

Usability Engineering = User Centered Design+Testing

Matthias Rauterberg

HG, 3.53

TU/e - ID - DI

g.w.m.rauterberg@tue.nl

Website:

<http://www.idemployee.id.tue.nl/g.w.m.rauterberg/>

Contents course (1)

- Introduction
 - Short introduction to design/evaluation
- Requirements and specification
 - Background knowledge to be able to conduct an evaluation (e.g. user profiles)
- This lecture course: details on how to choose, plan and optimize design and evaluation activities

Contents course (2)

- Why evaluate?
- Cost/benefits of evaluation methods
- Assessment of evaluation methods
- Evaluation metrics
- Evaluation methods:
 - Inspection methods (experts)
 - Empirical methods (users)
- Compare, choose and combine methods

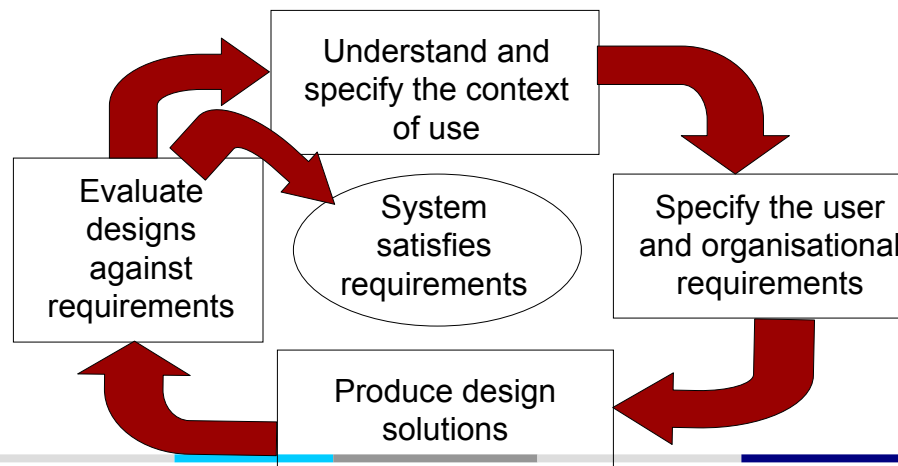
Incorrect / sub-optimal usability evaluation

- Examples:
 - Unaware of limitations of evaluation methods (validity, reliability, etc.)
 - Incorrect task selection, or task representation
 - Imprecise user profiling
 - Imprecise usability operationalisation
 - Incorrect data gathering, or data interpretation

Content today: introduction

- User-centered design
- Why evaluate?
- What is usability?
- Cost/benefits of usability evaluation
- Choosing methods/assessment
- Explanation of assignments

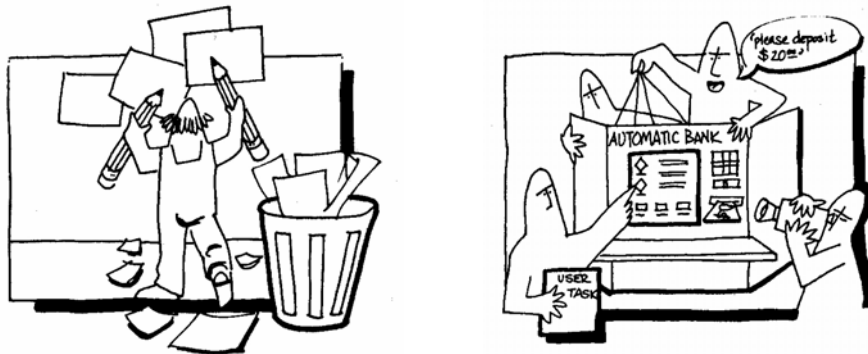
User-Centred Design



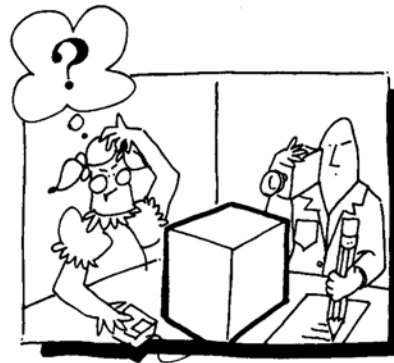
Why evaluate?

- You cannot get it right the first time
- Iteratively improve the design
 - Combine prediction of problems and observing problems

Evaluation phases (1)



Evaluation phases (2)



© M. Rauterberg, 2006

JFS-USI Primer-7

9/47

Usability: ISO DIS 9241-11

- Effectiveness, efficiency and satisfaction
- specified users with specified goals
- in particular environments
- Depending on type of product, more emphasis on fun and pleasure aspects (less task-oriented), e.g. see Jordan

© M. Rauterberg, 2006

JFS-USI Primer-7

10/47

Define usability measurements / goals

- Translate general usability definition to specific product, user groups and contexts of use
- What is effective? Reach goal
- What is efficient? Little effort
- What is satisfaction?

Maslow's hierarchy of needs



Maslow's Hierarchy of Needs is shown above. The pyramid illustrates the five levels of human needs. The most basic are physiological and safety/security, shown at the base of the pyramid. As one moves to higher levels of the pyramid, the needs become more complex. A collaborative IPT can provide self-actualization through team performance, with the ego and social needs reinforced by other team members.

Copyright © 1998 by Bellwether Learning Center

Hierarchy of consumer needs

Pleasure ↑

Usability ↑

Functionality ↑

- When people get used to something, they want more (Jordan, 2000)

Usability

- Relative importance of aspects differs between domains
- and sub-domains

- Professional applications versus games
- **Drawing application** versus drawing application for children (e.g. KidPix)

Evaluation

- In order to evaluate you have to:
 - Define what is ideal or intended behaviour, to know when user deviates
 - E.g., exploration or straight towards goal
 - Define what are usability goals of product, to know whether goal is met
 - E.g., easy to learn, efficient in the long run, or have fun

Two examples of preparation for evaluation

- User group: consumer **vs** DJ
- Ideal behaviour: simple steps to reach goals **vs** complex goals
- Usability goals:
easy to learn,
easy to use **vs**
efficient in the
long run



Why evaluate?

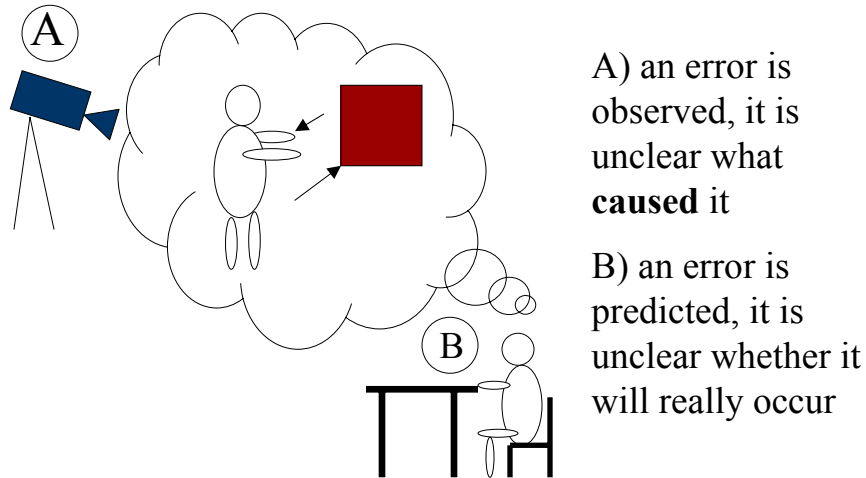
- Design is based on assumptions and predictions.
- Assumptions on different levels, e.g.:
 - About needs for a certain kind of product
 - About needs for functionality
 - About understanding of words and icons
- Amount of evidence for assumptions may vary.

- These assumptions have to be verified!

Why evaluate (2)?

- Design is based on providing a good mapping between goals of users and product
- Such a mapping may be wrong because of:
 - Incorrect assumptions
 - Unpredicted behavior
- How to determine what is deviating/ unpredicted behavior?
 - Predict errors, e.g. based on theory and models
 - Record errors that occur, e.g. using observations

Incomplete models of user system interaction



A) an error is observed, it is unclear what **caused** it

B) an error is predicted, it is unclear whether it will really occur

© M. Rauterberg, 2006

JFS-USI Primer-7

19/47

Example: cause and effect of problem

- Predicted problem:
 - The user will not know that an icon has to be double-clicked.

- Observed problem:
 - The user does not realized that a particular icon should be selected at all.

© M. Rauterberg, 2006

JFS-USI Primer-7

20/47

Norman's model of task performance

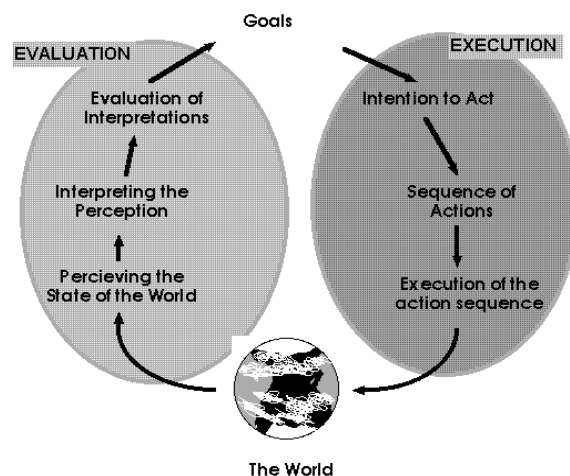
1. Forming the goal, e.g. enter room, make it brighter
2. Forming the intention, e.g. open the blinds
3. Specifying the actions, e.g. walk there, pull the cord and secure it
NOTE: none of these steps is visible - they go on inside the head...
4. Executing the action. e.g. carrying out those steps
This step is visible - it involves interacting with the world and in particular a device
5. Perceiving state of the world e.g. sight and sound of the blind rolling up, the change in brightness in the room
6. Interpreting state of the world e.g. making sense of what you perceive
7. Evaluating outcome against intention (have I rolled up the blind?) and the goal (is it lighter? if not, form another intention - maybe turn on the light)

© M. Rauterberg, 2006

JFS-USI Primer-7

21/47

Norman's action cycle



© M. Rauterberg, 2006

JFS-USI Primer-7

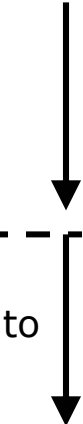
22/47

Interaction can fail at different stages of model

- The gulf of execution:
 - the difference between the intentions of the person and the perceived, allowable actions
- The gulf of evaluation:
 - the amount of effort that is exerted to interpret the physical state of the system and determine how well the expectation and intentions have been met.

How easily can the user:

1. Determine the function of the device?
 2. Tell what actions are possible?
 3. Determine mapping from intention to physical movement?
 4. Perform the action?

 5. Tell if system is in a desired state?
 6. Determine mapping from system state to interpretation?
 7. Tell what state the system is in?
- 

Choosing methods

- 1) Based on financial trade-offs:
Cost/benefit analysis
(Mantei and Teorey, 1988; Mayhew, 1999)
- 2) Based on quality of the method
(validity, reliability, etc.)
(Hartson et al., 2001)
- 3) Based on characteristics of the methods
(suitability for design phase, type of
design problem, etc.) (other lecture)

Cost/benefit analysis of evaluation methods

- Convince managers or clients of need for evaluation
- Plan a usability project (consider trade-offs)
- Compare costs for 'extra usability activities' with benefits of decreased costs

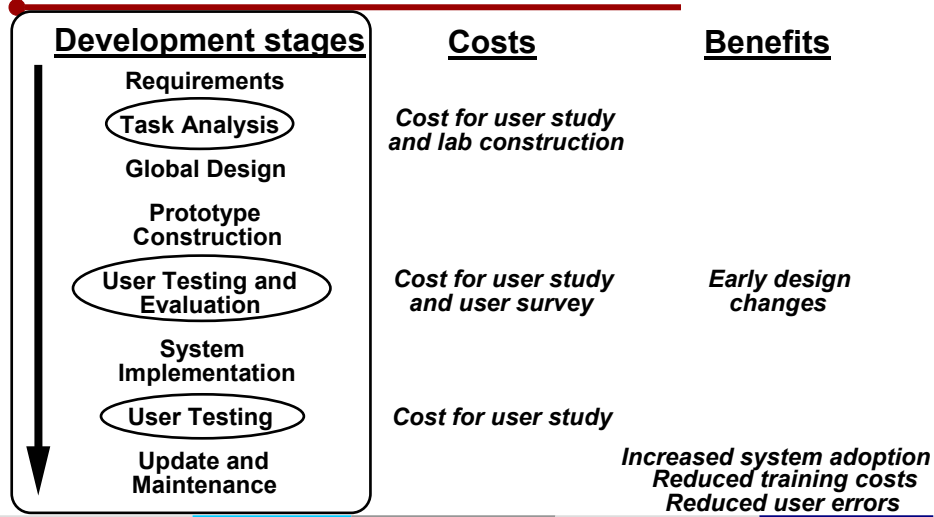
Sample benefits

- Decreased late design changes
- Decreased user training
- Increased user productivity
- Decreased user errors
- Decreased need for user support

Relation between cost reduction - design stages

- | | |
|---|----------------------------------|
| ■ Increased system adoption | ■ Market analysis |
| | ■ Product acceptance analysis |
| | ■ User testing and evaluation |
| ■ Reduced training costs | ■ Task analysis |
| | ■ User Testing |
| ■ Reduced user errors | ■ Task analysis |
| | ■ User Testing |
| ■ Transfer of design changes to an earlier stage in the project | ■ Prototype construction |
| | ■ User testing (on prototype) |
| | ■ Product survey (next redesign) |

Cost/benefit analysis of evaluation



© M. Rauterberg, 2006

JFS-USI Primer-7

29/47

Example: costs for user study with 5 subjects

Type of expense	Amount
Development of subject directions (40 hrs)	\$ 1600
Pilot testing (20 hrs)	800
Redesign of directions (20hrs)	800
Running experiment (40 hrs)	1600
Analyzing results (40hrs)	1600
Videotape	120
Cost of subjects	800
Total	\$ 7320

© M. Rauterberg, 2006

JFS-USI Primer-7

30/47

Example of cost savings

- Increased productivity for data entry system (sixty screens per day):
- Improvement: 1 second faster
- Benefit:
 - 60 transactions a day
 - Time per transaction decreased by 1 second
 - $250 \text{ users} * 230 \text{ days} * 60 \text{ transactions} * 1/3600 \text{ hours} * \$ 25 = \$ 23.958$ in first year

Comparing costs and benefits

- Investments: \$ 7.320
- Return: \$ 23.958 in first year

- Of course, real C/B analysis more items for costs and benefits.

How to assess methods?

(Hartson et al, 2001)

Are they all equally 'good'?

How do Usability Evaluation Methods (UEM) score on:

- Reliability, consistent results
- Thoroughness, complete findings
- Validity, correct findings
- Effectiveness, trade-off between complete and correct

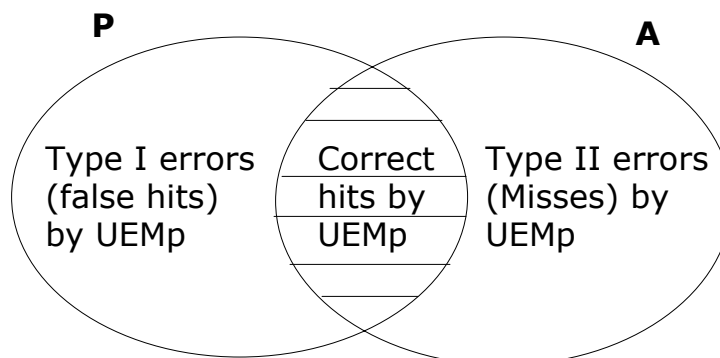
What is a real usability problem?

- If it predicts a problem that users will encounter in *real work-context usage*
 - that will have an impact on usability (performance, productivity, and/or satisfaction)
- => has to happen during real use, not just lab-use, or prediction!

How to determine the real set of usability problems?

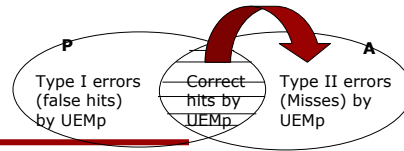
- Seeding with known usability problems
- Laboratory based usability testing
- Asymptotic lab-based testing
- Combine sets produced by two different UEM's
- Verify sets through field evaluation methods (e.g. observation, diary studies)

Relationship between problems found and existing problems



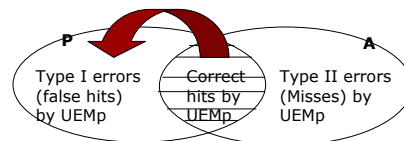
Venn diagram describing comparison of usability problem set by UEM [P] and Real usability set [A]

Thoroughness



- Proportion of real problems found to the real problems existing in the design
- E.g. 20 real problems exist, 10 real problems are found => thoroughness $10/20 = 0.5$

Validity



- Proportion of problems found that are real problems.
- 20 "problems" were found, 5 of these are real problems => validity $5/20 = 0.25$

Effectiveness

- Effectiveness = thoroughness * validity
 - E.g. 25 problems exist, UEMa finds 10 problems, of which 5 are real problems, UEMb finds 20 problems of which 10 are real
 - Effectiveness-a = $5/25 * 5/10 = 0.1$
 - Effectiveness-b = $10/25 * 10/20 = 0.2$
- Compromise target for optimisation

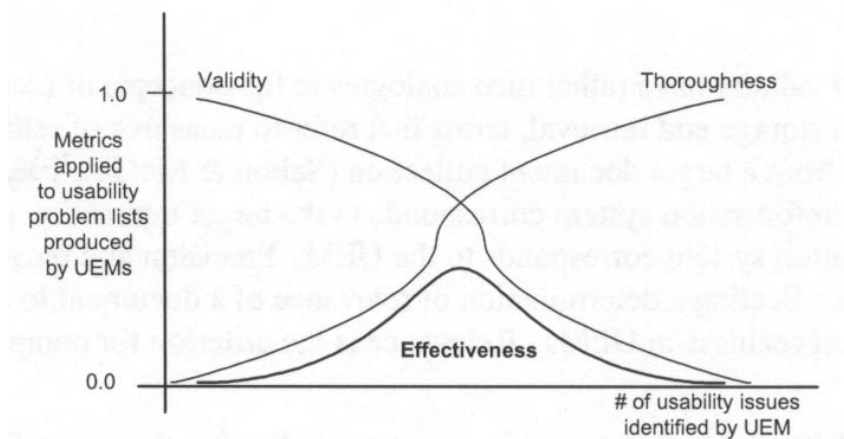
© M. Rauterberg, 2006

JFS-USI Primer-7

39/47

Speculation about the relationship among Thoroughness, Validity and Effectiveness

(Hartson et al., 2001)



© M. Rauterberg, 2006

JFS-USI Primer-7

40/47

Reliability

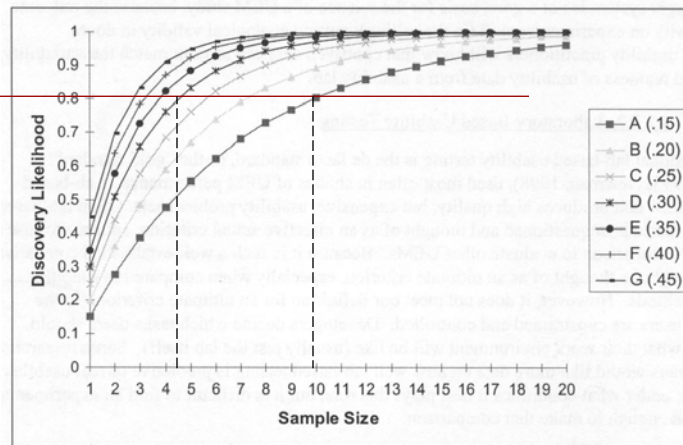
- Consistent findings
- (In)dependent of:
 - Users (Virzi, 1992; Lewis, 1994; Jacobsen, 1999)
 - Experts (Nielsen and Landauer, 1993), (Nielsen at http://www.useit.com/papers/heuristic/heuristic_evaluation.html)
 - Evaluators (Jacobsen, 1999)

The user effect in empirical evaluation

(Virzi, 1992; Lewis, 1994)

- Number of users needed to find percentage of total set of usability problems
- Formula: $1 - (1-p)^n$, with p as the probability of finding a usability problem
- Depends on probability of problem being detected
 - Virzi (1992) p between 0.32 and 0.42
 - Lewis (1994) p of 0.16

Asymptotic behavior of discovery likelihood as a function of the number of users for various detection rates (.15 -.45)



© M. Rauterberg, 2006

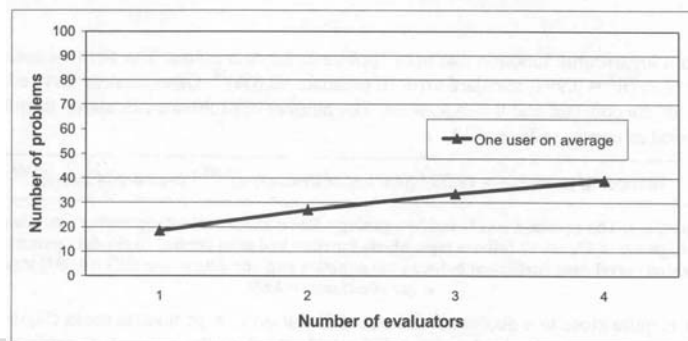
JFS-USI Primer-7

43/47

The evaluator effect in user testing

(Jacobsen et al., 1998)

- Number of problems found also depend on evaluator

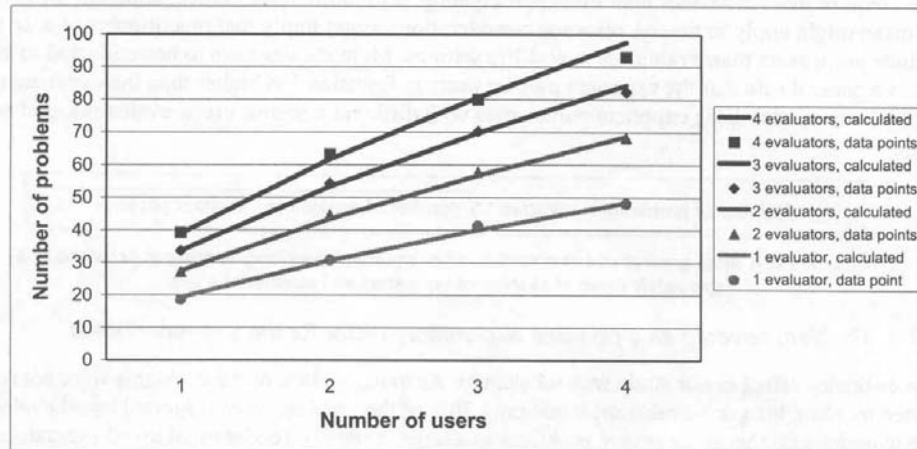


© M. Rauterberg, 2006

JFS-USI Primer-7

44/47

The combined user and evaluator effect for user testing (Jacobsen et al., 1998)



© M. Rauterberg, 2006

JFS-USI Primer-7

45/47

Optimizing evaluation

- Choose right set of methods
- Understand about pros and cons of methods
- Optimize implementation of method
 - E.g. choose the right tasks, subjects, experts, data gathering and analysis method!
 - Come to most valid conclusions!

© M. Rauterberg, 2006

JFS-USI Primer-7

46/47

References

- Hartson, H. R., Andre, T. S., & Williges, R. C. (2001). Evaluating usability evaluation methods. *International Journal of Human-Computer Interaction: Special issue on Empirical Evaluation of Information Visualisations*, 13 (4), 373-410.
- Jacobsen, N.E. , Hertzum, M. and John, B.E. (1998b) The evaluator effect in usability studies: problem detection and severity judgments. In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting: Santa Monica, CA: Human Factors and Ergonomics Society*, 1336-1340.
- Lewis, J.R. (1994) Sample sizes for usability studies: additional considerations, *Human Factors*, 36(2), 368-378.
- Mantei, M.M. and Teorey, T.J. (1988) Cost/Benefit Analysis for incorporating human factors in the software lifecycle. *Communications of the ACM*, 31(4), 428- 439.
- Mayhew, D. (1999) *The Usability Engineering Lifecycle. A practitioner's handbook for user interface design*. Chapter 20. Cost justification.449-481.
- Nielsen, J. and Landauer, T.K. (1993) A Mathematical model of the finding of usability problems, In *Proceedings of InterCHI 1993*, April 24-29, 206 -213.
- Nielsen's website: <http://www.useit.com/papers/heuristic/>
- Virzi, R.A. (1992) Refining the test phase of usability evaluation: how many subjects is enough?, *Human Factors*, 34(4), 457-468.