A Controlled Experiment for Evaluating Quality Guidelines on the Maintainability of Object-Oriented Designs

Briand, et al

IEEE TOSE June 2001
What is problem?
What did you learn?

- Quality Design
- Experimentation
Exp Context Guidelines

C1: Be sure to specify as much of the industrial context as possible. In particular, clearly define the entities, attributes, and measures that are capturing the contextual information.
Null Hypothesis

C2: If a specific hypothesis is being tested, state it clearly prior to performing the study and discuss the theory from which it is derived, so that its implications are apparent.

Standard significance testing was used to clearly specify these effects, the null hypothesis being stated as:

\[ H_0: \text{There is no significant difference between “good” and “bad” object-oriented design in terms of ease of understanding and impact analysis.} \]
Exp Context Guidelines

C3: If the research is exploratory, state clearly and, prior to data analysis, what questions the investigation is intended to address and how it will address them.

C4: Describe research that is similar to, or has a bearing on, the current research and how current work relates to it.
Exp Design Guidelines

D1: Identify the population from which the subjects and objects are drawn.

D2: Define the process by which the subjects and objects were selected.

D3: Define the process by which subjects and objects are assigned to treatments.
Experimental Design

- Subjects?
- Table 3 p518
  - 2 by 2 factorial design
  - within-subjects
  - learning and fatigue
Guidelines

- Coupling
- Cohesion
- Clarity of design
- Generalization-specialization depth
- Keeping objects and classes simple
Maintenance Task

◆ Questions
  – Structure
  – Specific question

◆ Impact Analysis – 21/22 changes
  – Requirements change
  – Enhancement of functionality
Exp Design Guidelines

D7: Use appropriate levels of blinding.

D8: If you cannot avoid evaluating your own work, then make explicit any vested interests (including your sources of support) and report what you have done to minimize bias.

D9: Avoid the use of controls unless you are sure the control situation can be unambiguously defined.
Exp Design Guidelines

D10: Fully define all treatments (interventions).

D11: Justify the choice of outcome measures in terms of their relevance to the objectives of the empirical study.
Dependent Variables

- Und_Time
- Und_Corr
- Mod_Time
- Mod_Comp
- Mod_Corr
- Mod_Rate
Table 1

<table>
<thead>
<tr>
<th>Counts</th>
<th>Bad Original</th>
<th>Bad Used</th>
<th>Good Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>classes</td>
<td>14</td>
<td>18</td>
<td>13</td>
</tr>
<tr>
<td>inheritance</td>
<td>4</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>relationships</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Associations</td>
<td>12</td>
<td>16</td>
<td>9</td>
</tr>
<tr>
<td>Attributes</td>
<td>24</td>
<td>30</td>
<td>16</td>
</tr>
<tr>
<td>Operations</td>
<td>20</td>
<td>25</td>
<td>18</td>
</tr>
<tr>
<td>max. inheritance</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Depth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inconsistencies</td>
<td>0</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Aggregations</td>
<td>2</td>
<td>2</td>
<td>8</td>
</tr>
</tbody>
</table>
How Good and Bad are created?
Table 2

<table>
<thead>
<tr>
<th>Measure</th>
<th>Bad Original</th>
<th>Bad Used</th>
<th>Good Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIT</td>
<td>0.29</td>
<td>0.67</td>
<td>0.22</td>
</tr>
<tr>
<td>NOC</td>
<td>0.29</td>
<td>0.45</td>
<td>0.22</td>
</tr>
<tr>
<td>CBO</td>
<td>2.90</td>
<td>3.10</td>
<td>2.6</td>
</tr>
</tbody>
</table>
Table 4

Descriptive Statistics for Each Dependent Variable

<table>
<thead>
<tr>
<th>Variable</th>
<th>Good OO Design</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Bad OO Design</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>$\bar{X}$</td>
<td>m</td>
<td>Min</td>
<td>Max</td>
<td>s</td>
<td>N</td>
<td>$\bar{X}$</td>
<td>m</td>
<td>Min</td>
<td>Max</td>
<td>s</td>
</tr>
<tr>
<td>Und_Time</td>
<td>22</td>
<td>48.18</td>
<td>50.50</td>
<td>13</td>
<td>69</td>
<td>14.12</td>
<td>19</td>
<td>46.74</td>
<td>45.00</td>
<td>25</td>
<td>75</td>
<td>9.61</td>
</tr>
<tr>
<td>Und_Comp</td>
<td>33</td>
<td>5.64</td>
<td>6.00</td>
<td>3</td>
<td>7</td>
<td>1.03</td>
<td>32</td>
<td>4.31</td>
<td>4.00</td>
<td>2</td>
<td>7</td>
<td>1.38</td>
</tr>
<tr>
<td>Mod_Time</td>
<td>21</td>
<td>29.71</td>
<td>30.00</td>
<td>9</td>
<td>65</td>
<td>13.41</td>
<td>19</td>
<td>30.21</td>
<td>30.00</td>
<td>10</td>
<td>50</td>
<td>9.74</td>
</tr>
<tr>
<td>Mod_Comp</td>
<td>33</td>
<td>0.66</td>
<td>0.77</td>
<td>0</td>
<td>1.00</td>
<td>0.27</td>
<td>32</td>
<td>0.45</td>
<td>0.48</td>
<td>0.10</td>
<td>0.76</td>
<td>0.22</td>
</tr>
<tr>
<td>Mod_Corr</td>
<td>31</td>
<td>0.95</td>
<td>1.00</td>
<td>0</td>
<td>1.00</td>
<td>0.18</td>
<td>32</td>
<td>0.93</td>
<td>1.00</td>
<td>0.50</td>
<td>1.00</td>
<td>0.12</td>
</tr>
<tr>
<td>Mod_Rate</td>
<td>21</td>
<td>0.65</td>
<td>0.62</td>
<td>0</td>
<td>1.69</td>
<td>0.40</td>
<td>19</td>
<td>0.37</td>
<td>0.27</td>
<td>0.08</td>
<td>1.40</td>
<td>0.31</td>
</tr>
</tbody>
</table>
Figure 2 and 3
Figure 4 and 5

Fig. 4. Number of participants who found change places for the good design.

Fig. 5. Number of participants who found change places for the bad design.
Analysis

A1: Specify any procedures used to control for multiple testing.

A2: Consider using blind analysis.

A3: Perform sensitivity analyses.
Improvements

- Correctness of impact analysis
- More debriefing
- More subjects
- Cleaner, more standard documentation
- More comparable tasks
Presentation

P1: Describe or cite a reference for all statistical procedures used.

P2: Report the statistical package used.

P3: Present quantitative results as well as significance levels. Quantitative results should show the magnitude of effects and the confidence limits.

P4: Present the raw data whenever possible. Otherwise, confirm that they are available for confidential review by the reviewers and independent auditors.

P5: Provide appropriate descriptive statistics.

P6: Make appropriate use of graphics.
Interpretation

I1: Define the population to which inferential statistics and predictive models apply.

I2: Differentiate between statistical significance and practical importance.

I3: Define the type of study.

I4: Specify any limitations of the study.
Evaluating Experiment

- Claim

- Convincing