Value-Based Peer Review

CSE Annual Research Review 2006

by

Keun Lee
(keunlee@sunset.usc.edu)

Monvorath Phongpaibul (Molly)
(phongpai@usc.edu)
Content

• Hypothesis

• Value-Based Review

• Experiment Results

• Conclusions of the experiment
Hypothesis

• Value-neutral software peer reviews misallocate effort
  • All requirements, use cases, objects, defects are equally important
  • Too much effort is spent on trivial issues

• Current status checklist, defect function-based reviews are largely value-neutral
Value-Based Review Process

Developers

Negotiation

Meetings

Users

Priorities of system capabilities

Criticalities of issues

Reviewing Artifacts

<table>
<thead>
<tr>
<th>Priority</th>
<th>Criticality</th>
<th>1</th>
<th>4</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>High</td>
<td>1</td>
<td>4</td>
<td>6</td>
</tr>
</tbody>
</table>
| Medium     | Medium      | 2 | 5 | opt
| Low        | Low         | 3 | opt | opt |

Number indicates the usual ordering of review*

* May be more cost-effective to review highly-coupled mixed-priority artifacts.

Domain Expert

General Value-based checklist

Artifact-oriented checklist

Customers

Other stakeholders

Priorities of system capabilities

Criticalities of issues

Artifacts

Number indicates the usual ordering of review*
## Artifact-oriented Value-Based Checklist

*Example: OCD 4.3 system capability*

<table>
<thead>
<tr>
<th>Question</th>
<th>Criticality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Are the system capabilities consistent with the system services provided as described in OCD 2.3?</td>
<td>3</td>
</tr>
<tr>
<td>Are there critical missing capabilities needed to perform the system services?</td>
<td>3</td>
</tr>
<tr>
<td>Are capabilities prioritized as High, Medium, or Low?</td>
<td>3</td>
</tr>
<tr>
<td>Are capability priorities consistent with current system shortcoming priorities (OCD 3.3.5)?</td>
<td>3</td>
</tr>
<tr>
<td>Are capabilities traced back to corresponding project goals and constraints (OCD 4.2)?</td>
<td>3</td>
</tr>
<tr>
<td>Are simple lower-priority capabilities (e.g., login) described in less detail?</td>
<td>2</td>
</tr>
<tr>
<td>Are there no levels of service goals (OCD 4.4) included as system capabilities?</td>
<td>2</td>
</tr>
</tbody>
</table>
Weight of Review Issues

Effectiveness Metric = \sum_{\text{issues}} (\text{Artifact Priority}) \times (\text{Issue Criticality})

<table>
<thead>
<tr>
<th>Artifact Priority</th>
<th>H</th>
<th>M</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>9</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>M</td>
<td>6</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>L</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Generally considered optional to review

* Numeric Value: H = 3, M = 2, L = 1
V&V Experiment Results

<table>
<thead>
<tr>
<th>By Number</th>
<th>P-value</th>
<th>% Gr A higher</th>
<th>By Impact</th>
<th>P-value</th>
<th>% Gr A higher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average of Concerns</td>
<td>0.202</td>
<td>34</td>
<td>Average Impact of Concerns</td>
<td>0.049</td>
<td>65</td>
</tr>
<tr>
<td>Average of Problems</td>
<td>0.056</td>
<td>51</td>
<td>Average Impact of Problems</td>
<td>0.012</td>
<td>89</td>
</tr>
<tr>
<td>Average of Concerns per hour</td>
<td>0.026</td>
<td>55</td>
<td>Average Cost Effectiveness of Concerns</td>
<td>0.004</td>
<td>105</td>
</tr>
<tr>
<td>Average of Problems per hour</td>
<td>0.023</td>
<td>61</td>
<td>Average Cost Effectiveness of Problems</td>
<td>0.007</td>
<td>108</td>
</tr>
</tbody>
</table>

- Group A averaged over twice the cost-effectiveness in finding concerns and problems.

- Group B had significantly higher numbers of trivial concerns and problems found (typo and grammar faults)
Conclusions of the experiment

Conclusions: At least in this small-team, remote IV&V context,

- Value-based reviews had significantly higher payoff than value-neutral reviews
  - Factor of 2 in cost-effectiveness

- With statistical significance for concerns and problems per hour, value impact, and value impact per hour

- VBR Required minimum effort comparing with CBR

- VBR checklists were helpful to understand and review artifacts.
Question & Answer

-Thank You-
Motivation

Value-neutral review vs. Value-based review

ROI

% Reviews Run

Value-based review

Value-neutral review
## General Value-Based Checklist

<table>
<thead>
<tr>
<th></th>
<th>High-Criticality Issues</th>
<th>Medium-Criticality Issues</th>
<th>Low-Criticality Issues</th>
</tr>
</thead>
</table>
| **Completeness**               | • Critical missing elements: backup/ recovery, external interfaces, success-critical stakeholders, critical exception handling, missing priorities  
• Critical missing processes and tools; planning and preparation for major downstream tasks (development, integration, test, transition)  
• Critical missing project assumptions (client responsiveness, COTS adequacy, needed resources) | • Medium-criticality missing elements, processes and tools: maintenance and diagnostic support; user help  
• Medium-criticality exceptions and off-nominal conditions; smaller tasks (review, client demos), missing desired growth capabilities, workload characterization | • Easily-deferrable, low-impact missing elements: straightforward error messages, help messages, GUI details doable via GUI builder, project task sequence details |
| **Consistency/Feasibility**    | • Critical elements in OCD, SSRD, SSAD, LCP not traceable to each other  
• Critical inter-artifact inconsistencies: priorities, assumptions, input/output, preconditions/post-conditions  
• Missing evidence of critical consistency/feasibility assurance in FRD | • Medium-criticality shortfalls in traceability, inter-artifact inconsistencies, evidence of consistency/feasibility in FRD | • Easily-deferrable, low-impact inconsistencies or inexplicit traceability: GUI details, report details, error messages, help messages, grammatical errors |
| **Ambiguity**                 | • Vaguely defined critical dependability capabilities: fault tolerance, graceful degradation, interoperability, safety, security, survivability  
• Critical misleading ambiguities: stakeholder intent, acceptance criteria, critical user decision support, terminology | • Vaguely defined medium-criticality capabilities, test criteria  
• Medium-criticality misleading ambiguities | • Non-misleading, easily deferrable, low-impact ambiguities: GUI details, report details, error messages, help messages, grammatical errors |
| **Conformance**               | • Lack of conformance with critical operational standards, external interfaces | • Lack of conformance with medium-criticality operational standards, external interfaces  
• Misleading lack of conformance with document formatting standards, method and tool conventions | • Non-misleading lack of conformance with document formatting standards, method and tool conventions, optional or low-impact operational standards |
| **Risk**                      | • Missing FRD evidence of critical capability feasibility: high-priority features, levels of service, budgets and schedules  
• Critical risks in top-10 risk checklist: personnel, budgets and schedules, requirements, COTS, architecture, technology | • Missing FRD evidence of mitigation strategies for low-probability high-impact or high-probability, low-impact risks: unlikely disasters, off-line service delays, missing but easily-available information | • Missing FRD evidence of mitigation strategies for low-probability, low-impact risks |
V&V Experiment Result
Effort Comparison

- No differences in effort between group A and group B at level 0.05 statistically

- VBR requires less effort (22.89%) for preparation, even both techniques are proved that require same preparation effort statistically.