

projects

- paper prototyping is next week
 - run the sessions on Wednesday and Friday
 - discussion sessions, 2:00 – 3:00, IERF B011
 - sign up

overview

- final “technique” segment
 - experimental evaluation
- comparing techniques
- review

experimental evaluation

- detailed answers to focused questions
- comparative studies
- manipulate one or more factors

experimental design

- subjects
 - between-subjects designs
 - independent subject design
 - matched subject design
 - within-subjects design
 - single-subject design
 - repeated measures design
- variables
 - independent variables
 - dependent variables

reviewing exp. procedure

- user preparation
 - adequate instruction and guidance?
- impact of variables
 - what do users experience as independent variable changes?
- structure of tasks
 - complex enough to reveal the phenomena?
- time taken
 - tasks long enough? (eliminate learning effects)
 - too long? (fatigue and boredom)

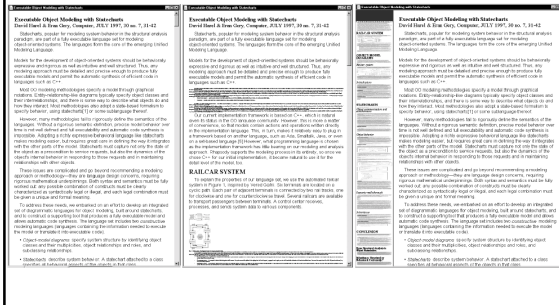
reviewing experimental results

- size of effect
 - statistical significance not always practical
- alternative interpretations
 - results explicable by other means?
- consistency between dependent variables
 - e.g. user preferences and task performance
- generalization of results
 - how general or specific?

example

- looking at on-line reading (CHI 2001)
 - we do more and more reading online
 - on the web
 - e-books
 - what is the most effective way to present material?
 - experimental evaluation
 - what are the display techniques?
 - what are the tasks?

three technologies



experimental design

- tasks
 - essay tasks and question tasks
 - reading IEEE journal papers
 - between 8 and 14 pages
 - average four figures
- dependent measures
 - effectiveness (independent blind grading)
 - satisfaction (evaluated by questionnaire)
 - efficiency (time to complete)

experimental design

- design
 - 2x3 within-subjects
 - task and interface types as independent variables
 - 20 subjects
 - three sessions, each 1h:45
- subjects
 - 15 males, 5 females
 - mean age 27
 - mean years studying computer science 6.5

results – effectiveness

Interface	Essay task (N=58)		
	Researcher's grading	Subject's grading	No. correct questions
Linear	2.00 -	2.35	4.20 +
Fisheye	1.95 -	2.32	3.42
Overview+detail	2.47 +	2.53	4.58

results – efficiency

Interface	Essay Tasks (N=58)	Question tasks (N=354)
Linear	44.4 -	5.9 +
Fisheye	37.4 +	6.6
Overview+ Detail	44.5 -	7.1 -

reviewing the experiment

- positive features
 - significant number of subjects
 - well-structured tasks
 - gave plenty of time
- negative features
 - results are pretty small
 - subjective evaluation of performance
 - essay grading

techniques

- we've covered a number
 - paper prototypes
 - interviews
 - questionnaires
 - predictive evaluation
 - heuristic evaluation
 - walkthroughs
 - contextual enquiry
 - ethnography
 - experimental evaluation

comparing techniques

- factors
 - the purpose of the evaluation
 - engineering towards a target
 - comparing alternatives?
 - understanding the world?
 - checking conformance?
 - stage of development
 - pre-requirements?
 - design?
 - pre-release?

comparing techniques

- factors
 - involvement of users
 - how many?
 - how much control?
 - type of data
 - quantitative or qualitative? both?

comparing techniques

	Observation	Experiments	User opinions	Interpretive	Predictive
Purpose					
Interface development					
User involvement					
Type of data					
Practical considerations					

technical criteria

- validity
 - does the technique measure what it should?
 - are the results what they purport to be?
- reliability
 - does it produce the same result every time?
- biases
 - are there systematic sources of error in the procedure?
 - e.g. selective data gathering

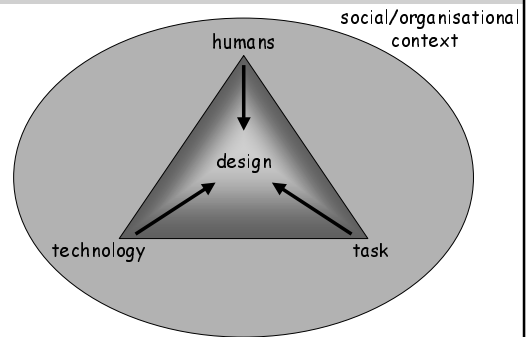
empirical comparison

- why not evaluate evaluation techniques?
- book shows studies comparing techniques
 - overall results indicate:
 - team techniques better than individual techniques
 - guidelines and heuristic evaluation highly effective
 - empirical testing uncovers major flaws others miss
 - but:
 - questions of validity
 - some of these studies were very short
 - recognition of context
 - more than simply overall effectiveness when deciding what technique to apply

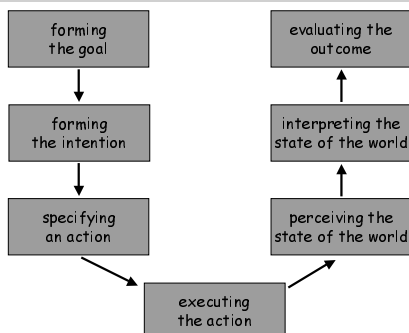
user-centered design



review – context



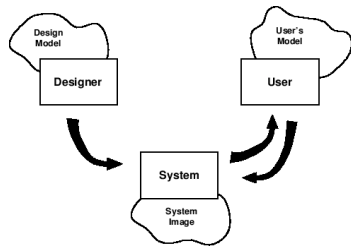
norman's seven-step model



mental models



there's more than one model



engineering for usability

- engineering for usability
 - a *systematic relationship* between design & use
 - uncovering that relationship
 - incorporating it into the process
 - affordances
 - properties of the environment (or the world, or an artifact) that afford action to appropriately equipped individuals
 - not simply physical action

next week

- Tuesday is the midterm
- Wednesday and Friday sessions
 - paper prototyping
- Thursday
 - start talking about UI programming (JFC/Swing)