Methodology: Experiment

	Exemplary 4	Proficient 3	Marginal 2	Unacceptable 1
Research question ("Is the research question interesting, significant, well-motivated?")	The research question is significant, interesting and well-motivated. It is of interest to the HCI community and makes noteworthy contributions extending prior work.	One aspect of the research question could be improved. For instance, the question is significant and interesting but not well-motivated.	Two aspects of the research question could be improved. For instance, the question is interesting but not significant nor well-motivated.	The question is neither interesting, significant, nor well-motivated. (Or no research question.)
Hypothesis ("Are the hypotheses testable, concise and constructed from prior work?")	Hypotheses are justified, testable, concise, name key constructs and constructed from prior work. OR: No hypothesis but a "hunting phenomena" approach (e.g., exploratory study) is justified.	One aspect of the hypotheses could be improved. For instance, hypotheses are not concise enough.	Multiple aspects of the hypotheses could be improved. For instance, hypotheses are not concise and not constructed from prior work.	The hypotheses are not well formulated (e.g., bad hypotheses blind us to interesting findings). Or no hypotheses formulated without justification.
Experimental Design - I ("Does it employ the right experimental design?")	The research employed the most appropriate experimental design (e.g., between-subject vs within-subject, controlled vs "in the wild", participant selection & assignment, no confounding effects, order effects, strong baseline...) for the research question.	The methodology is not the best one for addressing the research question but a realistic and still valid one given the constraints (e.g., time, number of participants).	Multiple minor aspects of the design could be improved and might impact the results but it's acknowledged and potential explanations are given. The finding, albeit imperfect, is interesting and makes some level of contribution.	The research used the wrong methodology for the question or does not control (potential) confounding factors that make the study invalid.
Experimental Design - II ("Does it recruit the right participants, uses the right tasks and setting?")	Task, setting and participants justified and appropriate for the study.	One aspect could be improved. For instance, the tasks could better reflect real-world tasks.	Multiple aspects could be improved. For instance, the recruited participants already have experience (that can impact the result) and the tasks are not realistic enough.	The study does not recruit representative participants, the setting or tasks are unrealistic. As a result, the result is not generalizable.

Analysis ("Does it use the right analytical methods?")	Uses a proper analytical method to provide strong, convincing evidence to back-up the analysis/conclusions. When appropriate, they add justification for their choice.	Strong evidence and appropriate analysis. The justification was necessary but not provided.	Evidence is not as strong because of unconvincing justification or a different analytical method would be better.	Uses incorrect analytical method, resulting in unconvincing analysis/conclusion.
Articulation ("Are the claims reported with enough details?")	The claims are articulated with details (e.g., with numbers, X was 7.2% faster than Y vs. X was faster than Y).	Some claims are not presented clearly with evidence (e.g., a statement is not presented along with numbers to support it).	While central claims are articulated with enough detail, most claims are not presented with enough details.	The paper does not articulate enough detail about its claims that they cannot be trusted.
Reproducibility ("Is enough information available for replication?")	The paper details enough information for others to be able to reproduce the exact experiment and analysis.	A few minor details are missing but they can be inferred in order to reproduce the experiment.	The experiment can be replicated but missing details can make it difficult for replicated studies to be properly compared with the original study.	The paper is missing important details for the study to be replicated.
Interpretation ("Does it correctly portray the findings?")	The paper accurately reports the analysis (e.g., covers most, if not all, potential interpretations of the presented analysis, reports limitations for future work, without misinterpretations). For instance, it does not magnify nor downsize the size or effect.	The paper reports the main analysis correctly but provides one or two potentially misleading analyses (e.g., equating low p-value with high effect size), or a few interpretations and limitations are not reported with a sufficient level of details.	The paper is missing a number of significant interpretations of the data and contains misleading analyses.	The paper misinterpreted the findings and reported wrong analyses for their research questions.