735 Experiment Fall 2002
Class Discussion

Jeff Carver
November 26, 2002

Families of Reading Techniques
Background and Motivation

- Study run by Jose Maldonado, et al in Brazil
  - Performed a replication of a study originally run at NASA comparing the effectiveness of PBR and checklist.
  - 4 Replications were Performed
    - Two with less experienced subjects (R1 & R2)
    - Two with more experienced subjects (R3 & R4)

<table>
<thead>
<tr>
<th></th>
<th>PBR</th>
<th>Checklist</th>
</tr>
</thead>
<tbody>
<tr>
<td>NASA 1994</td>
<td>22.9%</td>
<td>18.2%</td>
</tr>
<tr>
<td>NASA 1995</td>
<td>34.4%</td>
<td>33.3%</td>
</tr>
<tr>
<td>R1</td>
<td>15.3%</td>
<td>15.6%</td>
</tr>
<tr>
<td>R2</td>
<td>15.3%</td>
<td>13.5%</td>
</tr>
<tr>
<td>R3</td>
<td>17.7%</td>
<td>14.5%</td>
</tr>
<tr>
<td>R4</td>
<td>22%</td>
<td>17.7%</td>
</tr>
<tr>
<td>ATM</td>
<td>5.8%</td>
<td>5.7%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>PBR</th>
<th>Checklist</th>
</tr>
</thead>
<tbody>
<tr>
<td>NASA 1994</td>
<td>33.9%</td>
<td>24.4%</td>
</tr>
<tr>
<td>NASA 1995</td>
<td>35.2%</td>
<td>23.2%</td>
</tr>
</tbody>
</table>
Study Setting

- **In Vitro**: in the laboratory, under controlled conditions
  - *In Vivo* would mean that it was done in a real world environment
  - Classroom setting imposes its own constraints:
    - Examples
      - No control over subject population
      - Time and organization constraints of the classroom
      - Others??

Study Goals

- **GQM**
  - To analyze:
    - PBR techniques for the purpose of evaluating the level of detail with respect to effectiveness from the point of view of the highly experienced inspector.
    - PBR techniques for the purpose of evaluating the level of detail with respect to effectiveness from the point of view of the low experienced inspector.
    - Training for the purpose of evaluation with respect to effectiveness from the point of view of the researcher.
Study Design

- To evaluate the tailoring, multiple versions of PBR were necessary:

  - Two versions existed
    - One with a complex, detailed underlying model (Equivalence Partition Testing)
    - One with no underlying model
  
  - A new version was needed
    - Simpler, more abstract model (Category Partition Testing)

Study Design

- Potential treatments for this study:

<table>
<thead>
<tr>
<th>Perspective Experience</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process Detail Model</td>
<td>Low Detail</td>
<td>High Detail</td>
<td>High Detail</td>
<td>Low Detail</td>
</tr>
<tr>
<td>Model</td>
<td>Own Model</td>
<td>Abstract Model</td>
<td>Abstract Model</td>
<td>Own Model</td>
</tr>
</tbody>
</table>
Study Design

• Background questionnaire to determine experience level (High or Low)
  • Tester perspective was chosen
    • 6 high experience subjects
    • 16 low experience subjects
      • 4 were in Dr. Memon’s testing class and already learned about CPT
        • Placed in Group 3
      • 12 had never seen CPT
        • Split evenly between Groups 3 and 4

Study Design

• Treatments

<table>
<thead>
<tr>
<th>Perspective Experience</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process Detail</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Model</td>
<td>Own Model</td>
<td>Abstract Model</td>
<td>Abstract Model</td>
<td>Own Model</td>
</tr>
<tr>
<td>Number of Subjects</td>
<td>6</td>
<td>10</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>
GQM: Questions and Metrics

• Goal:
  - To analyze PBR techniques for the purpose of evaluating the level of detail with respect to effectiveness from the point of view of the highly experienced inspector.

• Question:
  - What affect does the level of detail in the PBR techniques have on the high experienced subjects?

• Metrics
  - Percentage of defects found by members of Group 1
  - Historical Data from previous 735 class (1997) and two replications from Brazil mentioned earlier
GQM: Questions and Metrics

- **Goal:**
  - To analyze PBR techniques for the purpose of evaluating the level of detail with respect to effectiveness from the point of view of the low experienced inspector.

- **Question:**
  - What affect does previous knowledge of the underlying model in the PBR techniques have on the subjects who use that model?

- **Metrics**
  - Percentage of defects found by members of Group 3 who were not in Dr. Memon’s class
  - Percentage of defects found by members of Group 3 who were in Dr. Memon’s class

GQM: Questions and Metrics

- **Goal:**
  - To analyze training for the purpose of evaluation with respect to effectiveness from the point of view of the researcher.

- **Question:**
  - What affect did the training have on the subjects?

- **Metrics**
  - Pretest answers
  - Posttest answers
  - Post-experiment questionnaire
GQM: Questions and Metrics

- **Goal:**
  - To analyze PBR techniques for the purpose of evaluating the level of detail with respect to effectiveness from the point of view of the high and low experienced inspector.

- **Question:**
  - How can the versions of PBR used in this study be improved?

- **Metrics**
  - Post-experiment questionnaire data

---

**Compare level of detail for Low Experience subjects**

- **Compare level of detail for Low Experience subjects**
- Determine if a complex model is better or worse than an abstract model
Data Collection

- Defect lists submitted after inspections
  - Experimenters went through the defect lists and ‘scored’ them based on the master defect list
    - Give an initial idea of the defects
  - Master defect list sent out for comments
    - Understand quality of our defect list
  - Mapping of defects to master list sent out for comments
    - Verifying the mapping
## Initial Results

<table>
<thead>
<tr>
<th>Perspective Experience</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process Detail</td>
<td>Low Detail</td>
<td>High Detail</td>
<td>High Detail</td>
<td>Low Detail</td>
</tr>
<tr>
<td>Model</td>
<td>Own Model</td>
<td>Abstract Model</td>
<td>Abstract Model</td>
<td>Own Model</td>
</tr>
<tr>
<td>Number of Subjects</td>
<td>6</td>
<td>10</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Defect Rate</td>
<td>22.1%</td>
<td>20.6%</td>
<td>26.5%</td>
<td></td>
</tr>
</tbody>
</table>

**Percent Defects Found**

- No Model - High Experience
- No Model - Low Experience

**Outlier?**
Outlier?

- A data point that is either from a different population or a very unusual member of the same population
- Computed as follows:
  - Lower Quartile $Q_L = 25^{th}$ percentile of data
  - Upper Quartile $Q_U = 75^{th}$ percentile of data
  - Interquartile Range (IQR) = $Q_U - Q_L$
  - Inner Fence = $Q_L + 1.5(IQR)$; Outer Fence = $Q_U + 3(IQR)$
- Our data (No model – Low Experience group):
  - $Q_L = .235$; $Q_U = .257$; IQR = .257 - .235 = .022
  - Inner = .257 + 1.5(.022) = .29; Outer = .257 + 3(.022) = .324

Initial Results
Initial Results

<table>
<thead>
<tr>
<th>Perspective Experience</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process Detail -- Model</td>
<td>LowDetail -- Own Model</td>
<td>HighDetail -- Abstract Model</td>
<td>HighDetail -- Abstract Model</td>
<td>Low Detail -- Own Model</td>
</tr>
<tr>
<td>Number of Subjects</td>
<td>5</td>
<td>10</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Defect Rate</td>
<td>24.1%</td>
<td>20.6%</td>
<td>21.8%</td>
<td></td>
</tr>
</tbody>
</table>

Question:

What affect does the level of detail in the PBR technique have on the high experienced subjects?

- Low Process Detail
  - This study – 24.1%

- High Process Detail
  - 735(97) [Equivalence Partitioning] – 27.1%
  - R3 [Equivalence Partitioning] – 17.7%
  - R4 [Equivalence Partitioning] – 8.7%

(R3 and R4 results were significant)
Initial Results

• Question:
  • What affect does the level of detail in the PBR technique have on the low experienced subjects?
    • Low Process Detail
      • No Model Group (this study) – 21.8%
    • High Process Detail
      • Abstract, simple model
        • CPT Group (this study) – 20.6%
      • Concrete, complex model
        • 735(99) [Equivalence Partitioning] – 27.1%
    • (None of the results are significant)

Initial Results

• Question:
  • What affect does previous knowledge of the underlying model in the PBR techniques have on the subjects who use that model?
    • Subjects with experience – 22.8%
    • Subjects without experience – 19.1%
    • (Not statistically significant)
Initial Results

- **Question:**
  - What affect did the training have on the subjects?
    - **Quantitative**
      - Defects
        - Pretest average = 3.7 → Posttest average = 4.4
      - Test cases
        - Pretest average = 2.7 → Posttest average = 4.3
    - **Qualitative**
      - Effectiveness
        - 12 out of 22 said it was effective and adequate
        - 3 said it was not (only 1 was using CPT)
      - 10 out of 22 wanted more examples
      - 4 out of 22 wanted more test case examples

- **Question:**
  - How can the versions of PBR used in this study be improved?
    - 9 out of 10 users of CPT thought there was enough detail in the technique to do their job
    - Members of the high experience thought the detail was sufficient, although most wished for more example test cases
    - Members of the low experience group, not given CPT, were split
      - Half thought there was enough detail
      - Half thought there was not enough detail
Threats to Validity

• Internal
  • History
    • Results of later treatments may be attributed to events that occurred between treatments
  • Maturation
    • Processes occurring within subjects may change over time
  • Testing
    • Results may vary over time as subjects get more comfortable with testing procedures

• Instrumentation
  • Results may differ with different measures

• Selection
  • Results may differ because of the type of subjects in different groups

• Process Conformance
  • Results may differ because procedure was not followed.
Threats to Validity

• External
  • Are the results valid outside of this class?
    • Professional developers?
    • Real projects?
  • Are results valid for other requirements documents?
    • Different formats?
    • Different domains?
    • Different languages?

What is left to do?

• More analysis on your comments about our defect list
• Analyze results looking at:
  • Individual Defects
  • Classes of Defects
  • Make sure the experience levels of this class and the historical data are comparable
Discussion

• How was the Process Conformance?