KEEP SCORE BEFORE YOU TRUST THE READ

Measure what self-coverage catches, misses, and overcalls.

Self-coverage becomes useful when its claims are recorded before review and compared with the notes that actually return. Page one records a scorable preflight before send-up. Page two compares it with returned notes, logs misses and false alarms, and allows only repeated patterns to reach human rubric review. *Anything set as dialogue is an instruction you can give the agent word for word; adapt names, dates, and file paths to the production.*

PREPARED FOR THE AVS SCREENWRITING TEAM · JULY 2026

FADE IN:

TURN THE PREFLIGHT INTO A RECORD

Self-coverage already makes provisional claims about what a committee may flag. Without a dated record, those claims disappear when the real notes arrive, and memory quietly grades the agent in its own favor. Calibration begins by separating what the preflight actually said from what hindsight wishes it had said.

Rule 1: Score only specific concerns filed before send-up.

Rule 2: Preserve ambiguity, conflict, and contrary evidence. A generous match teaches nothing.

Rule 3: Patterns earn proposals, not automatic rules. A person decides what changes the rubric.

1. **Give every production its own ledger.** Save self-coverage under `projects/[production]/coverage/` and create `projects/[production]/predictions/` beside it. In `predictions/`, keep a dated file per draft, `scorecard.md`, and `watch-list.md`. Predictions record expectations; the scorecard records outcomes; the watch-list holds repeated results worth examining. Use one draft identifier across coverage, predictions, notes, and scorecard so versions cannot cross.

2. **Make each concern scorable.** A prediction must name the draft and filing date, page or beat, substantive issue, rubric criterion, original source note, and confidence. “The middle may drag” is too vague. “On p. 9, the time jump may obscure chronology under the clarity criterion” can be tested. Confidence records uncertainty; it does not rescue a weak claim. If the agent cannot point to a location, criterion, and source, keep the thought in coverage but out of the ledger. Split bundled concerns: pacing and motive at the same beat are two claims, and either may survive while the other fails.

3. **Save the coverage before extracting predictions.**

The Writer
(to the agent)

> PROPOSE only. Read `projects/return-visit/draft-3.pdf` against `rubrics/committee.md` and `rubrics/[stage]-reviewer-profile.md`. Show the full contents proposed for `projects/return-visit/coverage/2026-07-10-draft-3.md`, separating evidence, inference, and conflict. Change nothing; save nothing.

*You should get: full cited coverage and no file changes. **Next:** verify citations, then APPLY only the named coverage file.*

4. **Freeze the record before the draft leaves.**

The Writer
(to the agent)

> PROPOSE only. From `projects/return-visit/coverage/2026-07-10-draft-3.md`, copy every concern into a full preview of `projects/return-visit/predictions/2026-07-10-draft-3.md`. Record page or beat, criterion, original source note, and confidence. Add nothing. Change no source or draft; save nothing.

*You should get: a dated prediction preview traceable to coverage and evidence, with no file changes. **Next:** correct it, APPLY only that file, verify the saved copy, then send the draft up.*

5. **Protect the before-and-after line.** Once the draft goes up, do not revise its prediction file. When notes return, archive them verbatim with reviewer, date, stage, and draft. Work from the scorecard, never by cleaning the frozen record. If you discover an error later, append it to `projects/return-visit/predictions/corrections.md` with discovery time and affected prediction.

The Writer
(to the agent)

> After I attach or paste the returned notes: PROPOSE only. Stop if `evidence/return-visit/draft-3-notes.md` already exists. Preview that file in chat with reviewer, date, stage, and draft metadata; preserve the note text verbatim. Do not create or edit files.

*You should get: a named evidence-file preview with verbatim notes and no file changes. **Next:** verify it, APPLY only that evidence file, and freeze the saved copy.*

ONE SCORABLE CONCERN

PREDICTION 03 *p. 9 time jump*
Criterion *clarity / chronology*
Source *Ryan note, 2026-05-14, draft 2*
Confidence *medium*
Outcome *pending*

READY WHEN

You can open one frozen prediction file and identify exactly what was expected, where, why, and when it was recorded — without looking at the later notes.

CUT TO:

GRADE THE PREFLIGHT AGAINST RETURNED NOTES

The scorecard measures usefulness against observed notes, not objective story truth and not access to a reviewer's mind. Silence does not prove a predicted concern was artistically wrong; a committee may notice an issue and choose not to mention it. Score only what the record supports.

Pair the records in both directions. Match each prediction to notes by substance and location. Then check each returned note against the frozen predictions. Compare issues, not wording. Repeated misses, false alarms, disputed matches, and contrary examples may enter the watch-list; one round does not rewrite a rubric.

Use four classifications.

- **Hit:** a returned note raises the same substantive issue in the same scene, beat, or page range.
- **Miss:** a returned note raises a material issue absent from the frozen predictions.
- **False alarm:** no returned note raises the prediction. This describes only the observed review; it does not prove the concern wrong or unnoticed.
- **Unresolved:** wording or location overlaps, but the evidence is too ambiguous to classify.

Keep unresolved cases unresolved. Do not turn a thematic resemblance into a hit or treat every unmentioned concern as a definitive false alarm. Classification describes the alignment between two records; it does not rank the note, the writer, or the draft. Quote enough context to let another person challenge the match.

The Writer
(to the agent)

```
ASK only. Compare evidence/return-visit/draft-3-notes.md with
projects/return-visit/predictions/2026-07-10-draft-3.md.
Classify each prediction as hit, false alarm, or
unresolved. Quote it and any match; state when none
exists. Then classify material notes absent from
predictions as misses, quoting each note. Quote
both files only when both are relevant. Change
neither file.
```

You should get: cited classifications, with ambiguity unresolved. Next: approve or correct each before it enters `scorecard.md`.

The Writer
(to the agent)

```
APPLY my approved classifications to
projects/return-visit/predictions/scorecard.md and, when
warranted,
projects/return-visit/predictions/watch-list.md. Append to
the scorecard; do not rewrite prior rounds. Update
the watch-list only for a specific pattern in at
least two independent rounds; cite both. Show the
diff. Leave notes and predictions unchanged.
```

You should get: one appended round and, only when warranted, a traceable watch-list entry. Next: inspect the diff before saving the snapshot.

Read categories, not a vanity percentage. Hits identify checks that usefully direct attention. Misses reveal where the rubric or its application may be blind. False alarms consume attention and can pressure a writer away from a sound choice. There is no universal acceptable hit rate: stage, reviewer mix, draft condition, and note-taking habits all change what can be observed.

The Writer
(to the agent)

```
ASK only. Read the last six rounds in
projects/return-visit/predictions/scorecard.md. Name
recurring misses and false alarms by rubric
category. Show supporting rounds and contrary
examples. Do not propose a rule from one outcome.
Do not create or edit files.
```

You should get: categories with cited rounds, contrary examples, and no forced verdict. Next: decide what belongs on the watch-list.

Independence matters more than volume. Six drafts shaped by one repeated note may be one signal, not six. Before changing a criterion, ask whether the pattern survives different rounds, stages, or reviewers and whether an exception explains it.

The Writer
(to the agent)

```
PROPOSE only. For watch-list items supported by at
least two independent rounds, preview an addition,
softening, or retirement for rubrics/committee.md.
Cite the underlying notes, state confidence and
exceptions, and show the proposed diff in chat. Do
not create or edit files.
```

You should get: inspectable proposals citing multiple independent rounds. Next: adopt, revise, defer, or reject each; APPLY only an approved rubric change.

Habits for every scoring round.

- File before send-up; score only after notes return.
- Quote both sides of every claimed match.
- Count unresolved as unresolved and retain contrary examples.
- Review category patterns, not one dramatic round.
- Keep raw predictions and notes unchanged.

WHEN THE SCORE CANNOT BE TRUSTED

If the agent rewrites a prediction, matches by vague theme, or cannot cite the record behind a classification, discard it. Restore only the unauthorized change, preserve later human work, and rescore narrowly. An explanation cannot repair contaminated evidence.

Guardrail. Calibration is not clairvoyance. The ledger cannot prove what the committee noticed, what a reviewer privately thought, or whether an unmentioned concern was artistically valid. It can show where self-coverage has been useful, noisy, or blind under observed review conditions. Use that record to direct human attention, never to replace judgment.

PROOF OF DONE

After six independent rounds, or more if evidence is thin, show which categories produced useful hits, any recurring misses or false alarms, and each human-approved rubric decision, with supporting notes and frozen predictions. If no pattern appears, record that; do not manufacture one.

FADE OUT.