

Fulfillment of SIGSOFT Empirical Standard for Controlled Experiments (with Human Participants)

Essential Attributes

- states formal hypotheses and
 - ✓ See Study protocol - points 4&5
- uses two-sided hypotheses
 - ✓ See Study protocol - point 5
- describes the dependent and independent variable(s) and justifies how they are measured (including units, instruments)
 - ✓ See Variables under Section 3.3; see Study protocol - Point 6
- describes extraneous variables and how they are controlled or not
 - ✓ See Study protocol - point 6
- describes how characteristics of the phenomenon under investigation relate to experimental constructs
 - ✓ See Section 3.3, see Study protocol - point 7
- describes the research design and protocol, including treatments, materials, tasks, design (e.g., 2x2 factorial), participant allocation, period and sequences (for crossover designs), and logistics
 - ✓ See Experiment design under Section 3.3, and See study protocol - Points 7, 8, 9
- uses random assignment and explains logistics (e.g. how random numbers were generated)
 - ✓ See the Study Protocol - point 8.
- describes experimental objects (e.g., real or toy system) and their characteristics (e.g., size, type)
 - ✓ See Tools Selection under Section 3.3
- justifies selection of experimental objects; acknowledges object-treatment confounds.
 - ✓ See Tools Selection under Section 3.3
- design and protocol appropriate (not optimal) for stated research questions and hypotheses
 - ✓ See study protocol document
- describes participants (e.g. education, relevant experience or preferences)
 - ✓ See Study protocol - point 10 and Section 3.4 in the paper.

- reports distribution-appropriate descriptive and inferential statistics; justifies tests used
 - ✓ See study protocol - point 13., Section 3.5 of the paper.
- reports effects sizes with confidence intervals (if using frequentist approach)
- discusses construct, conclusion, internal, and external validity
 - ✓ See Section 6 of the paper for external/internal/construct/conclusion validity.
- discusses alternative interpretations of results

Desirable Attributes

- provides supplementary material such as complete, algorithmic research protocol; task materials; raw, de-identified dataset; analysis scripts
 - ✓ See online appendix for task materials, and study protocol
- named experiment design (e.g. simple 2-group, 2x2 factorial, randomized block)
 - ✓ We employed a cross-over design
- reports manipulation checks
 - See study protocol - point 12.
- clearly distinguishes evidence-based results from interpretations and speculation³
 - ✓ We evaluate whether the error and completion rate differences are significant using Kruskal-Wallis/Bonferroni analysis. See Section 4.4.