

RESEARCH DEFENSE • EVIDENCE BEFORE CERTAINTY

The quiet value of a pause

Does a brief reflection prompt help novice analysts notice uncertainty before making a recommendation?

Stillwater Research / fictional study • October 2026 • Synthetic study



What remains unanswered.

WHAT IS ALREADY KNOWN

Prior classroom exercises suggest that novices can identify caveats when explicitly asked. Those exercises do not show whether the habit transfers into an unprompted decision.

THE UNRESOLVED QUESTION

Can one short reflection prompt improve uncertainty recognition during an unfamiliar analysis task, without making the task impractically slow?

Does a brief reflection prompt help novice analysts notice uncertainty before making a recommendation?

How the question was studied.

A method explains what the evidence can — and cannot — support.

01

Within-session comparison

24 fictional participants completed two matched analysis tasks. Prompt order was counterbalanced; the task domain remained constant.

02

Blind rubric coding

Two fictional reviewers coded named uncertainties against a predefined rubric. Coding differences were discussed before scoring.

03

Descriptive analysis

Report paired summaries and observed spread. The sample was designed for exploration, not population-level causal estimates.

Make the study reproducible.

- 01

Recruit & balance

Include novice analysts with less than one year of practice. Assign the prompt-first condition using a fixed, recorded alternating sequence.
- 02

Run the two tasks

Allow 12 minutes per task and a 3-minute break. The prompt asks for missing information before the recommendation; no answer is supplied.
- 03

Code & preserve

Save the de-identified response, prompt order, coding rubric and reviewer disagreement notes. Exclude incomplete paired responses before comparison.

Three observations. Traceable origins.

+1.4

MEAN UNCERTAINTIES PER RESPONSE

More uncertainties were named after the prompt.

Descriptive paired difference; not a generalizable treatment effect.

SOURCE / S1

+38

MEDIAN SECONDS PER TASK

The reflection step added a modest amount of time.

A short task cannot establish the time cost in professional work.

SOURCE / S3

88%

RAW CODING AGREEMENT

Reviewers agreed on most coded observations.

Raw agreement is reported; no chance-corrected reliability claim.

SOURCE / S2

SUPPORTED INTERPRETATION

In this small synthetic study, prompted analysts named more uncertainties. Decision quality itself remains unproven.

DECISION THIS SUPPORTS

Proceed to a small transfer pilot; do not yet recommend the prompt as a production decision standard.

Observed results are not causal proof unless the study design establishes causality. Source details follow on slide 12.

What else could explain this?

01

Practice from the first task

Compare prompt-first and prompt-second sequences separately. A shared order effect would weaken the prompt interpretation.

02

Demand characteristics

Use a neutral instruction of similar length in a new session and conceal the uncertainty-recognition hypothesis.

03

Rubric-specific learning

Ask independent reviewers to score a transfer task with a different uncertainty taxonomy.

Where the answer stops.

- 01

Small exploratory sample

The observed averages are unstable. Replication needs a prospectively defined sample and analysis plan.
- 02

One task domain

Transfer to financial, medical or other consequential decisions was not studied.
- 03

Recognition is not quality

Listing caveats may improve a response without improving the eventual decision or its outcome.

Use the finding with its conditions.

ANALYST TRAINING

Pilot the prompt as a reflective exercise and invite discussion of missing evidence.

DO NOT INFER

Do not describe the exercise as a validated decision-quality intervention.

RESEARCH PLANNING

Use the paired protocol to design a larger transfer study with blinded outcome scoring.

DO NOT INFER

Do not use these synthetic values in a real paper or public performance claim.

Questions worth preparing for.

QUESTION 01 • SOURCE S1

Why use paired tasks?

Pairing reduces between-person variation in this exploratory exercise. Counterbalancing helps inspect, but cannot eliminate, order effects.

QUESTION 02 • SOURCE S2

Does 88% establish reliability?

No. It describes raw agreement in the synthetic coding pass. A future study should report the planned reliability statistic and uncertainty.

QUESTION 03 • SOURCE S3

Did recommendations improve?

That endpoint was not measured. The supported claim concerns named uncertainties and observed task time only.

THE NEXT USEFUL UNCERTAINTY



Run a preregistered transfer task with neutral instructions, blinded decision-quality scoring and a prospectively justified sample.

DECISION UNDER REVIEW

Proceed to a small transfer pilot; do not yet recommend the prompt as a production decision standard.

Trace the argument to its sources.

Source identifiers connect the observations and defense answers to their provenance.

S1

Synthetic paired response register

Fictional study appendix A · 24 paired observations · generated for this template.

S2

Uncertainty recognition rubric

Fictional methods appendix B · rubric v1 · two reviewers.

S3

Timing and debrief notes

Fictional process appendix C · session logs and exit comments.