3.2.3	Experimental studies on groups
3.2.4	The Data gathering and Analysis processes
3.3	Field studies
3.3.1	Observational Methods
3.3.2	Query Techniques
3.3.3	Physiological methods
4.0	Conclusion
5.0	Summary
6.0	Tutor Marked Assignment
7.0	Further Reading/References

1.0 INTRODUCTION

Evaluation tests usability and functionality of system and can be carried out in the laboratory, in the field and/or in collaboration with users.

The evaluation technique which covers both design and implementation should be considered at all stages in the design life cycle.

The goals of evaluation are to assess extent of system functionality, to assess effect of interface on user, and to identify specific problems from design and implementation.

2.0 OBJECTIVES

By the end of this unit, you should be able to:

- Express your understanding of the three main evaluation techniques viz a viz the cognitive walkthrough, the heuristic, and the review-based evaluations.
- Master available methods of evaluating interaction design and its implementation
- Select the best evaluation methods among the options available

3.0 MAIN CONTENT

3.1 Evaluation Techniques

The Evaluation Design Techniques include:

The Cognitive Walkthrough The Heuristic Evaluation
The Review-based evaluation User Participation

3.1.1 Cognitive Walkthrough

This technique was proposed by Polson et al It evaluates design on how well it supports user in learning task and is usually performed by expert in cognitive psychology. The design expert 'walks through' the design to identify potential problems using psychological principles with forms used to guide the analysis. For each task, the walkthrough considers what impact the interaction will have on the user, the cognitive processes required and the learning problems that may occur. The analysis focuses on goals and knowledge to establish whether the design leads the user to generate the correct goals.

3.1.2 Heuristic Evaluation

This was proposed by Nielsen and Molich. Here, usability criteria (heuristics) are identified; the designs are examined by experts to see if these are violated. Examples of heuristics include:
Testing whether the system behaviour is predictable Testing whether the system behaviour is consistent Testing whether a feedback is provided
Heuristic evaluation 'debugs' design.

3.1. 3 Review-based evaluation

This evaluation reviews results from the literature that are used to support or refute parts of the design. However, care is needed to ensure the results are transferable to new design. It is a model-based evaluation which in addition encompasses cognitive models that can be used to filter design options. An example is the GOMS prediction of user performance. The design rationale can also provide useful evaluation information

3.1.4 Evaluating through user Participation

Evaluation could be carried out in two ways;

1. through laboratory studies and
2. through the field studies

Laboratory studies

Laboratory studies are appropriate if system location is dangerous or impractical for constrained single user systems to allow controlled manipulation of use

The advantages of carrying out laboratory studies are appropriate specialist equipment available and are utilised in an uninterrupted environment.

Disadvantages could be lack of context and difficulty in observing several users cooperating.

Field Studies approach

This approach is appropriate where context is crucial for longitudinal studies

The advantages are that studies are carried out in a natural environment where context of evaluation is retained. Though observation may alter such context. Advantageously, longitudinal studies are also possible.

Obvious disadvantages are that there could be distractions and noise particularly in which the location of the field is within a public place.

3.2 Evaluating Implementations

To evaluate implementations, the evaluator uses artefacts such as simulation, prototypes and the full implementation.

3.2.1 Experimental evaluation

This is a controlled evaluation of specific aspects of interactive behaviour. Here the evaluator chooses the hypothesis to be tested with a number of experimental conditions considered different only in the value of some controlled variable.

The changes in behavioural measure are attributed to the different conditions.

Experimental evaluation factors

The following experimental factors are given consideration:

Subjects: This identifies who the representative is, and the measure of the sufficient sample for the experiment.

Variables: These are the things to modify and measure Hypothesis: This considers what you would like to show

Experimental design: This looks at how you are going to do it

Variables experimental factors

There are two variables; Independent variable (IV) and Dependent variable (DV)

The independent variable characteristics are changed to produce different conditions.

Examples of the characteristics include the interface style and number of menu items.

The dependent variable (DV) characteristics are those measured in the experiment. Examples of such characteristics include the time taken and number of errors.

Hypothesis experimental factor

This is a prediction of outcome framed in terms of IV and DV. For example, "error rate will increase as font size decreases".

For example, if the null hypothesis states that there is no difference between conditions, our aim is to disprove this

e.g. A null hypothesis may be stated that there is "no change with font size". So we disprove.

Experimental design factors

Within groups design: Here each subject performs experiment under each condition.

The advantage here is that transfer of learning is made possible. It is also less costly and less likely to suffer from user variation.

Between groups design: Each subject here performs under only one condition hence there is no transfer of learning. Also more users are required and variation can bias results.

3.2.2 Analysis of data

It is necessary that before you start to do any statistics, you have to look at the data and you have to save the original data.

The choice of statistical technique depends on type of data and the information required.

Type of data : This could either be discrete, that is, comprising finite number of values or continuous, comprising any value.

Analysis - types of test Parametric test:

This assumes a normal distribution and it is robust and powerful. Non-parametric test:

This does not assume a normal distribution. It is less powerful but more reliable Contingency table test:

This classifies data by discrete attributes. It counts number of data items in each group.

The information required is to establish whether there is a difference and how big the difference is. It also seeks to establish how accurate the estimate is.

However, parametric and non-parametric tests are used to mainly establish whether there is a difference.

3.2. 3 Experimental studies on groups

These are more difficult than single-user experiments.

These studies identify problems associated with subject groups, choice of task, the data gathering and the analysis of the data gathered.

Some of the problems identified with subject groups are:

The larger the number of subjects the more expensive the experimental design becomes. It also takes a longer time to 'settle down'. It creates an even more variation that makes it difficult to adhere to timetable.

Therefore, only three or four groups are recommended.

The tasks involved in experimental studies on groups are the needs to encourage cooperation among the groups through the use of multiple channels.

The options that may be adopted are:

Creative task such as *writing a short report on a particular experiment*

Decision games such as desert survival task games modelling a decision phenomenon Control

task such as demonstrated in a particular firm.

3.2.4 The Data gathering and Analysis processes

This can be done using several video cameras with direct logging of application data.

The problems with data gathering are synchronisation of data and the sheer volume of data required. The one solution to this is to record from each perspective.

Analysis of data

Because of the vast variation between groups, Carry out experiments within groups

Conduct a micro-analysis such as gaps in speech.

Conduct anecdotal and qualitative analysis and look at interactions between group and media. Realise that controlled experiments may 'waste' resources!

3. 3 Field studies

In field studies, experiments are dominated by group formation but are more realistic because: There is a distributed cognition with the work studied in context

The real action is a situated action having both the physical and social environment being crucial. Contrast:

Psychology — controlled experiment

Sociology and anthropology comprises open study and rich data

3. 3.1 Observational Methods These involve the following:

Think Aloud Cooperative evaluation Protocol analysis Automated analysis Post-task walkthroughs

Think Aloud method

The user observed performing task as he is asked to describe what he is doing and why, what he thinks is happening etc.

The advantages of this method are :- It is simple and requires little expertise It can provide useful insight

It can show how system is actually in use The disadvantages are:-

It is subjective and selective

The act of describing may alter task performance

Cooperative evaluation method

This is a variation on think aloud. The user collaborates in evaluation such that both the user and the evaluator can ask each

other questions throughout. Additional advantages here are that:- It is less constrained and easier to use

The user is encouraged to criticize the system

Clarification is possible between the user and collaborator

Protocol analysis method

This requires paper and pencil; it is therefore cheap and limited to writing speed.

It uses audio that is good for think aloud but difficult to match with other protocols.

It uses video that enables an accurate and realistic analysis but needs special equipment.

It is obtrusive. Other analysis tools involve computer logging that is automatic and unobtrusive in which large amounts of data may be difficult to analyze

It requires a user notebook that is coarse and subjective, providing useful insights and good for longitudinal studies.

However, mixed use of these tools is carried out in practice.

The audio or video transcription is difficult and requires skill. Some automatic support tools are similarly available.

Automated analysis

This is a workplace project involving a post task walkthrough where the user reacts on action after the event. It is used to fill in intention.

Advantages

The analyst has time to focus on relevant incidents It helps avoid excessive interruption of task

Disadvantages

There is a lack of freshness

There may be post-hoc interpretation of events

Post-task walkthroughs

Here transcript is played back to participant for comment

The advantages are that the response to transcript playback is immediate and is fresh in mind.

The evaluator has time to identify questions and hence useful to identify reasons for actions and alternatives considered

It is mostly necessary in cases where think aloud is not possible.

3. 3.2 Query Techniques

Query technique comprises Interviews and Questionnaires

Interviews

The analyst questions the user on one-to-one basis and is usually based on prepared questions.

The interviews are informal, subjective and relatively cheap to conduct.

The advantages are:

It can be varied to suit context Issues can be explored more fully

It can elicit user views and identify unanticipated problems

The disadvantages are that it is very subjective and time consuming.

Questionnaires

In this method, set of fixed questions are given to users.

The advantages are that it is quick and reaches large user group. It can be analyzed more rigorously.

The disadvantages are

It is less flexible and less probing

There is a need for a careful design on what information is required and how answers are to be analyzed. Styles of question are: general, open-ended, scalar, multi-choice, and ranked.

3. 3. 3 Physiological methods

These comprise Eye tracking and Physiological measurement

Eye tracking

With this method, the head or desk mounted equipment tracks the position of the eye. The eye movement reflects the amount of cognitive processing a display requires.

Measurements include

Fixations: Here, the eye maintains a stable position. The number and duration of measurements indicate level of difficulty with display

Saccades: In this case, there is a rapid eye movement from one point of interest to another.

Scan paths: This involves moving straight to a target with a short fixation at the target being optimal.

Physiological measurements

In physiological measurement, the emotional response is linked to physical changes which may help determine a user's reaction to an interface.

The measurements include:

Heart activity, including blood pressure, volume and pulse. Activity of sweat glands such as in Galvanic Skin Response (GSR) Electrical activity in muscle called electromyogram (EMG) Electrical activity in brain called electroencephalogram (EEG)

However, some difficulties are always experienced in interpreting these physiological responses; therefore, more research is required in this area.

4.0 Conclusion

Guides towards choosing an evaluation method comprise:

Commencement of evaluation process: Design versus Implementation
Style of evaluation: Laboratory versus Field

Nature of evaluation: Subjective versus Objective
Type of measures: Qualitative versus Quantitative

Level of information: High level versus Low level
Level of interference:

Obtrusive versus Unobtrusive

Resources available: Time, Subjects, Equipment and Expertise

5.0 Summary

Cognitive Walkthrough evaluates design on how well it supports user in learning task and is usually performed by expert in cognitive psychology.

In heuristic evaluation, usability criteria (heuristics) are identified and the designs are examined by experts to see if these are violated.

Review-based evaluation reviews results from the literature which are used to support or refute parts of design.

User participation evaluation is carried out through laboratory studies and field studies

Experimental evaluation is a controlled evaluation of specific aspects of interactive behaviour by choosing the hypothesis to be tested with a number of experimental conditions.

Analysis of data is done through parametric test, non-parametric test, and contingency table test of data. Experimental studies on groups identifies problems associated with subject groups, choice of task, the data gathering and the analysis of the data gathered.

Observational methods involve think aloud, cooperative evaluation, protocol analysis, automated analysis, and post-task walkthroughs

Query technique comprises Interviews and Questionnaires

Physiological methods of evaluation comprise Eye tracking and Physiological measurements.

6.0 Tutor Marked Assignment

1. What is the purpose of carrying out evaluation tests? Enumerate the available techniques used in carrying out the evaluation.
2. What is the objective of the "Cognitive walkthrough" and how is it carried out?
3. Provide 3 examples of Heuristics
4. Explain the two ways by which evaluation is carried out through user participation
5. Carrying out evaluation through the laboratory and field studies has some obvious advantages and disadvantages. What are they?
6. Describe the 4 experimental factors to consider while carrying out an experimental evaluation.
7. Describe 3 types of tests that can be carried out for analyzing data
8. "Experimental studies on subject groups are more difficult than single-user experiments". What are the specific problems associated with subjects groups to justify this statement?
9. Describe the Query techniques of evaluating design. What are their advantages and disadvantages?

10. Distinguish between the "Think Aloud" and Cooperative observational methods of evaluating designs?
11. Describe the physiological methods employed to evaluate interactive design.

7.0 Further Readings / References

- J. Dumas, J. Redish: A practical guide to usability testing. Ablex, 1993.
- D. Freedman, G. Weinberg: Walkthroughs, Inspections, and technical reviews. Dorset, 1990.
- ISO 9241 (Part 11: Guidance on usability, Part 13: User guidance)
- Monk, P. Wright, J. Haber, L. Davenport: Improving your Human-Computer Interface: a practical technique. Prentice Hall, 1993.
- J. Nielsen, R. Mack (ed.): Usability inspection methods. Wiley, 1994.
- D. Norman: The psychology of everyday things. Basic Books, 1988.