FORMATIVE EVALUATION
OF USER INTERACTION:
DURING EVALUATION SESSION

Chapter 10 (c)

TOPICS:

• Generating and collecting data
• Quantitative techniques
• Qualitative techniques
• Observational data collection techniques

Copyright © 2001 H. Rex Hartson and Deborah Hix.

All rights reserved. No part of this material may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, microfilming, recording, or otherwise, without prior written permission of its authors.
DURING EVALUATION SESSION: GENERATING AND COLLECTING DATA

• Preliminaries with participants

* Explain protocol to participant, including any compensation

* Show participant the lab and experimental set-up if they are interested

* Have participant sign informed consent form and NDA
GENERATING AND COLLECTING DATA

- Some *quantitative* techniques (for numeric data, results)

* Benchmark tasks
  - Measuring time on task, number of errors, etc.
  - Quantitative measures such as timing can be valuable even with paper prototype, though not very precise
  - Facilitator usually in room with participant

* User satisfaction scores
GENERATING AND COLLECTING DATA

• Some *qualitative* techniques (for non-numeric data, results)

* Verbal protocol taking

  - *Verbal protocol*: participant thinks aloud, talking while performing tasks
    
    Can be intrusive, but effective
    
    Some participants not good at talking

  - Can be used for both timed and non-timed tasks
    
    Studies show verbal protocol has minimal effect on time of user performance

  - Facilitator may need to prod participant who stops talking, but don't get into discussion during timed tasks
    
    Too much prodding may affect time of user performance

  - Answer questions about what to do with a hint, not a direct answer

15.4 Eval During
GENERATING AND COLLECTING DATA

• Some *qualitative* techniques (continued)

* Critical incident taking

- *Critical incident:* something that happens while participant is working that has significant effect, either positive or negative, on task performance or user satisfaction

- Arguably single most important kind of formative evaluation data

- *Negative* critical incident: problem encountered

- *Positive* critical incident: occurrence that causes participant to express satisfaction or closure
GENERATING AND COLLECTING DATA

• Some qualitative techniques (continued)

* Critical incident taking (continued)

- Pay attention to detailed participant behavior
  
  It's easy to miss them! It's a skill; takes experience.

- Example: User wasn't sure what alarm clock icon meant.
  
  Could have had to do with time of day. Solution: show it "ringing" to emphasize alarm part.

- Example: User confused by word "Cancel" on button in dialogue box showing appointments
  
  Subtle: Normally understand Cancel, but with appointment in domain of calendars cancel has meaning of removing appointment.
GENERATING AND COLLECTING DATA

• Some *qualitative* techniques (continued)

* Structured interviews
  - Post-session questioning
  - Typically obtain general information

* Co-discovery
  - More than one participant, using system together, thinking aloud together
GENERATING AND COLLECTING DATA

• Some observational data collection techniques

* Video taping

  Captures every detail, but tedious to analyze

  Generally one camera on hands/keyboard/mouse/screen; if a second camera, on user's face

  Scan converter with inset camera image

  - Can be used with paper prototypes, too

    Audio-only taping is more agile.

  - Use primarily as backup

* Audio taping

* Most important: Real-time note taking (e.g., pencil and paper, on-line)

  Nothing beats this for effective data gathering.
GENERATING AND COLLECTING DATA

- Form for collecting both quantitative and qualitative data during session

<table>
<thead>
<tr>
<th>TASK NAME:</th>
<th>PARTICIPANT ID:</th>
<th>Date:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Task start time:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Task end time:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Time to perform task:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>No. of errors:</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Critical Incident Description</th>
<th>Tape Counter</th>
<th>Evaluator's Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

15.9 Eval During
GENERATING AND COLLECTING DATA

• When developers watch users evaluating a product

"Stupid user!"
  • "It's his/her (another developer's) fault!"
  • "I'm mud!"
  • "Let's fix it!"

Dilbert: user rep ->->->

• Videoclip: Formative evaluation of Envision

15.10 Eval During