

VE usability evaluations: design and issues to consider

Doug Bowman

Outline

- Two major considerations:
 - Formality of evaluation
 - What is being evaluated?
- General experiment design issues
- Formal experiment design issues
- Distinctives of VE evaluation
- Analyzing results
- Case study

Outline

- Two major considerations:
 - Formality of evaluation
 - What is being evaluated?
- General experiment design issues
- Formal experiment design issues
- Distinctives of VE evaluation
- Analyzing results
- Case study



Formality of evaluation

- Formal: independent & dependent variables, statistical analysis, strict adherence to procedure, hold constant all other variables, usually done to compare multiple techniques or at the end of the design process
- Informal: looser procedure, often more qualitative, subject comments very important, looking for broad usability issues, usually done during the design process to inform redesign

What is being evaluated?



- Application:
 - Prototype - consider fidelity, scope, form
 - Complete working system
 - Controlled experiments are rare
- Interaction techniques / UI metaphors
 - Can still evaluate a prototype
 - More generic context of use
 - Formal experiments more often used
- Consider “Wizard of Oz” evaluation

Outline

- Two major considerations:
 - Formality of evaluation
 - What is being evaluated?
- **General experiment design issues**
- Formal experiment design issues
- Distinctives of VE evaluation
- Analyzing results
- Case study

Subjects / participants

- How many?
- What backgrounds?
 - Technical vs. non-technical
 - Expert vs. novice VE users
 - Domain experts vs. general population
- What age range?
- Recruiting
 - Flyers
 - Email/listservs/newsgroups
 - Psychology dept.
 - CS classes



Number of evaluators

- Multiple evaluators often needed for VE evaluations
- Roles
 - cable wrangler
 - software controller
 - note taker
 - timer
 - behavior observer
 - ...

Procedure

- Welcome
 - Informed consent
 - Demographic/background questionnaire
 - Pre-testing
 - Familiarize with equipment
 - Exploration time with interface
 - Tasks
 - Questionnaires / post-testing
 - Interviews
- Subject “packets” are often useful for organizing information and data
 - Pilot testing should be used in most cases to:
 - “debug” your procedure
 - identify variables that can be dropped from the experiment

Instructions



- How much to tell the subject about purposes of experiment?
- How much to tell the subject about how to use the interface?
- Always tell the subject what they should try to optimize in their behavior.
- If using think-aloud protocol, you will have to remind them many times.
- If using trackers, you will have to help users “learn” to move their heads, feet, and bodies – it doesn’t come naturally to many people.
- Remind subjects you are NOT testing them, but the interface.

Outline

- Two major considerations:
 - Formality of evaluation
 - What is being evaluated?
- General experiment design issues
- **Formal experiment design issues**
- Distinctives of VE evaluation
- Analyzing results
- Case study

Formal experiment issues

- Choosing independent variables
- Choosing dependent variables
- Controlling (holding constant) other variables
- Within- vs. between-subjects design
- Counterbalancing order of conditions
- Full factorial or partial designs



Independent variables

- Main variable of interest (e.g. interaction technique)
- Secondary variables
 - Task characteristics
 - Environment characteristics
 - System characteristics
 - User characteristics

Metrics (dependent variables)

- Task performance time
- Task errors
- User comfort (subjective ratings)
- Observations of behavior (e.g. strategies)
- Spoken subject comments (e.g. preferences)
- Surveys/questionnaires
- Interviews

Formal experiment issues



- Choosing independent variables
- Choosing dependent variables
- Controlling (holding constant) other variables
- Within- vs. between-subjects design
- Counterbalancing order of conditions
- Full factorial or partial designs

Outline

- Two major considerations:
 - Formality of evaluation
 - What is being evaluated?
- General experiment design issues
- Formal experiment design issues
- **Distinctives of VE evaluation**
- Analyzing results
- Case study

Distinctives of VE evaluation (physical environment)

- Novel I/O devices
- Whole-body input, standing/walking users
- HMD users can't see physical world
- CAVE users can't see CAVE walls
- Cables
- Seeing both the subject and the VE
- Think-aloud not compatible with speech
- Video/audio of users may be difficult
- How to evaluate distributed and/or collaborative VEs?

Distinctives of VE evaluation (evaluator issues)

- Evaluator can break sense of presence
- If no evaluator intrusion, procedure must be air-tight
- Multiple evaluators needed
- Making sense of subject behavior in multi-modal interfaces

Distinctives of VE evaluation (user issues)

- What is the target population for an interaction technique?
- Very few expert users
- High variability due to novelty - many subjects needed
- How to measure presence?
- Sickness and discomfort issues

Sickness

- No exposure should last more than 20 minutes continuously
- If experiment longer than 20 minutes, plan rest breaks
- Ask subject often how they are feeling
- Allow subjects to quit anytime they want
- Measure levels of discomfort several times during long experiments
- Warn subjects not to drive immediately afterwards if they experience strong symptoms

Distinctives of VE evaluation (evaluation type issues)

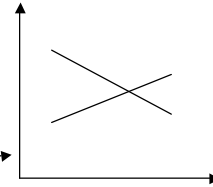
- Hard to do heuristic or guideline-based evaluation
- No predictive performance models
- Many independent variables could potentially have an effect
- Potential to over-generalize the results of controlled VE studies

Outline

- Two major considerations:
 - Formality of evaluation
 - What is being evaluated?
- General experiment design issues
- Formal experiment design issues
- Distinctives of VE evaluation
- **Analyzing results**
- Case study

Data analysis

- Averages (means) of quantitative metrics
- Counts of errors, behaviors
- Correlate data to demographics
- Analysis of variance (ANOVA)
- Visual analysis of trends (esp. learning)



- *Interactions between variables* are often important
- Expect high variance in VE interaction studies

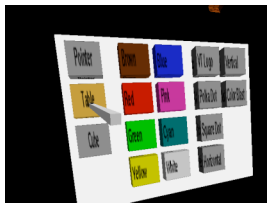
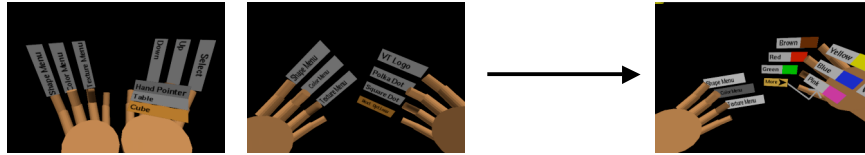
Analysis tools

- Excel
 - Basic aggregation of data
 - Correlations
 - 1- and 2-way ANOVAs
- SPSS, SAS, etc.
 - Full statistical analysis packages

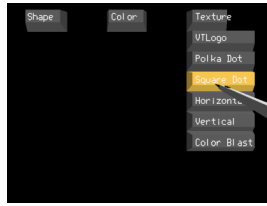
Outline

- Two major considerations:
 - Formality of evaluation
 - What is being evaluated?
- General experiment design issues
- Formal experiment design issues
- Distinctives of VE evaluation
- Analyzing results
- **Case study**

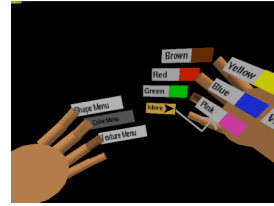
Case study: TULIP



VS.




VS.



(C) 2006 Doug Bowman, Virginia Tech

26



Resources

- 3DUI chapter 11
- VE class web page on IRB approval
 - <http://people.cs.vt.edu/~bowman/cs5754/irb.html>
- ETS standardized tests (e.g. of spatial ability)
- Usability engineering textbooks
- Statistics classes, statistics center
- Psych. or ISE classes on experimental design