

Experiment Phases (Wohlin) / Approaches	Experiment Scoping (Goal definition)	Experiment Planning (Context selection, Hypothesis, Variables, Selection of subjects, design type, instrumentation and validity evaluation)	Experiment Operation (Preparation, Execution and Data validation)	Analysis & Interpretation (Descriptive statistics, Data set reduction and Hypothesis testing)
Checklist - Kitchenham e Charters (2007)	1 Are the aims clearly stated?	<p>2 Do the study measures allow the questions to be answered?</p> <p>3 Is the sample representative of the population to which the results will generalise?</p> <p>4 Were treatments randomly allocated?</p> <p>5 Is there a comparison or control group?</p> <p>6 If there is a control group, are participants similar to the treatment group participants in terms of variables that may affect study outcomes?</p> <p>7 Was the sample size justified?</p> <p>8 If the study involves assessment of a technology, is the technology clearly defined?</p> <p>9 Could the choice of subjects influence the size of the treatment effect?</p> <p>11 Are the variables used in the study adequately measured (i.e. are the variables likely to be valid and reliable)?</p> <p>12 Are the measures used in the study fully defined?</p> <p>13 Are the measures used in the study the most relevant ones for answering the research questions?</p> <p>14 Is the scope (size and length) of the study sufficient to allow for changes in the outcomes of interest to be identified?</p> <p>21 What was the response rate?</p> <p>22 Was the denominator (i.e. the population size) reported?</p> <p>23 Do the researchers explain the data types (continuous, ordinal, categorical)?</p> <p>24 Are the study participants or observational units adequately described? For example, SE experience, type (student, practitioner, consultant), nationality, task experience and other relevant variables.</p> <p>52 Do the researchers explain the consequences of any problems with the validity/reliability of their measures?</p>	<p>10 Could lack of blinding introduce bias?</p> <p>15 Did untoward events occur during the study?</p> <p>16 Was outcome assessment blind to treatment group?</p> <p>17 Are the data collection methods adequately described?</p> <p>18 If two groups are being compared, were they treated similarly within the study?</p> <p>19 If the study involves participants over time, what proportion of people who enrolled at the beginning dropped out?</p> <p>20 How was the randomisation carried out?</p>	<p>25 Were the basic data adequately described?</p> <p>26 Have "drop outs" introduced bias?</p> <p>27 Are reasons given for refusal to participate?</p> <p>28 Are the statistical methods described?</p> <p>29 Is the statistical program used to analyse the data referenced?</p> <p>30 Are the statistical methods justified?</p> <p>31 Is the purpose of the analysis clear?</p> <p>32 Are scoring systems described?</p> <p>33 Are potential confounders adequately controlled for in the analysis?</p> <p>34 Do the numbers add up across different tables and subgroups?</p> <p>35 If different groups were different at the start of the study or treated differently during the study, was any attempt made to control for these differences, either statistically or by matching?</p> <p>36 If yes, was it successful?</p> <p>37 Was statistical significance assessed?</p> <p>38 If statistical tests are used to determine differences, is the actual p value given?</p> <p>39 If the study is concerned with differences among groups, are confidence limits given describing the magnitude of any observed differences?</p> <p>40 Is there evidence of multiple statistical testing or large numbers of post hoc analysis?</p> <p>41 How could selection bias arise?</p> <p>42 Are all study questions answered?</p> <p>43 What do the main findings mean?</p> <p>44 Are negative findings presented?</p> <p>45 If statistical tests are used to determine differences, is practical significance discussed?</p> <p>46 If drop outs differ from participants, are limitations to the results discussed?</p> <p>47 How are null findings interpreted? (I.e. has the possibility that the sample size is too small been considered?)</p> <p>48 Are important effects overlooked?</p> <p>49 How do results compare with previous reports?</p> <p>50 How do the results add to the literature?</p> <p>51 What implications does the report have for practice?</p>
Checklist - Kampenes (2007)		<p>1.1 Sample Size;</p> <p>1.3 Type (Student/Professionals);</p> <p>1.4 Recruitment (Voluntarily/Mandatory);</p> <p>1.5 Some kind of background information;</p> <p>1.5.1 Programming experience;</p> <p>1.5.2 Work experience;</p> <p>1.5.3 Task related experience;</p> <p>1.5.4 Grades;</p> <p>2.1 Task;</p> <p>2.2 Duration;</p> <p>2.4 Size of materials;</p> <p>2.5 Location;</p> <p>2.6 The use of tools;</p> <p>3.1 Well-defined population;</p> <p>3.5 Assignment procedure (randomized or quasi);</p> <p>3.6 Randomization method;</p> <p>4.1 Discussion of internal validity;</p> <p>4.2 Threats of internal validity;</p> <p>4.3 Discussion of external validity;</p> <p>4.4 Discussion of statistical conclusion validity;</p> <p>4.5 Discussion of construct validity;</p>	2.3 Application system;	<p>1.2 Mortality rate;</p> <p>3.2 Statistical power;</p> <p>3.3 Effect size;</p> <p>3.4 Information available for estimation of at least one effect size;</p>

Experiment Phases (Wohlin) / Approaches	Experiment Scoping (Goal definition)	Experiment Planning (Context selection, Hypothesis, Variables, Selection of subjects, design type, instrumentation and validity evaluation)	Experiment Operation (Preparation, Execution and Data validation)	Analysis & Interpretation (Descriptive statistics, Data set reduction and Hypothesis testing)
Checklist - Kitchenham et al. (2010)	<p>1 Do the authors clearly state the aims of the research?</p> <p>1.1 Do the authors state research questions, e.g., related to time-to-market, cost, product quality, process quality, developer productivity, and developer skills?</p>	<p>1.2 Do the authors state hypotheses and their underlying theories?</p> <p>2 Do the authors describe the sample and experimental units (=experimental materials and participants as individuals or teams)?</p> <p>2.1 Do the authors explain how experimental units were defined and selected?</p> <p>2.2 Do the authors state to what degree the experimental units are representative?</p> <p>2.3 Do the authors explain why the experimental units they selected were the most appropriate for providing insight into the type of knowledge sought by the experiment?</p> <p>2.4 Do the authors report the sample size?</p> <p>3 Do the authors describe the design of the experiment?</p> <p>3.1 Do the authors clearly describe the chosen design (blocking, within or between subject design, do treatments have levels)?</p> <p>3.2 Do the authors define/describe all treatments and all controls?</p> <p>4 Do the authors describe the data collection procedures and define the measures?</p> <p>4.1 Are all measures clearly defined (e.g., scale, unit, counting rules)?</p> <p>6.1 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 authors developed the treatments (or parts of them) without discussing the implications, the answer to question 6 is "not at all".)</p> <p>6.2 Was there random allocation to treatments?</p> <p>7 Do the authors discuss the limitations of their study?</p> <p>7.1 Do the authors discuss external validity with respect to subjects, materials, and tasks?</p> <p>7.2 If the study was a quasi-experiment, do the authors discuss the design components that were used to address any study weakness?</p> <p>7.3 If the study used novel measures, is the construct validity of the measures discussed?</p>	<p>4.2 Is the form of the data clear (e.g., tape recording, video material, notes, etc.)?</p> <p>4.3 Are quality control methods used to ensure consistency, completeness and accuracy of collected data?</p> <p>4.4 Do the authors report drop-outs?</p> <p>6 Do the authors discuss potential experimenter bias?</p> <p>6.3 Was training and conduct equivalent for all treatment groups?</p> <p>6.4 Was there allocation concealment, i.e., did the researchers know to what treatment each subject was assigned?</p>	<p>5 Do the authors define the data analysis procedures?</p> <p>5.1 Do the authors justify their choice/describe the procedures/provide references to descriptions of the procedures?</p> <p>5.2 Do the authors report significance levels and effect sizes?</p> <p>5.3 If outliers are mentioned and excluded from the analysis, is this justified?</p> <p>5.4 Do the authors report or give references to raw data and/or descriptive statistics?</p> <p>8 Do the authors state the findings clearly?</p> <p>8.1 Do the authors present results clearly?</p> <p>8.2 Do the authors present conclusions clearly?</p> <p>8.3 Are the conclusions warranted by the results and are the connections between the results and conclusions presented clearly?</p> <p>8.4 Do the authors discuss their conclusions in relation to the original research questions?</p> <p>8.5 Are limitations of the study discussed explicitly?</p> <p>9 Is there evidence that the E/QE can be used by other researchers/practitioners?</p> <p>9.1 Do the authors discuss whether or how the findings can be transferred to other population, or consider other ways in which the research can be used?</p> <p>9.2 To what extent do authors interpret results in the context of other studies/the existing body of knowledge?</p>
Scale - Dieste et al. (2011)	<p>1 Does the introduction contain the industrial context (entities, attributes, and measures) and description of the techniques to be reviewed? For experiments that evaluate techniques developed in industry;</p> <p>2 Does the report summarize and discuss earlier similar experiments that have been conducted?</p>	<p>3 Are the hypotheses being laid and are they synonymous with the goal discussed before in introduction?</p> <p>4 Does the researcher define the population from which objects and subjects are drawn?</p> <p>5 Does the researcher define the process from which the objects and subjects are selected (e.g. random sampling)?</p> <p>6 Does the researcher define the process by which he applies the treatment to objects and subjects (e.g. randomization)?</p> <p>7 Was randomization used for selecting the population and applying the treatment?</p> <p>8 Is mention made of the threats to validity and also how these threats affect the results and findings?</p> <p>10 Is an appropriate blinding procedure used (e.g. blind allocation of materials, blind marking)?</p>		<p>9 Are the statistical significances mentioned with the results?</p> <p>10 Is an appropriate blinding procedure used (e.g. blind allocation of materials, blind marking)?</p>