Session 10
Research and Evaluation results
Joshua Gutwill
Senior Researcher
Evaluation & Research Goals

Evaluation
Describe visitors’ APE behavior
Determine whether APE is happening
Compare APE & Planned Discovery exhibits

Research
Study designs that promote APE behavior
APE Studies

- Formative Evaluation
  - Describing behavior
  - Comparing iterations of same exhibit
- Summative Evaluation
  - Comparing APE to Planned Discovery
- Research
  - Studying design characteristics that lead to APE
APE Studies

• Formative Evaluation
  – Describing behavior
  – Comparing iterations of same exhibit

• Summative Evaluation
  – Comparing APE to Planned Discovery

• Research
  – Studying design characteristics that lead to APE
Comparing 1st & Last Iterations

- 8 Completed APE exhibits
- First iteration to last
  - Group holding times
Comparing 1st & Last Iterations

• 8 Completed APE exhibits
• First iteration to last
  – Group holding times

Mann-Whitney p < .01
APE Studies

• Formative Evaluation
  – Describing behavior
  – Comparing iterations of same exhibit
• Summative Evaluation
  – Comparing APE to Planned Discovery
• Research
  – Studying design characteristics that lead to APE
APE Studies

• Formative Evaluation
  – Describing behavior
  – Comparing iterations of same exhibit

• Summative Evaluation
  – Comparing APE to Planned Discovery

• Research
  – Studying design characteristics that lead to APE
Comparing APE to PD

- Holding time study
- Tracking and timing (mid-project)
- Question / Response study
- Why do visitors leave the exhibit?
Comparing APE to PD

• Holding time study
  – Videotaped exhibits
    • 14 completed APE exhibits
    • 5 PD exhibits
  – Coded for group holding time
Holding time preliminary results

\[ F_{1,971} = 90.6, \ p < .0001 \]
Comparing APE to PD

• Holding time study
  – Videotaped exhibits
    • 14 completed APE exhibits
    • 5 PD exhibits
  – Coded for group holding time
Comparing APE to PD

- Holding time study
- Tracking and timing (mid-project)
Comparing APE to PD

- Holding time study
- Tracking and timing (mid-project)
  - Do APE exhibits somehow attract diligent Vs?
  - Follow same visitors around a small area
    - 4 APE exhibits
    - 6 PD exhibits
  - Analyze behavior of visitors who go to both
    - Individuals, not groups
Tracking & timing sheet

SAMPLE

Date: Day 3
Start Time: 08:00
Finish Time: 10:30
Sex: Male
Age: 25
Size: 1,232 lbs
Total Time: 20:30
Group: A, B, C, D

Comments:

Take the Spring to the Left

10:9

Bike Stage: 156
Calories: 1,102
Distance: 25 miles
Time: 2:10

Total Time: 3:03

Groups:

A: 1
B: 2,6
C: 3
D: 4,5

Water 100%
Tracking & timing results

Mann-Whitney Test $p < .001$
Comparing APE to PD

- Holding time study
- Tracking and timing (mid-project)
Comparing APE to PD

- Holding time study
- Tracking and timing (mid-project)
- Question / Response study
  - What questions do visitors ask at APE & PD?
    - Do APE exhibits promote shift from “Why?” to “What would happen if...?”
  - How do they respond to their own questions?
Question Codes

• Action request
  – *What will happen if we turn it? Can you make it go faster?*
Question Codes

- Action request
- Explanation request
  - Why did that happen? How come it’s faster?
Question Codes

• Action request
• Explanation request
• Orientation request
  – What’s this one? What are you supposed to do?
Question Codes

- Action request
- Explanation request
- Orientation request
- Perception request
  - *Did you see that? Isn’t that beautiful? Is it hot?*
Question Codes

- Action request
- Explanation request
- Orientation request
- Perception request
- Off-task
  - Did you know we’re being videotaped?
Question Codes

- Action request
- Explanation request
- Orientation request
- Perception request
- Off-task

- Label quote / paraphrase
  - *What’s Going On?  What’s Hot, What Not?*
Response Codes

• Uses label
  – Reads label aloud
  – Stares at label (appears to read silently)
  – Paraphrases label in discussion
Response Codes

- Uses label
- Uses exhibit without label
  - Manipulates exhibit components
  - Discusses without involving label
  - Moves body/changes focus of attention
Response Codes

- Uses label
- Uses exhibit without label
- Off-task
Response Codes

- Uses label
- Uses exhibit without label
- Off-task
- No response to question
Groups in Q/R Study

<table>
<thead>
<tr>
<th></th>
<th>Exhibit 1</th>
<th>Exhibit 2</th>
<th>Exhibit 3</th>
<th>Exhibit 4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>APE</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>160</td>
</tr>
<tr>
<td>PD</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>160</td>
</tr>
</tbody>
</table>
## Groups Coded So Far

<table>
<thead>
<tr>
<th></th>
<th>Exhibit 1</th>
<th>Exhibit 2</th>
<th>Exhibit 3</th>
<th>Exhibit 4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>APE</td>
<td>11</td>
<td>11</td>
<td>10</td>
<td>0</td>
<td>32</td>
</tr>
<tr>
<td>PD</td>
<td>12</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>39</td>
</tr>
</tbody>
</table>

Questions coded; responses not yet coded
Q / R Preliminary Results

$F_{1,69} = 7.5, p < .01$
Q / R Preliminary Results

$F_{1,69} = 7.9$, $p < .01$
Q / R Preliminary Results

- How many groups pose each question?
## Q / R Preliminary Results

<table>
<thead>
<tr>
<th>Question type</th>
<th>APE</th>
<th>PD</th>
<th>Chi-square value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action</td>
<td>50%</td>
<td>21%</td>
<td>6.8</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>Explanation</td>
<td>44%</td>
<td>38%</td>
<td>0.2</td>
<td>.65</td>
</tr>
<tr>
<td>Label</td>
<td>16%</td>
<td>10%</td>
<td>Fisher</td>
<td>.72</td>
</tr>
<tr>
<td>Orientation</td>
<td>44%</td>
<td>44%</td>
<td>0.0002</td>
<td>.99</td>
</tr>
<tr>
<td>Perception</td>
<td>72%</td>
<td>46%</td>
<td>4.8</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>Off-task</td>
<td>28%</td>
<td>38%</td>
<td>0.8</td>
<td>.36</td>
</tr>
</tbody>
</table>
Comparing APE to PD

- Holding time study
- Tracking and timing (mid-project)
- Question / Response study
Comparing APE to PD

- Holding time study
- Tracking and timing (mid-project)
- Question / Response study
- Why do visitors leave the exhibit?
  - Exit interviews - 10 APE and 4 PD exhibits
Comparing APE to PD

- Holding time study
- Tracking and timing (mid-project)
- Question / Response study
- Why do visitors leave the exhibit?
  - Exit interviews - 10 APE and 4 PD exhibits
  - Intrinsic
    - “Felt like I had done everything”
  - Extrinsic
    - “Go to the bathroom” or “Wanted to see more”
Why Leave? Preliminary results

<table>
<thead>
<tr>
<th></th>
<th>Intrinsic</th>
<th>Extrinsic</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>APE</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 Exhibits</td>
<td>22%</td>
<td>60%</td>
<td>18%</td>
</tr>
<tr>
<td>208 Visitors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PD</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Exhibits</td>
<td>45%</td>
<td>45%</td>
<td>10%</td>
</tr>
<tr>
<td>40 Visitors</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Chi Square = 9.8; p < .0
Comparing APE to PD

- Holding time study
- Tracking and timing (mid-project)
- Question / Response study
- Why do visitors leave the exhibit?
Comparing APE to PD Summary

- Holding time study
  - APE more engaging (APE > PD)
- Tracking and timing (mid-project)
  - APE more engaging (APE > PD, same visitors)
- Question / Response study
  - Different Q-asking behavior (PD > APE questions/min)
  - Different Qs (APE more Action & Perception Qs)
- Why do visitors leave the exhibit?
  - APE more extrinsic; PD more intrinsic
APE Studies

- Formative Evaluation
  - Describing behavior
  - Comparing iterations of same exhibit

- Summative Evaluation
  - Comparing APE to Planned Discovery

- Research
  - Studying design characteristics that lead to APE
APE Studies

• Formative Evaluation
  – Describing behavior
  – Comparing iterations of same exhibit

• Summative Evaluation
  – Comparing APE to Planned Discovery

• Research
  – Studying design characteristics that lead to APE
APE Research

• Goal
  – Study features that promote APE

• Methods
  – Videotape and interview same visitor groups

• Research focus
  – Multiple Stations
Disparate tools at each station

Circuit Workbench
Single Station

Circuit Workbench
Multiple Station
Identical tools at each station

Spinning Patterns
Single Station

Spinning Patterns
Multiple Station
Mix of tools at each station

Pulley Table
Single Station

Pulley Table
Multiple Station
Types of multiple stations

Different Materials

Circuit Workbench

Identical Materials

Pulley Table

Spinning Patterns
Visitor groups in study

<table>
<thead>
<tr>
<th>Exhibit</th>
<th>Single Station</th>
<th>Multiple Station</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circuit Workbench</td>
<td>46</td>
<td>51</td>
</tr>
<tr>
<td>Spinning Patterns</td>
<td>49</td>
<td>50</td>
</tr>
<tr>
<td>Pulley Table</td>
<td>50</td>
<td>54</td>
</tr>
</tbody>
</table>
Visitor groups analyzed

<table>
<thead>
<tr>
<th>Exhibit</th>
<th>Single Station</th>
<th>Multiple Station</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circuit Workbench</td>
<td>46</td>
<td>51</td>
</tr>
<tr>
<td>Spinning Patterns</td>
<td>49</td>
<td>50</td>
</tr>
<tr>
<td>Pulley Table</td>
<td>50</td>
<td>54</td>
</tr>
</tbody>
</table>
Preliminary Results: Spinning Patterns

• Better Physical Engagement at Multi
  - More groups made patterns
    • 92% at Multi vs. 73% at Single (p < .05)
  - Trend: more groups made complex patterns
    • 44% Multi vs. 27% Single (p < .07)
  - Holding time doubled
    • 7.7 minutes vs. 3.4 minutes (p < .0001)
Preliminary Results: Spinning Patterns

• Better Physical Engagement at Multi
• Better Intellectual Engagement at Multi
  – Discussion of concepts or techniques increased
    • 56% at Multi vs. 38% at Single (p=.05)
Preliminary Results: Spinning Patterns

- Better Physical Engagement at Multi
- Better Intellectual Engagement at Multi
- Better Social Engagement at Multi
  - Less interference
    • 26% at Multi vs. 65% at Single (p < .0001)
  - More teacher/learner collaboration
    • 62% vs. 41% (p < .05)
  - More of group used exhibit together (p < .05)
Preliminary Results: Spinning Patterns

- Better Physical Engagement at Multi
- Better Intellectual Engagement at Multi
- Better Social Engagement at Multi
- Higher Overall Engagement at Multi
Preliminary Conclusions

• Multiple stations solved problems at S.P.
  – Provided visitors with individual control
  – Promoted social interactions; kept families together
  – Building on Minda Borun’s PISEC study
    • Multi-sided and multi-user

• Remaining questions
  – Do disparate tools yield different behaviors?
  • Do visitors move from one station to the next?