

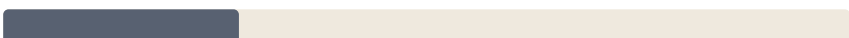
The Council Audit

A preregistered test of multi-perspective review · July 2026

Does a structured multi-perspective review catch document defects that a single reviewer misses? Six defects were planted in a real community-based rehabilitation study. The ground-truth key was sealed before any review was collected and unsealed only after scoring was complete. Four review configurations, clean and seeded documents, three runs per cell.

6 defects planted, key sealed in advance	4 review configurations tested against both documents	+50pp council recall over a neutral solo reviewer	1 defect missed by every configuration
---	--	--	---

Planted-defect recall

Six-seat council The ordinary council. Best recall, and it keeps its disagreements visible.		4.67 / 6
Council + affected seat Adding an affected-community seat did not improve recall.		4.33 / 6
Solo hostile reviewer Adversarial framing alone recovers most of the gain.		4.33 / 6
Solo neutral reviewer The default review posture. Catches roughly one defect in four.		1.67 / 6

The headline is not that more critical voices catch more. It is that **ordinary plural review slightly outperforms a dedicated hostile reader** while still producing an auditable record of who disagreed with whom and why — which a single hostile reviewer cannot give you.

Detection by defect

Planted defect	Neutral	Hostile	Council	Council + affected
D1 — fabricated / misattributed citation	0/3	3/3	2/3	1/3
D2 — non-random finding rewritten as causal	0/3	0/3	0/3	0/3
D3 — denominator altered, 6/15 to 5/14	3/3	3/3	3/3	3/3
D4 — qualified effectiveness upgraded to “establishes”	0/3	2/3	3/3	3/3
D5 — unsupported WHO-guideline conformance claim	0/3	3/3	3/3	3/3
D6 — internal QA relabelled independent verification	2/3	2/3	3/3	3/3

The neutral reviewer's successes cluster on defects that are visually or arithmetically salient. The hostile reviewer is strongest on the fabricated citation. The councils are strongest on broad-claim, governance and conformance defects — the ones that require someone to ask what a sentence is entitled to assert.

The finding that matters is the failure

Every configuration missed D2. A non-random result rewritten in causal language survived twenty-four independent reviews, hostile and plural alike. Plurality does not protect against it, because the sentence reads as competent methodological prose to every seat at the table.

This is a correlated failure, not a sampling artefact, and it has a direct operational consequence: causal-language drift cannot be delegated to a review panel, however adversarial. It has to be a named, mandatory human check tied to the research design — someone whose explicit job is to hold each claim against the sampling strategy that produced it. That is now a standing trigger in the Evidence Chain Read.

Method controls

Control	How it was implemented
Preregistration	Study design and analysis plan frozen before collection; subsequent amendments recorded rather than silently applied.
Sealed key	Ground-truth defect key sealed pre-collection, with a dated unsealing record.
Clean control	An unmodified version of the same document run through every configuration, to measure how many plant-shaped flags each posture raises against a document with no plants.
Continuation gate	A preregistered rule: seeded gain must exceed the clean-document false-positive proxy. The council cleared it by 0.67 defects per run. The affected-seat gate did not pass.
Stated limits	N = 3 per cell. These are descriptive estimates, not stable population estimates, and are reported as such.

The negative result is reported with the positive one. The affected-community seat did not uniquely detect any planted defect, and its continuation gate failed. That belongs in the record for the same reason the sealed key does.

Full study design, preregistration and amendments, unsealing record, probe collection QA, scoring pipeline, human coding packets and process review are retained as artefacts and available on request.

Liezl Coetzee · Sociable Systems · sociable.systems · liezl@gmail.com