Investigating the Temporal Behavior of Defect Detection in Software Inspection and Inspection-Based Testing

Dietmar Winkler Stefan Biffl Kevin Faderl

Institute of Software Technology and Interactive Systems,
Vienna University of Technology

dietmar.winkler@tuwien.ac.at

http://qse.ifs.tuwien.ac.at

Motivation & Goals

- Early detection and removal of defects, e.g., in the design phase, helps increasing software quality and decrease rework effort and cost.
- Analytical Quality Assurance (QA) typically includes
  - Reviews and Inspection for systematic V&V in early phases.
  - Software Testing, i.e., test case definition and execution in late phases.

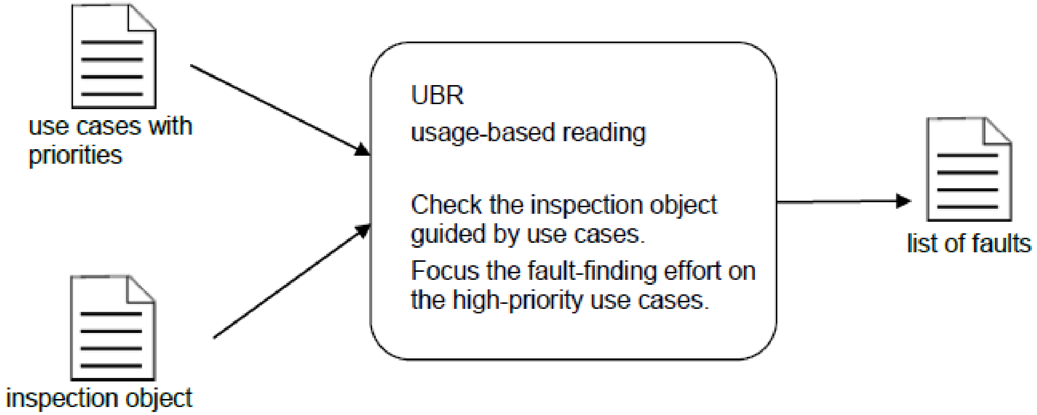
- Goal is the early definition of test cases based on inspection results (test-first approach).
  - Early defect detection as contribution of software inspection.
  - Improved understanding of customer requirements.
  - Test case generation based on requirements and inspection results.

- Key research questions focus on:
  - How can inspection support early test case definition?
  - What are the effects on defect detection performance (inspection vs. inspection-based test case generation)?

Software Inspection

- Software Inspection …
  - is a static analysis technique to verify quality properties of software artifacts.
  - does not require executable code (applicable to design documents).
  - focuses on defined defect types and locations in the inspected object.
  - provides active guidance of inspectors with reading techniques and guidelines (how to traverse a software document).

- "Best-practice" approach: Usage-Based Reading (UBR)
  - Well-investigated reading technique approach in business IT software development projects.
  - Focus on users and use cases.
  - Prioritization acc. to value/risk.
  - Application of use cases and scenarios.



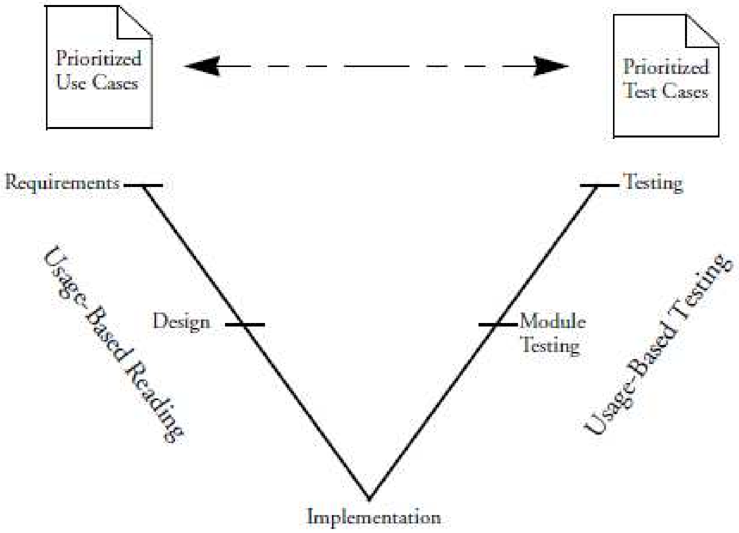
UBR
usage-based reading

Check the inspection object guided by use cases.
Focus the fault-finding effort on the high-priority use cases.

Usage Based Testing with Inspection

Usage-Based Testing (UBT)

- Test case generation based on use cases.
- Prioritized test cases late in the development process.
- Focus on executable code.



Usage-Based Testing with inspection (UBT-i)

- Bundling benefits of early Inspection and UBT
  - Early defect detection with inspection.
  - Early test-case definition based on prioritized use cases.

- Previous studies showed benefits of UBT-i with respect to isolated best-practice inspection.
- Empirical study on the temporal behavior of defect detection performance.

Dependent Variables and Hypothesis

Performance measures:

- Inspection effort includes individual preparation time and inspection / test-case generation duration.
- Effectiveness is the number of defects related to the overall number of seeded (and important) defects.
- Efficiency is the number of defects found per time interval.
- False Positives is the number of "wrong defects detected" by individual participants.

Time limitations:

- Upper study execution was 300 min (5hrs).
- Suggestions for review/inspection duration: 120 min (focus of the evaluation)

Hypothesis:

- Higher effectiveness (H1) & efficiency (H2) for UBR within 120 min of method application.
- Higher number of false positives (H3) in UBR within 120 minutes.

Experiment Description

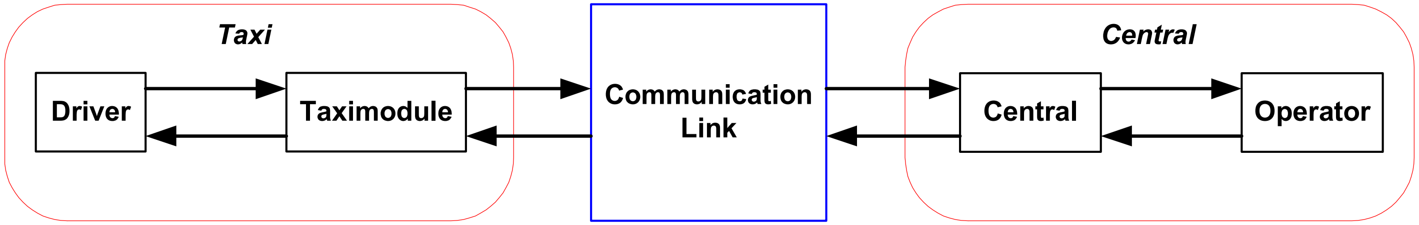
- Experiment Phases

  (a) Training & preparation.

  (b) Study execution: briefing, session 1 (taxi) and session 2 (central).

  (c) Data submission and evaluation.

- Subjects are 41 master students with SE and QA background.



- Study Material is a Taxi Management System in 2 parts
  - Snapshot of an agile software development project.
  - Textual requirements specification (8p), 2 Component diagrams, design document (8p), 24 prioritized use cases, appx. 1400 LOCs.
  - 60 seeded defects at 3 defect severity classes (29 defects in the taxi part and 31 in the central part)
  - Supporting material: guidelines and questionnaires.

Threats to Validity

Internal validity:

- Avoidance of communication between individuals during the study execution.
- Participants could take individual brakes, whenever necessary (break durations reported).
- Limitation of the overall study duration was 300 minutes, focus on the first 120 minutes (suggested inspection durations).
- Experience questionnaire to get an insight on prior knowledge.
- Feedback questionnaire to see if the participants followed the study process properly and to capture individual strategies.

External validity:

- Well-known application domain to avoid domain-specific interpretation problems.
- Pilot test and reviews to assure correctness of experiment material.
- Control of variables due classroom setting.

Effort and First Real Defect Reported

Effort

- No significant differences between UBR and UBT-i in both sessions.

Duration [min]	Session 1 (Taxi)		Session 2 (Central)	
	UBR	UBT-i	UBR	UBT-i
No of Subjects	20	21	21	20
Mean	272,5	268,8	281,3	276,2
Std.Dev.	38,01	29,13	35,32	30,11

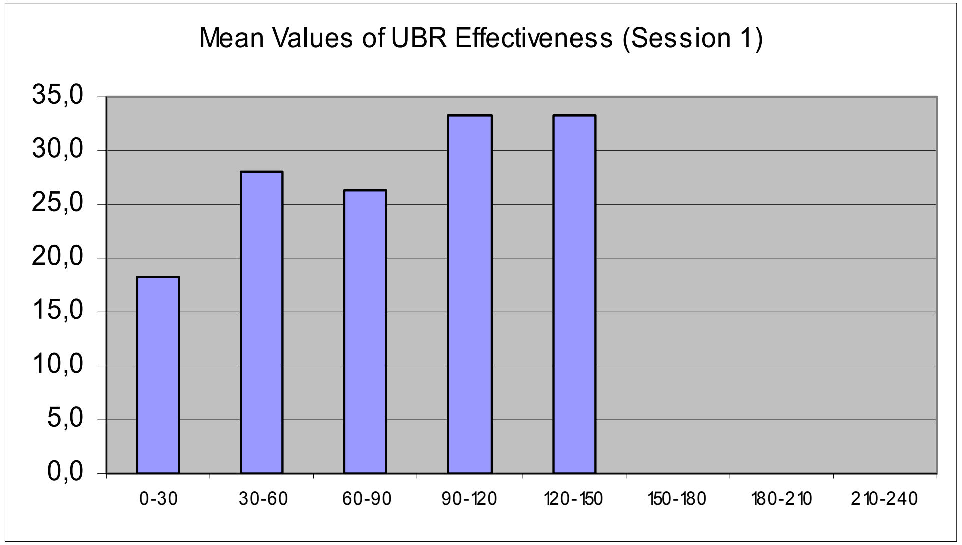
First Real Defect Reported

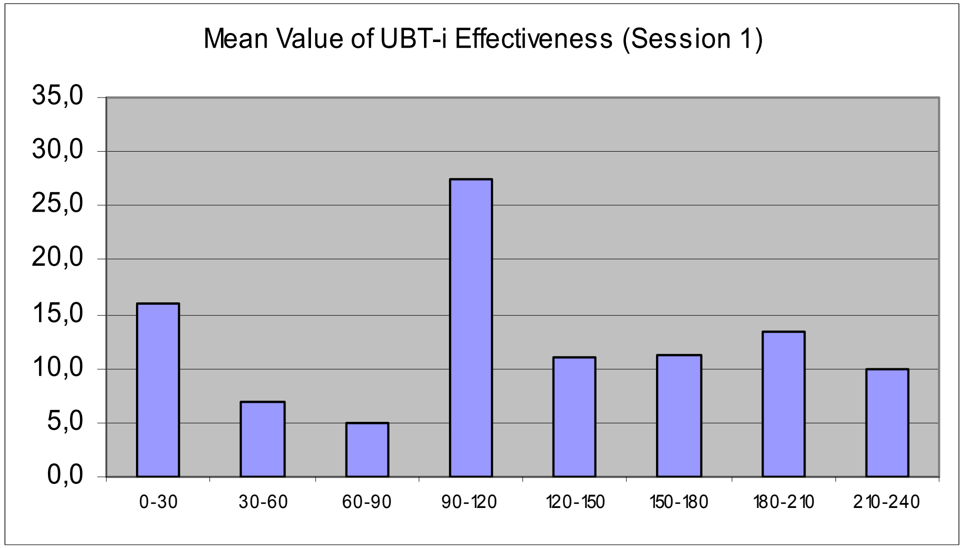
- Significant differences in session 1 and session 2

	Session 1 (Taxi)		Session 2 (Central)	
	UBR	UBT-i	UBR	UBT-i
Mean	12,2	17,6	15,4	17,4
Std.Dev.	10,59	10,39	10,93	10,42

Results: Effectiveness

- Expectations: UBR is significantly more effective than UBT-i.
- Comparable overall effectiveness for critical and important defects
  - 18.9 (UBR) and 16.9 (UBT-i), no significant differences.
- Time Interval Evaluation (first session)
  - No significant differences during the first 30 min.
  - Significant differences for all other time intervals.
  - No matched defects for t>240





- Assuming additional effort for test case generation might limit defect detection effectiveness.

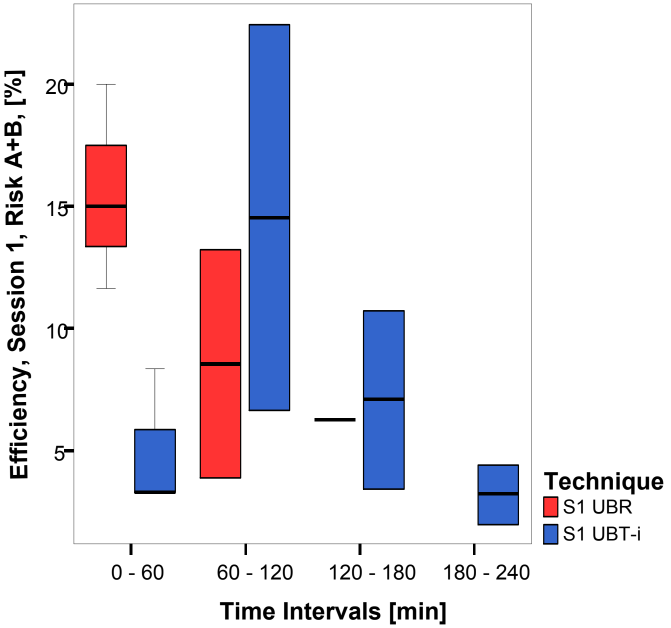
Results: Efficiency

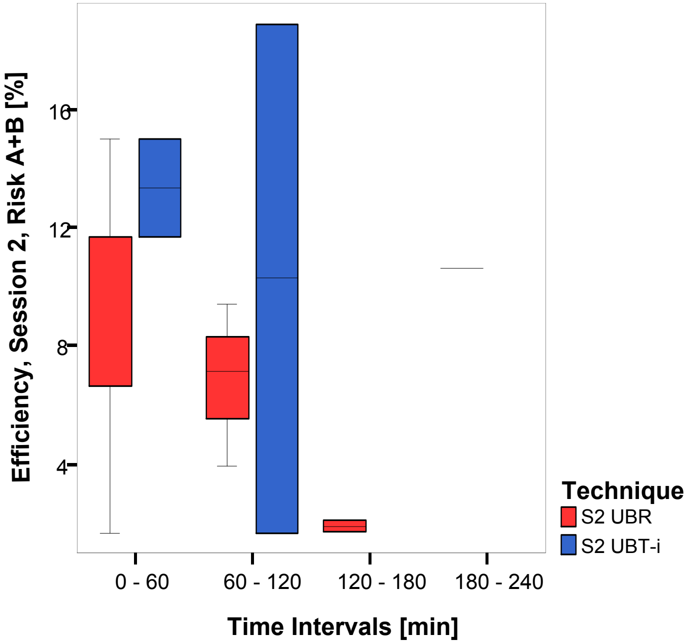
- Expectations: UBR is significantly more efficient than UBT-i.
- Measurement: Defects per hour.

- Session 1:
  - UBR are most efficient in the first hour.
  - UBT-i is most efficient in the second hour.
- Session 2:
  - UBT-i outperforms UBR inspection in the first 2 time intervals.

- Significant differences between all groups.

- Possible explanation is a changed defect detection approach of UBT-i in the second session: defect detection and test case generation in sequential order.

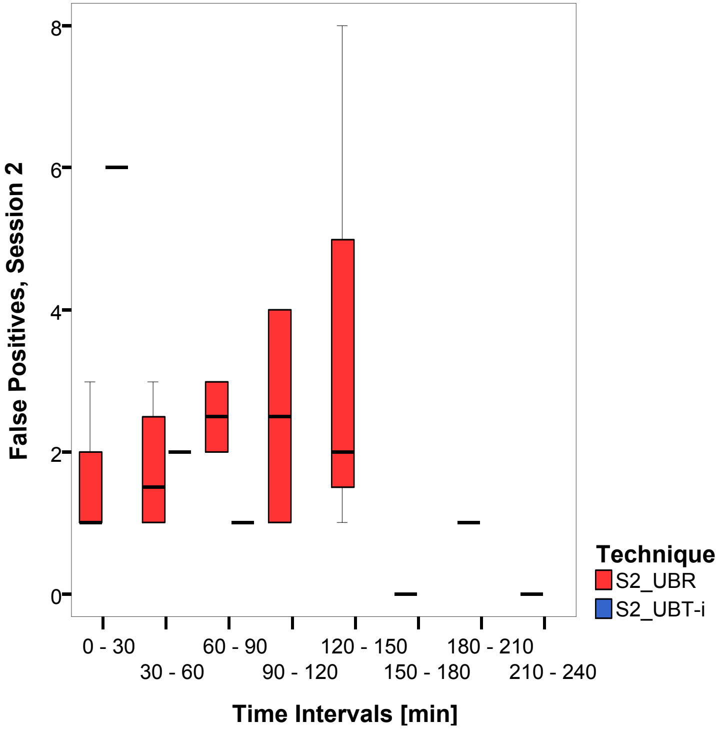




Results: False Positives

- Expectations: UBR report significantly more false positives (FP).

- Session 1:
  - Higher amount of FP at the begin and at the end of the study duration.
- Session 2:
  - Decreasing number of FP during the course of the study for UBT-i.
  - Increasing number of FP during the course of the study (up to 150 min) for UBR.



- Significant differences for all time intervals.
- Possible explanation seems to be a strong advantage for UBT-i participants who focused on test case generation (i.e., testability considerations).
- Nevertheless, results of effectiveness and false positives must be investigated in more detail for verification of the results strengthening the findings.

Summary and Further Work

Summary:

- Test case generation based on inspection results is a promising approach for bundling benefits of early defect detection.
- UBR performed very effective and efficient in a time interval up to 120 min.
- UBT-i requires more time for test case generation to achieve comparable defect detection results.
- We observed partly benefits in certain time intervals and notable differences between the two sessions → further investigations are required.
- Support of planning QA activities in SE projects.

Further Work:

- More detailed investigation of the study outcome regarding performance measures, defect types and document locations.
- In-depth analysis of the temporal behavior of defect detection performance because of partly contradictory results in two study sessions.

Thank you ...

Investigating the Temporal Behavior of Defect Detection in Software Inspection and Inspection-Based Testing

Dietmar Winkler, Stefan Biffl, Kevin Faderl

Vienna University of Technology
Institute of Software Technology and Interactive Systems

http://qse.ifs.tuwien.ac.at
Dietmar.Winkler@tuwien.ac.at