Checklist for Assessing the Quality of Papers Reporting Controlled Experiments in Software Engineering Research

Group ID: ___________________________

Paper Title: _____________________________________________________________

<table>
<thead>
<tr>
<th>#</th>
<th>Question</th>
<th>How to answer the question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category: Questions on Aims</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| 1. | Do the authors clearly state the aims of the research? | **Consider:**  
• Do the authors state research questions, e.g., related to time-to-market, cost, product quality, process quality, developer productivity, developer skills?  
• Do the authors state hypotheses and their underlying theories? | To full extent (fully)  
To a large extent (mostly)  
To a small extent (somewhat)  
Not at all |
| Category: Questions on Design, Data Collection, and Data Analysis | | | |
| 2. | Do the authors describe the sample and experimental units (=experimental materials and participants as individuals or teams)? | **Consider:**  
• Do the authors explain how experimental units were defined, selected?  
• Do the authors explain to what degree the experimental units are representative?  
• Do the authors explain why the experimental units they selected were the most appropriate to provide insight into the type of knowledge sought by the experiment?  
• Do the authors report the sample size? | To full extent (fully)  
To a large extent (mostly)  
To a small extent (somewhat)  
Not at all |
| 3. | Do the authors describe the design of the experiment? | **Consider:**  
• Do the authors clearly describe the chosen design (blocking, within or between subject design, do treatments have levels)?  
• Do the authors define/describe all treatments and all controls? | To full extent (fully)  
To a large extent (mostly)  
To a small extent (somewhat)  
Not at all |
| 4. | Do the authors describe the data collection procedures and define the measures? | **Consider:**  
• Are all measures clearly defined (e.g., scale, unit, counting rules)?  
• Is the form of the data clear (e.g., tape recording, video material, notes etc.)?  
• Are quality control methods used to ensure consistency, completeness and accuracy of collected data?  
• Do the authors report drop-outs? | To full extent (fully)  
To a large extent (mostly)  
To a small extent (somewhat)  
Not at all |
| 5. | Do the authors define the data analysis procedures? | **Consider:**  
• Do authors justify their choice / describe the procedures / provide references to descriptions of the procedures?  
• Do the authors report significance levels and effect sizes?  
• If outliers are mentioned and excluded from the analysis, is this justified?  
• Do the authors report or give reference to raw data and/or descriptive statistics? | To full extent (fully)  
To a large extent (mostly)  
To a small extent (somewhat)  
Not at all |

(continues on next page)
| 6. | Do the authors discuss potential experimenter bias? | **Consider:**  
- Were the authors the developers of some or all of the treatments? If yes, do the authors discuss the implications anywhere in the paper? (if the answer to the second question is ‘no’, then check “not at all”)  
- Was there random allocation to treatment?  
- Was training and conduct equivalent for all treatment groups?  
- Was there allocation concealment (i.e. the researchers did not know what treatment each subject was assigned)? | To full extent (fully)  
Not at all |

| 7. | Do the authors discuss the limitations of their study? | **Consider:**  
- Do the authors discuss external validity with respect to subjects, materials and tasks?  
- Do the authors discuss internal, construct, and conclusion validity?  
- If the study used novel measures, is the construct validity of the measures discussed? | To full extent (fully)  
Not at all |

Category: Questions on Study Outcome

| 8. | Do the authors clearly state the findings? | **Consider:**  
- Do the authors present results clearly?  
- Do the authors present conclusions clearly?  
- Are the conclusions clearly linked to the results?  
- Do the authors discuss their conclusions in relation to the original research questions?  
- Are limitations of the study discussed explicitly? | To full extent (fully)  
Not at all |

| 9. | Is there evidence that the experiment can be used by other researchers / practitioners? | **Consider:**  
- Do the authors discuss whether or how the findings can be transferred to other populations, or consider other ways in which the research can be used?  
- To what extent do authors interpret results in the context of other studies / existing body of knowledge? | To full extent (fully)  
Not at all |

**Scale:**  
- **To full extent (fully)**: all questions listed in the "consider" column can be answered with "yes"  
- **To a large extent (mostly)**: the majority of all (but not all) questions listed in the "consider" column can be answered with "yes"  
- **To a small extent (somewhat)**: some (but the minority) of the questions listed in the "consider" column can be answered with "yes"  
- **Not at all**: none of the questions listed in the "consider" column can be answered with "yes"

**Important:**  
For each item, take notes about the underlying rationale for your score, i.e., why did you score ‘fully’ or ‘mostly’ or ‘somewhat’ or ‘not at all’.