Methodology: Experiment

	Exemplary 4	Proficient 3	Marginal 2	Unacceptable 1

Research Question [2, 3]	• **Testable**
	• **3 of 3:**
	  ○ **Clear:** implication of results is well-defined
	  ○ **Significant:** contributes to scientific knowledge
	  ○ **Interesting:** The authors' contributions are novel and insightful
		• **Testable**
		• **2 of 3:**
		  ○ **Clear:** implication of results is well-defined
		  ○ **Significant:** contributes to scientific knowledge
		  ○ **Interesting:** The authors' contributions are novel and insightful
			• **Testable**
			• **1 of 3:**
			  ○ **Clear:** implication of results is well-defined
			  ○ **Significant:** contributes to scientific knowledge
			  ○ **Interesting:** The authors' contributions are novel and insightful
				• **Not testable**
				• **Not clear:** implication of results is not well-defined
				• **Not significant:** trivial results that do not add to theory or practice
				• **Not interesting:** contributions are not novel or results are predictable

Independent Variable Level Selection [3] **Terminology:** • ***Levels of an independent variable*:** the conditions experienced by the participants • ***Constructs*:** the key ideas of the research question	**5 of 5:** • **Representative:** The conditions match the constructs • **Complete:** More than two condition for each construct • **Strong baseline:** state-of-the-art baseline • **No bias from non-essential choices:** No effect or bias from UI features irrelevant to the research question • **Comparable:** the different conditions only differ in aspects under study	**4-3 of 5:** • **Representative:** The conditions match the constructs • **Complete:** More than two condition for each construct • **Strong baseline:** state-of-the-art baseline • **No bias from non-essential choices:** No effect or bias from UI features irrelevant to the research question • **Comparable:** the different conditions only differ in aspects under study	**2-1 of 5:** • **Representative:** The conditions match the constructs • **Complete:** More than two condition for each construct • **Strong baseline:** state-of-the-art baseline • **No bias from non-essential choices:** No effect or bias from UI features irrelevant to the research question • **Comparable:** the different conditions only differ in aspects under study	• **Not representative:** The conditions do not match the constructs • **Incomplete:** Only one condition used for each construct • **Weak baseline:** outdated or inferior baseline • **Bias from non-essential choices:** features irrelevant to the research question affect results • **Not comparable:** the different conditions differ in more aspects than the ones under study

Experimental Design and Procedure [1, 2, 3] **Terminology:** ● ***Experiment Design*:** ○ within subjects ○ between subjects ○ mixed factorial ● ***Bigger investigations*:** Building up to a bigger series of experiments that probes the phenomenon of interest more deeply	● **Experiment Design:** Reviewed and selected carefully ● **Formal procedure** (2 of 2): ○ Consistent experiment ○ Replicable experiment ● **Confounding variables** (3 of 3): Minimise by: ○ Controlling the order in which we test the interfaces ○ Devise different and well-defined tasks ○ Controlling context ● **Robust experiment** (3 of 3): ○ Careful design of instructions ○ Piloting ○ Careful collection and management of data ● **Bigger investigations**	● **Experiment Design:** Justified ● **Formal procedure** (2 of 2): ○ Consistent experiment ○ Replicable experiment ● **Confounding variables** (2 of 3): Minimise by: ○ Controlling the order in which we test the interfaces ○ Devise different and well-defined tasks ○ Controlling context ● **Robust experiment** (2 of 3): ○ Careful design of instructions ○ Piloting ○ Careful collection and management of data	● **Experiment Design:** Justified ● **Formal procedure** (1 of 2): ○ Consistent experiment ○ Replicable experiment ● **Confounding variables** (1 of 3): Minimise by: ○ Controlling the order in which we test the interfaces ○ Devise different and well-defined tasks ○ Controlling context ● **Robust experiment** (1 of 3): ○ Careful design of instructions ○ Piloting ○ Careful collection and management of data	● **Experiment Design:** <u>Not</u> justified ● **No formal procedure** ● **Confounding variables:** <u>No</u> minimisation ● **Robust experiment**: <u>No</u>
Participant selection [1, 3]	**4 of 4:** ● **No confounding:** No confounding as a result of participant choice ● **Generalizable:** Participant choice leads to generalizability of results ● **Appropriate number:** An appropriate number of participants chosen (power analysis) ● **Ethical:** All ethical considerations taken	**3 of 4:** ● **No confounding:** No confounding as a result of participant choice ● **Generalizable:** Participant choice leads to generalizability of results ● **Appropriate number:** An appropriate number of participants chosen (power analysis)	**2-1 of 4:** ● **No confounding:** No confounding as a result of participant choice ● **Generalizable:** Participant choice leads to generalizability of results ● **Appropriate number:** An appropriate number of participants chosen (power analysis) ● **Ethical:** All ethical considerations taken	● **Confounding:** Participant choice has a major impact on the integrity of the research and introduces confounding ● **Not generalizable:** Participant choice makes results ungeneralizable ● **Inappropriate number:** Number of participants chosen arbitrarily or randomly

	into account (VIP) • **Constraints discussion:** Any constraints on participant choice are discussed in detail	• **Ethical:** All ethical considerations taken into account (VIP) • **Constraints discussion:** Any constraints on participant choice are discussed in detail	into account (VIP) • **Constraints discussion:** Any constraints on participant choice are discussed in detail	• **Unethical:** No ethical considerations taken into account • **No constraints discussion:** Any constraints on participant choice are ambiguous or not addressed
Dependent Variable Selection [2, 3] **Example:** • ***Construct*:** usability of a user interface • ***Measurement dimensions*:** time to learn, subjective satisfaction, etc.	**3 of 3:** • **Well-defined:** The constructs the researcher is interested in measuring are clearly defined • **Representative:** Actual collected measurements reflect the constructs of interest • **No mono-operationalization bias:** If necessary, several dimensions of a construct are measured	**2 of 3:** • **Well-defined:** The constructs the researcher is interested in measuring are clearly defined • **Representative:** Actual collected measurements reflect the constructs of interest • **No mono-operationalization bias:** If necessary, several dimensions of a construct are measured	**1 of 3:** • **Well-defined:** The constructs the researcher is interested in measuring are clearly defined • **Representative:** Actual collected measurements reflect the constructs of interest • **No mono-operationalization bias:** If necessary, several dimensions of a construct are measured	• **Ill-defined:** The constructs the researcher is interested in measuring are not defined or left ambiguous • **Not representative:** Actual collected measurements do not reflect the constructs of interest • **Mono-operationalization bias:** Only one dimension of a multi-dimensional construct is measured
Results and Analysis [2]	**3 of 3:** • **Credibility:** authors demonstrate competence in the collection and analysis of results • **Relevance:** results are used to draw a conclusion about the research question • **Generality:** implication of the results on future or out-of-scope work is	**2 of 3:** • **Credibility:** authors demonstrate competence in the collection and analysis of results • **Relevance:** results are used to draw a conclusion about the research question • **Generality:** implication of the results on future or out-of-scope work is	**1 of 3:** • **Credibility:** authors demonstrate competence in the collection and analysis of results • **Relevance:** results are used to draw a conclusion about the research question • **Generality:** implication of the results on future or out-of-scope work is	• **No Credibility:** authors are unconvincing in their competence collecting and analyzing results • **No Relevance:** results are not related to the research question • **No Generality:** results have no implication on out-of-scope work. Potential future work left undiscussed

	discussed	discussed	discussed	

References

[1] Blandford, A., Cox, A. L. & Cairns, P. A. (2008) Controlled Experiments. In Cairns, P.A., & Cox, A.L. (eds.) *Research Methods for Human Computer Interaction*. CUP. 1-16.

[2] Gergle, D., & Tan, D. S. (2014). Experimental research in HCI. In *Ways of Knowing in HCI* (pp. 191-227). Springer, New York, NY.

[3] Hornbæk, K. (2013). Some whys and hows of experiments in human–computer interaction. *Foundations and Trends in Human-Computer Interaction*, *5*(4), 299-373.