EVALUATION DURING DATA COLLECTION: GENERATING AND COLLECTING DATA

• Some *quantitative* techniques

  * Benchmark tasks

  * Representative tasks
  * Need serious pretesting

  - Measuring time on task, number of errors, etc.

  - Quantitative measures such as timing can be valuable even with paper prototype, though not very precise

  * Long times still are convincing evidence of problems.

  * User satisfaction scores

  * Need serious pretesting, must be correctly done;
GENERATING AND COLLECTING DATA

• Some qualitative techniques

* 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

  - Evaluator 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

  - Evaluator often sits in room with participant to collect this kind of data
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
    - Pay attention to detailed participant behavior

  * 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.*

It's easy to miss them! It's a skill; takes experience.
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 *data collection* techniques

  * Video taping
    * Captures every detail, but tedious to analyze
    * Generally one camera on hands/keybd/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: Pencil and paper (e.g., real-time note taking)
      * 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></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time to perform task:</th>
<th>Task start time:</th>
<th>No. of errors:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></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>
</tbody>
</table>

*Task name: Schedule appt; Participant ID: P3 (No names here, to preserve anonymity)*

*Time for task: end time - start time; No. of errors: tick marks*

*Not enough for CI: user couldn’t delete appt*

*Better: didn’t select before deleting*

- Evaluation leader
  
  * In room with participant to collect qualitative data, encourage thinking aloud

  * Not in room if collecting quantitative data

- **Videoclip:** Formative evaluation of Envision
GENERATING AND COLLECTING DATA

• When developers watch users evaluating a product

"Stupid user!"

denial

⇒ "It's his/her (another developer's) fault!"

finger pointing

⇒ "I'm mud!"

acceptance

⇒ "Let's fix it!"