An Experimental Validation of MetricView Visualizations

Christian Lange

www.win.tue.nl/~clange
c.f.j.lange@tue.nl

EmpAnADa Project
Empirical Analysis of Architecture and Design Quality

TU/e
Outline

- About *MetricView Evolution*
- Purpose of the Experiment
- Design of the Experiment
- Results
- Conclusions & Future Work
- Discussion
About MetricView Evolution

- UML models are becoming central artefacts
- Larger variety of activities performed on models
- Can UML and existing tools deals with these activities…?
  - Traceability
    - Which use cases require class X?
    - In which Sequence Diagrams is class X instantiated?
  - Metrics
    - Why has class X a coupling metric of 33?
  - External Data
    - Which class contains the most bugs?
  - Evolution
    - How did the quality of class X evolve over the past versions?
  - …
Example Views

- Demo
  - MetricView Evolution
Purpose of the Experiment

- **Purpose**
  - Validation of MetricView visualizations for comprehension of UML models

- **Research Questions**
  - RQ 1: Is the **correctness** improved by using MV?
  - RQ 2: Is the **effort** reduced by using MV?

- **Focus on**
  - Structural Visualizations (Navigation)
  - Metrics Visualization (Quality Assessment)
Experimental Task

- **Task:** Subjects had to answer a questionnaire about a UML model using a tool.
Questionnaire

Task: Subjects had to answer a questionnaire about a UML model using a tool

- Comprehension Task
- Multiple-choice
- 29 questions
- 4 categories of questions
  - Structure within class diagram(s)
  - Interactions (SD – CD relations)
  - Functionality (UC – CD – SD relations)
  - Metrics
Questionnaire II (Examples)

- **Category 1 (Class Structure)**
  - To which classes is the class *Obstruction_DB* coupled by associations?

- **Category 2 (Interaction)**
  - Which classes call methods of class *CurrentLocation*?

- **Category 3 (Functionality)**
  - Which classes contribute to the use case “*Leave Route*”?

- **Category 4 (Metrics)**
  - In which class diagram does the class with the largest “*Number of Attributes*” occur?
Task: Subjects had to answer a questionnaire about a UML model using a tool.

- 2 different models
- Application Domain
  - Insurance Information System
  - Car Navigation System
- 39 classes
  - 5 class diagrams
  - 5 sequence diagrams
  - 11 use cases
Experimental Treatment

- **Task:** Subjects had to answer a questionnaire about a UML model using a tool.

- **Treatment:**
  - MetricView Evolution

- **Control Group**
  - Poseidon UML & SDMetrics
Subjects

Task: Subjects had to answer a questionnaire about a UML model using a tool

- 100 subjects
- MSc students in the course `Software Architecting`
- Sufficient background in UML
Operation

- **Preparation**
  - Lectures
  - 2 assignments
  - Pilot run

- **Details**
  - Randomization
  - Exam-like setting
  - Debriefing questionnaire after each run

- **Design**

  First Run
  - Insurance Model
  - Group A
  - Group B

  Second Run
  - Car Navigation Model
  - Group B
  - Group A

  MetricsView Evolution
  - Poseidon & SDMetrics
Results: Correctness

Variable:

\[
\frac{\text{# Correct answers}}{\text{# Questions}}
\]

Test:
- Mann-Whitney &
- Student t

<table>
<thead>
<tr>
<th></th>
<th>First Run</th>
<th>Second Run</th>
</tr>
</thead>
<tbody>
<tr>
<td>MVE</td>
<td>.837</td>
<td>.921</td>
</tr>
<tr>
<td>PoS</td>
<td>.771</td>
<td>.881</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>MVE</th>
<th>PoS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>.837</td>
<td>.771</td>
</tr>
<tr>
<td>Difference</td>
<td>+ 8.5 %</td>
<td>+4.5 %</td>
</tr>
<tr>
<td>P-value</td>
<td>&lt;0.001</td>
<td>0.002</td>
</tr>
</tbody>
</table>
Results: Effort

- **Variable:**
  - Time (minutes)

- **Test:**
  - Mann-Whitney &
  - Student t

<table>
<thead>
<tr>
<th></th>
<th>First Run</th>
<th>Second Run</th>
</tr>
</thead>
<tbody>
<tr>
<td>MVE</td>
<td>54.6</td>
<td>52.5</td>
</tr>
<tr>
<td>PoS</td>
<td>67.6</td>
<td>66.0</td>
</tr>
</tbody>
</table>

| Difference  | -19.2 %   | -20.6 %    |

| P-value     | <0.001    | <0.001     |
Debriefing Questionnaire

- No differences between groups
  - Skill & knowledge
  - Motivation

- No differences between models/runs
  - Quality, understandability, completeness

- Differences between runs for Correctness
  - Learning effects
    - Tool
    - Task
Conclusions & Future Work

- MetricView visualizations improve Model Comprehension
  - Effort (20%)
  - Correctness (4% - 8%)

- Subjective Evaluation Results → low Adoption Threshold

- Future work
  - Investigate Learning Effects
  - Variations in Comprehension Task + Change Task
  - Further Improvement of Visualizations
An Experimental Validation of *MetricView* Visualizations

www.win.tue.nl/empanada

*EmpAnADa Project*
*Empirical Analysis of Architecture and Design Quality*

TU/e technische universiteit eindhoven