Value-Based Peer Review

CSE Annual Research Review 2006

by

Keun Lee
(keunlee@sunset.usc.edu)
Monvorath Phongpaibul (Molly)
(phongpai@usc.edu)

Content

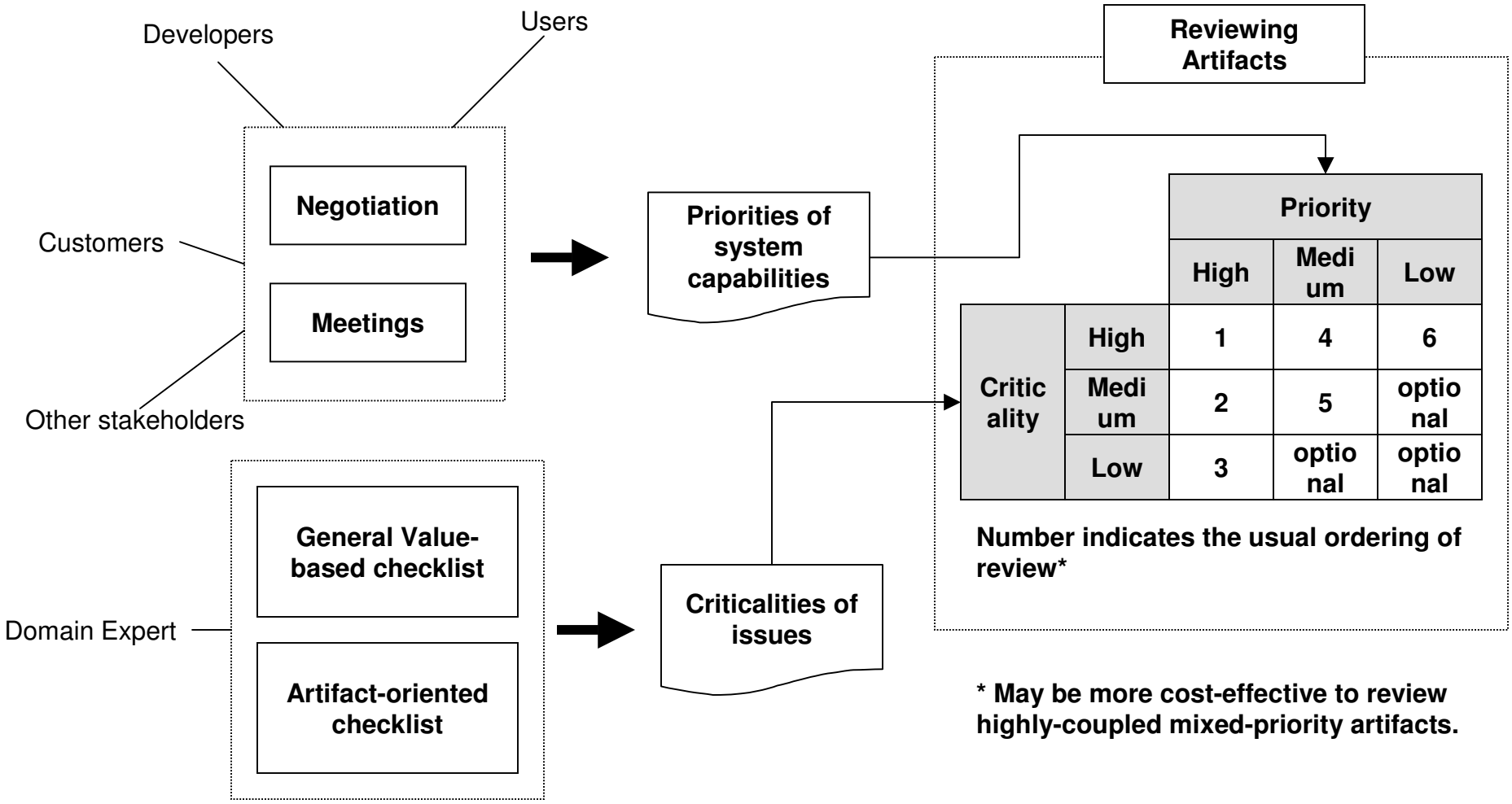
- Hypothesis
- Value-Based Review
- Experiment Results
- Conclusions of the experiment

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



Artifact-oriented Value-Based Checklist

<Example : OCD 4.3 system capability>

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

Weight of Review Issues

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

Issue Criticality \ Artifact Priority	H	M	L
H	9	6	3
M	6	4	2
L	3	2	1

Generally considered optional to review

* Numeric Value: H = 3, M = 2, L = 1

V&V Experiment Results

By Number	P-value	% Gr A higher	By Impact	P-value	% Gr A higher
Average of Concerns	0.202	34	Average Impact of Concerns	0.049	65
Average of Problems	0.056	51	Average Impact of Problems	0.012	89
Average of Concerns per hour	0.026	55	Average Cost Effectiveness of Concerns	0.004	105
Average of Problems per hour	0.023	61	Average Cost Effectiveness of Problems	0.007	108

- 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,

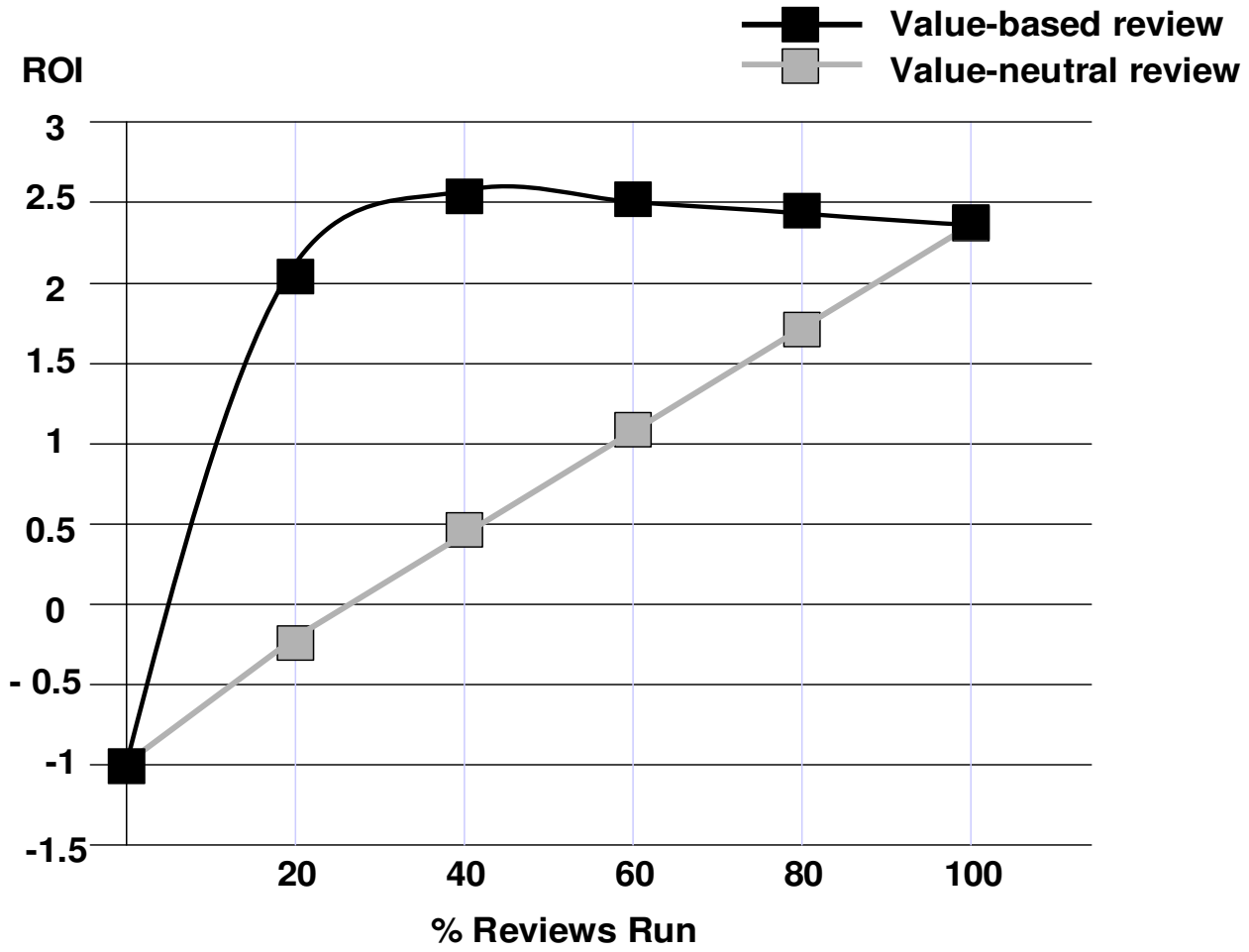
- Value-based reviews had significantly higher payoff than value-neutral reviews
  - Factor of 2 in cost-effectiveness
- With statistical significance for concerns and problems per hour, value impact, and value impact per hour
- VBR Required minimum effort comparing with CBR
- VBR checklists were helpful to understand and review artifacts.

Question & Answer

-Thank You-

Motivation

Value-neutral review vs. Value-based review

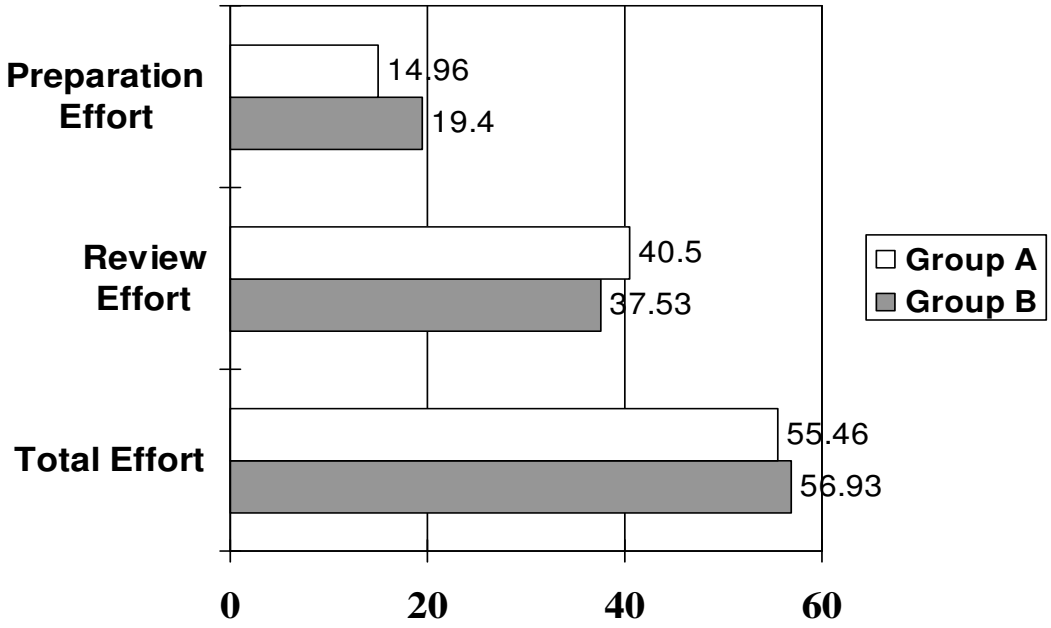


General Value-Based Checklist

	High-Criticality Issues	Medium-Criticality Issues	Low-Criticality Issues
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



Effort	P-value
Preparation Effort	0.093
Review Effort	0.633
Total Effort	0.726

- 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.