

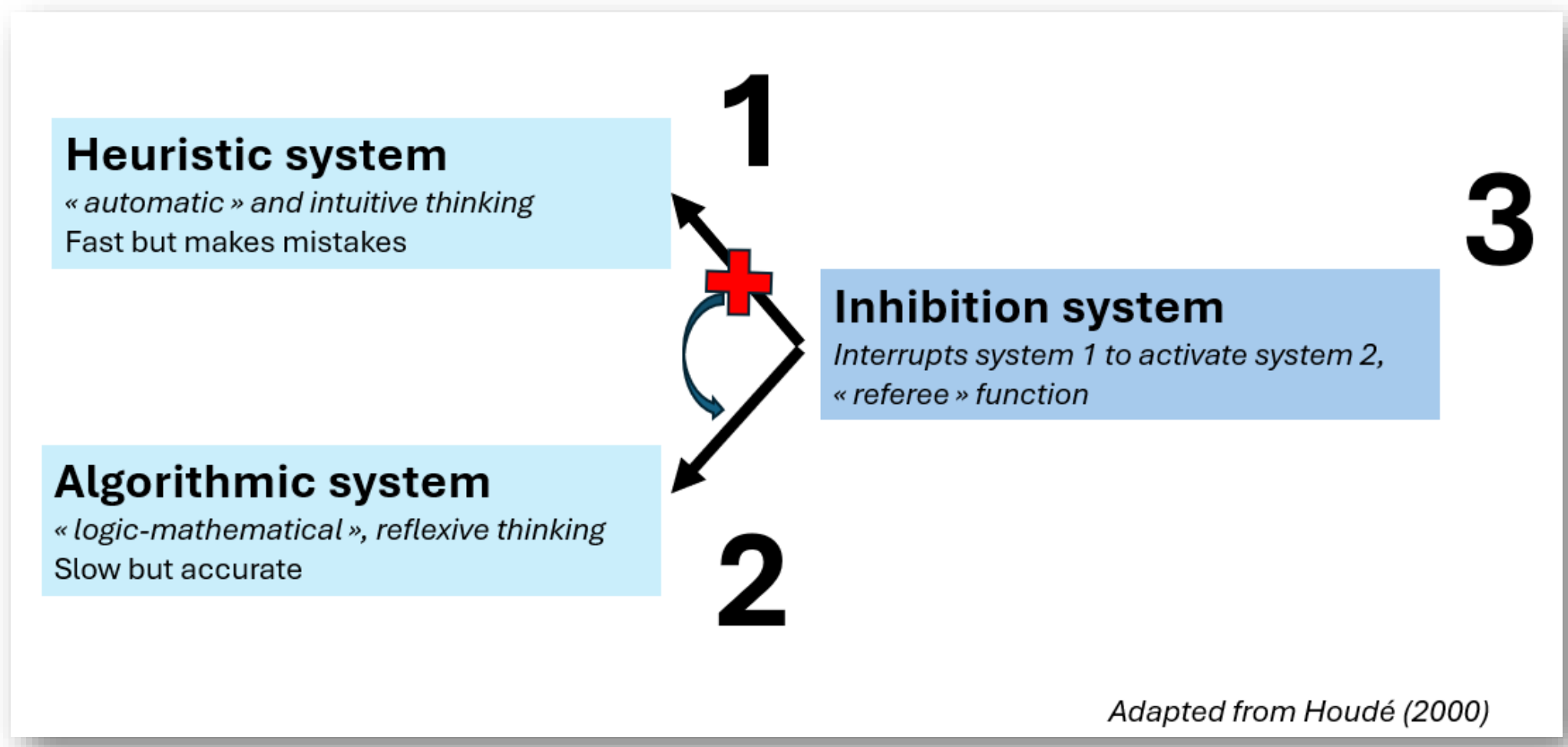
"Think of the Wrong Answer": Extending Inhibition Effects in Chemistry Multiple-Choice Questions

Vincent Natalis^{1,2}

Abstract

This study extends Talanquer's (2017) finding that prompting students to identify the intuitively appealing but incorrect answer—alongside their actual response—improves performance on chemistry concept inventories. Grounded in dual-process theory (System 1/System 2 reasoning) and cognitive inhibition research, the project replicates this effect among French-speaking Belgian students, addressing a gap highlighted by Ott et al. (2025), whose attempt to replicate the effect in physics (Force Concept Inventory) failed, suggesting domain-dependence. Three research questions guide the work: (1) whether the specific "predict the wrong answer" prompt outperforms a general awareness-raising instruction ("this is a trick question"); (2) whether the effect generalizes to conventional exam-style multiple-choice questions, beyond the original Intuitive Concept Inventory (ICI) format; and (3) whether the prompt's effectiveness diminishes across successive questions due to fatigue or habituation. The ICI was translated into French and validated by chemistry education experts. Preliminary data collection (April–May 2025) will involve ~100 first-year biology undergraduates and 50 secondary school students (ages 15–17), randomly assigned to control, wrong-answer-prediction, or general-awareness conditions using a double-questionnaire design. Later phases will expand to larger and more diverse cohorts across educational levels. The author acknowledges that, since data collection was incomplete at submission, evaluation must rest on the theoretical and methodological framing rather than results.

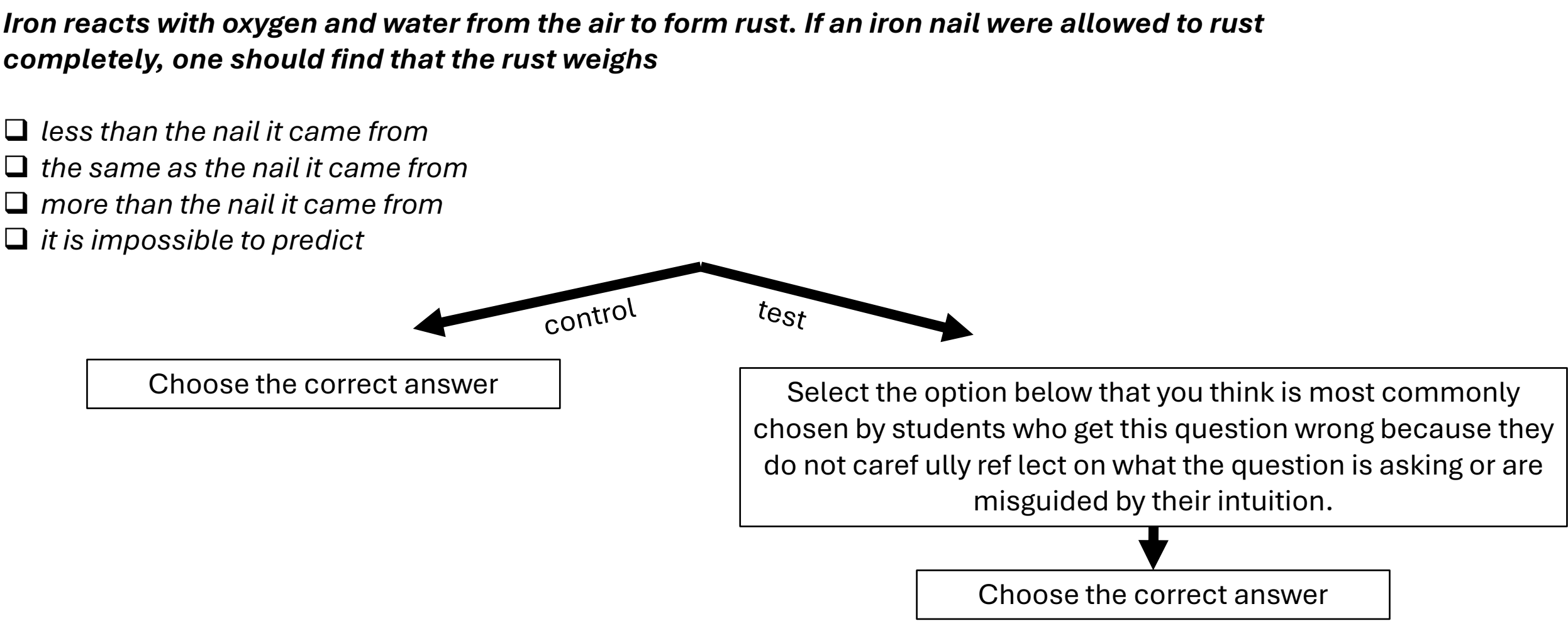
Context



Objectives

- 1) Replicate Talanquer's findings given Ott's non-conclusive results
- 2) Test new hypotheses
 - I. Broad vs. precise meta instructions
 - II. Conceptual vs. classical questions
 - III. Test the effect of fatigue

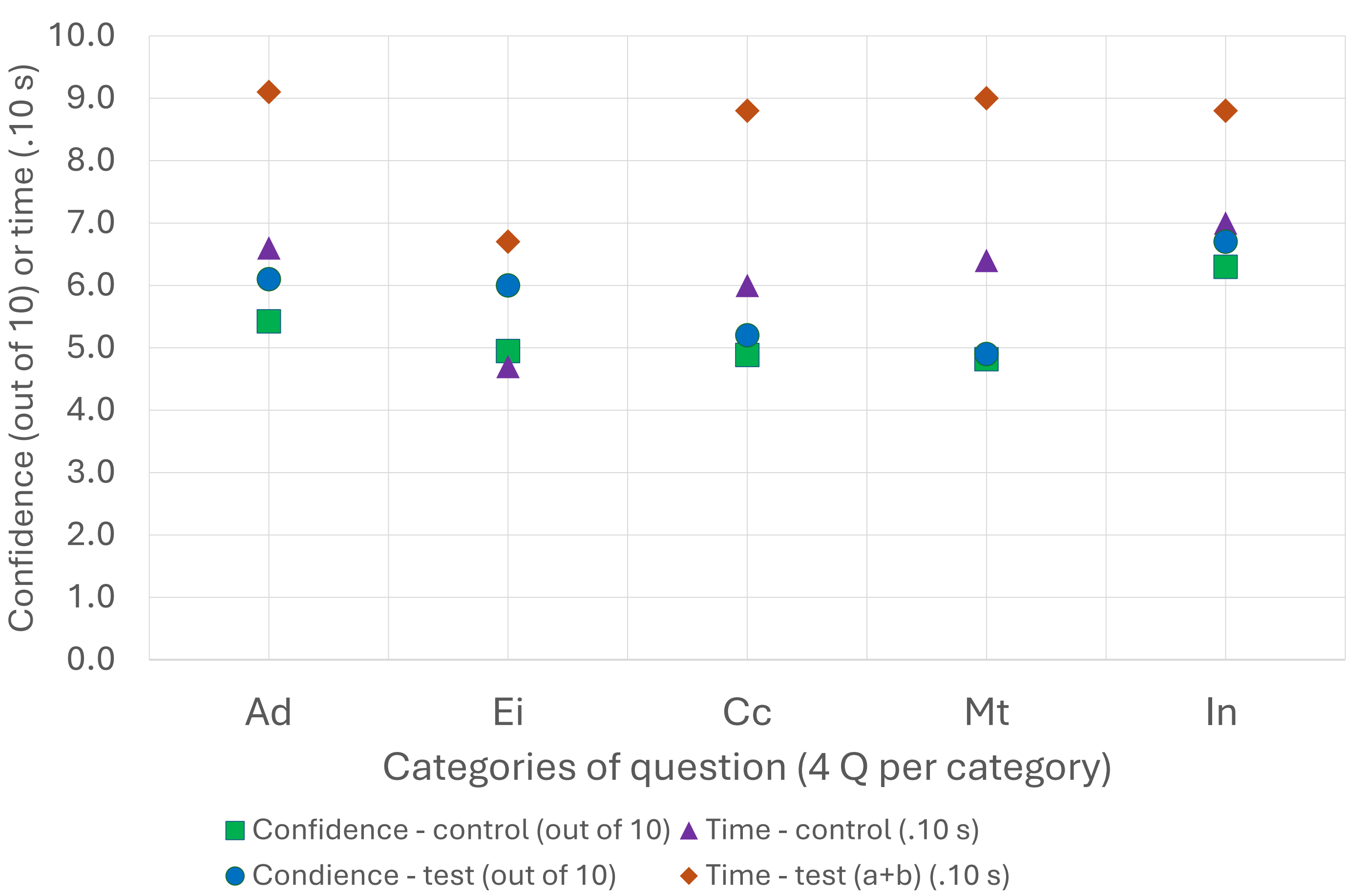
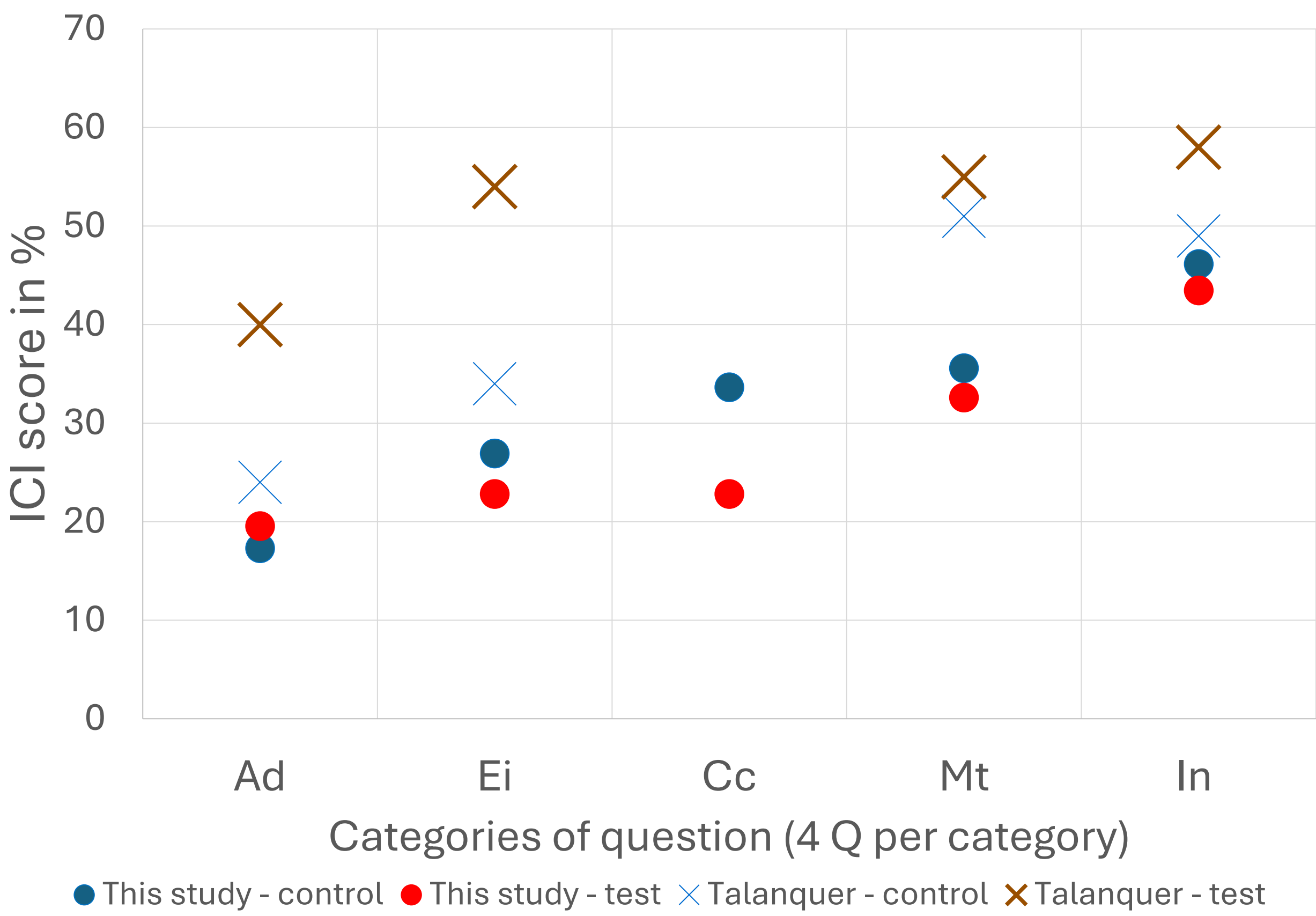
Method : Questionnaire



Method : Categories of misconceptions

- *Inheritance* (In): Tendency to attribute macroscopic properties to submicroscopic components
- *Additivity* (Ad): Tendency to assume that the properties of a composite entity are an average of the properties of its components
- *Matter tracking* (Mt): Tendency to over-rely on surface features and appearances to track matter during processes
- *Energy investment* (Ei): Tendency to think of molecules as springy structures that take effort to assemble but give back energy when broken apart
- *Centralized causality* (Cc): Tendency to think of chemical processes as driven by active agents with specific purposes and desires

Results



Perpective : what would you test?

Take home message

Vincent Natalis vincent.natalis@uliege.be Wrong answer prediction in chemistry	Vincent Natalis vincent.natalis@uliege.be Wrong answer prediction in chemistry	Vincent Natalis vincent.natalis@uliege.be Wrong answer prediction in chemistry	Vincent Natalis vincent.natalis@uliege.be Wrong answer prediction in chemistry	Vincent Natalis vincent.natalis@uliege.be Wrong answer prediction in chemistry	Vincent Natalis vincent.natalis@uliege.be Wrong answer prediction in chemistry	Vincent Natalis vincent.natalis@uliege.be Wrong answer prediction in chemistry	Vincent Natalis vincent.natalis@uliege.be Wrong answer prediction in chemistry
Small sample does not confirm the hypothesis of a - wrong answer prediction - effect, bigger dataset needed. Confidence and time measurements point to possible inconsistencies between types of questions and thus of questionnaire validity. Multiple extensions of research can be: broad vs. precise meta prompts, conceptual vs. classical questions, test the effect of fatigue	Small sample does not confirm the hypothesis of a - wrong answer prediction - effect, bigger dataset needed. Confidence and time measurements point to possible inconsistencies between types of questions and thus of questionnaire validity. Multiple extensions of research can be: broad vs. precise meta prompts, conceptual vs. classical questions, test the effect of fatigue	Small sample does not confirm the hypothesis of a - wrong answer prediction - effect, bigger dataset needed. Confidence and time measurements point to possible inconsistencies between types of questions and thus of questionnaire validity. Multiple extensions of research can be: broad vs. precise meta prompts, conceptual vs. classical questions, test the effect of fatigue	Small sample does not confirm the hypothesis of a - wrong answer prediction - effect, bigger dataset needed. Confidence and time measurements point to possible inconsistencies between types of questions and thus of questionnaire validity. Multiple extensions of research can be: broad vs. precise meta prompts, conceptual vs. classical questions, test the effect of fatigue	Small sample does not confirm the hypothesis of a - wrong answer prediction - effect, bigger dataset needed. Confidence and time measurements point to possible inconsistencies between types of questions and thus of questionnaire validity. Multiple extensions of research can be: broad vs. precise meta prompts, conceptual vs. classical questions, test the effect of fatigue	Small sample does not confirm the hypothesis of a - wrong answer prediction - effect, bigger dataset needed. Confidence and time measurements point to possible inconsistencies between types of questions and thus of questionnaire validity. Multiple extensions of research can be: broad vs. precise meta prompts, conceptual vs. classical questions, test the effect of fatigue	Small sample does not confirm the hypothesis of a - wrong answer prediction - effect, bigger dataset needed. Confidence and time measurements point to possible inconsistencies between types of questions and thus of questionnaire validity. Multiple extensions of research can be: broad vs. precise meta prompts, conceptual vs. classical questions, test the effect of fatigue	Small sample does not confirm the hypothesis of a - wrong answer prediction - effect, bigger dataset needed. Confidence and time measurements point to possible inconsistencies between types of questions and thus of questionnaire validity. Multiple extensions of research can be: broad vs. precise meta prompts, conceptual vs. classical questions, test the effect of fatigue

Affiliations

¹Université de Liège, chemistry department
²DIDACTIfen Research Unit

References

Houdé, O. (2000). Inhibition and cognitive development: Object, number, categorization, and reasoning. *Cognitive Development*, 15(1), 63–73. [https://doi.org/10.1016/S0885-2014\(00\)00015-0](https://doi.org/10.1016/S0885-2014(00)00015-0)
Ott, Y., Cauet, E., & Kautz, A. (2025). *Between Intuition and Reflection: A Critical Look at the impact of Metacognitive Prompts on solving the Force Concept Inventory*. ESERA 2025.
Talanquer, V. (2017). Concept Inventories: Predicting the Wrong Answer May Boost Performance. *Journal of Chemical Education*, 94(12), 1805–1810. <https://doi.org/10.1021/acs.jchemed.7b00427>