

Is experimental philosophy ready for meta-analyses?

A quantitative approach to assessing the quality of methodological and statistical reporting in x-phi

Bartosz Maćkiewicz, Katarzyna Kuś, Joanna Gęgotek, Monika Gurak, Franciszek Kittel (Faculty of Philosophy, University of Warsaw)

Why Does Good Reporting Matter?

- It is the bedrock of **cumulative science**, allowing us to move from isolated findings to robust knowledge.
- It enables **meta-analyses** to verify effects, estimate their true size, and detect publication bias.
- Without it, direct **replication** and **data reuse** become impossible, hindering scientific efficiency.
- It is a prerequisite for the audits that allow a scientific field to **assess its own quality and standards**.

Our Aims: While other fields routinely audit their reporting quality, experimental philosophy has a blind spot, with only limited prior work. Our study is a large-scale attempt to assess our field's reporting.

Methods & Sample

Tool: We built the X-Phi Transparency and Quality of Reporting Inventory (X-PHI TQRI), a tool with ~100 questions to systematically assess reporting across several domains: Transparency, Theoretical Background, Methods, Results, and Conclusions.

Sample: We drew a sample of 400 papers from the PhilPapers database (2007-2024). The sample was stratified across six 3-year bins to ensure representation over time.

Process: Coders were trained and calibrated on a common set of papers. To ensure reliability, 5% of papers were double-coded, with disagreements resolved by consensus. Our goal was an inter-coder agreement of >70%.

Our Assessment Tool: The X-PHI TQRI

X-Phi Transparency and Quality of Reporting Inventory (X-PHI TQRI) is designed to enable systematic tracking of reporting practices over time, comparison across journals, and identification of strengths and blind spots.

Transparency

Checks for open access, pre-registration, and the availability of data, materials, and code.

Theoretical Background

Assesses the clarity of the research problem, grounding in existing literature, and formulation of hypotheses.

Methods (general)

Covers reporting of study aims, theoretical grounding measures, participant recruitment, and analytic approach.

Methods (questionnaire)

Covers reporting of stimuli and probes, survey tools, response scales, instructions, timing, and group assignment.

Results

Evaluates the completeness of descriptive statistics, inferential statistics, and effect sizes.

Conclusions

Gauges how well the results are interpreted in light of the hypotheses and whether their limitations are discussed.

Results: Reporting Quality is Improving, but Progress is Uneven (2007-2018)



Takeaways & Limitations

The Good News: Reporting of concrete Methods and Results is getting significantly better. The field has made measurable progress in describing procedures and inferential statistics.

The Bad News: Key details for replication are still missing. Critical information like the exact wording of study questions is not reported any more frequently, and descriptive statistics are often inadequate. Progress is also uneven; for example, reporting on participant education has declined.

Limitations: The work for 2019-2024 is still in progress. The boundaries of X-Phi are blurry, and some coding items are inherently subjective despite reliability protocols. The scoring system is based on our intuitions about best practices.

