CATALOG 5 / REDESIGN AND PAIRED REVIEW

On the Record Comparative Review

Version 4 centers the assessment on a single political-economy question. Students trace what officials did and what a coalition may have changed. The first judgment stays visible when later Ecuador records open. The Colombia file asks students to apply the same distinction to a related decision with a different outcome. The final rubric totals 100 points, and the writing floor is 340 words.

Rank	Release	Design score
1	Version 4	26/28
2	Version 3	22/28
3	Version 2	21/28
4	Version 1.1	18/28

5 of 5 paired AI reviewers placed Version 4 first. The strongest gains were fit to the assessed reasoning and the quality of the evidence that bears on it. The score records design judgments across the same criteria.

Why the panel chose the staged audit

Paper candidate	Observation from the common-case trace
Full evidence brief	Supports a qualified final explanation. It can show how a record bears on a claim, while the first warrant has to be reconstructed afterward.
Fixed reveal	Makes the provisional inference and its response to a dated challenge visible in the report. Everyone works from the same opening records.
Choice of first inquiry	Can expose evidence strategy if students predict what would matter. Putting the settlement or dispatch record first would give an unequal preview. The extra choice earned a weaker case for this assignment.

What the team preserved

The original disciplinary, assessment, learner, interaction, art and testing perspectives returned. They qualified the cases, tried the implications of opening each record at a different time, wrote and filed responses, and tested recovery. The redesign reuses the evidence reader, sealed filing, separate practice mode, progress backup and human-scored report. Version 3 remains available as an exercise in comparing uneven sources.

COMPARATIVE RATINGS

Seven Shared Catalog Criteria

Each reviewer used the closest common task in the exact four releases. They worked with the earlier fuel route and the rebuilt Ecuador and Colombia file. The common task was to identify a bounded decision, connect evidence to a mechanism, evaluate the rival or limit, revise the warrant, and transfer the distinction. Review order varied. Differences in the actual files were recognizable.

Criterion	V1.1	V2	V3	V4
Format fit and learning validity	3	3	3	4
Content and rigor	3	3	3	4
Engagement and agency	2	3	3	4
Language	3	3	3	3
Doability and accessibility	2	3	4	4
Art and design	2	3	3	4
Assignment feasibility	3	3	3	3
Sum of medians	18/28	21/28	22/28	26/28

A 0 marked an absent task, a 1 a serious problem, and a 2 a workable task with a concrete weakness. A 3 required an observed ordinary task and a 4 an additional relevant challenge. The median uses reviewers who scored the dimension for all four releases. Adding medians gives the descriptive design rank.

Paired lens	V1.1	V2	V3	V4
Political economy and comparative state institutions	20/28	24/28	24/28	27/28
E-learning and assessment design	18/28	21/28	22/28	24/28
Simulated first-year comparative politics student after the lecture on the state	18/28	22/28	22/28	27/28
UX and accessibility	18/28	21/28	22/28	25/28
Art direction and evidence design	20/28	24/28	25/28	28/28

The differences that drove the rank

Version 1.1 already preserved initial prose and a gradeable report. Versions 2 and 3 made evidence and navigation easier to use, with Version 3 improving recovery and rival-testing directions. The earlier cases still asked a learner to explain a common result across records that supported different causal inferences. Version 4 offers one documented bargain and a new file about the same type of decision. That changes what the student can show the grader.

Matched ratings are AI design observations. Feasibility remains sensitive to actual first-year reading time, accommodations and grading agreement.

EVIDENCE AND SCORING

What the Reviewers Changed

Review finding	Redesign response	Test or boundary
A stronger contrast needs a shared decision.	Use the Ecuador settlement about diesel and the first effective Colombia rule after the trucker deal.	The earlier Ecuador fare agreement remains important evidence, with a different policy outcome.
A disputed letter would make the claim fragile.	Use CONAIE's October 12 public announcement calling for repeal or revision, as reported by two outlets.	A demand identifies what the organization put on the agenda. What officials privately intended remains open.
A negotiated term and delivery can be inferred from different records.	Add EC5 dispatch and checking activity, CO3's published reference schedule, and CO4's local reporting.	The local prices come from separate visits. Preserve credit for reported administrative activity and the limit on general compliance.
Two Colombian figures made the baseline easy to mix.	Use the same Cali reference in the prompt and table. Preserve the government's distinct headline in the context record.	Cali moved 9,892 to 11,800 to 10,292 COP per gallon. A learner must reason from the common reference series.
A fluent story can mix policy timing and causality.	Hold the first claim fixed and ask what a specific record changes. Keep one rival for the same decision.	A supported change, a justified retention and a fluent overclaim all filed. The instructor must judge the reasoning.

A limit can be a strong answer

One response can retain its opening judgment that the fare agreement settled only the transport leaders' dispute. The new records strengthen it and clarify what the remaining coalition secured. A different student can revise a prediction that partial accommodation would preserve the price. Both can earn full credit by explaining how the new record affects that inference.

A proposal for government retreat independently motivated by supply problems is a rival about the decision to repeal. A forecast that fares alone could end the strike while keeping the rule would predict a different outcome. The guide gives a model of how a dated decision record could distinguish the plausible causes of the observed repeal.

Ecuador chronology · October 12 public statement · Colombia's agreement · Cali reference table

VERIFICATION

Results on the Exact Files

Measured item	V1.1	V2	V3	V4
Matched workflow suite	43/43	43/43	43/43	43/43
Required words	450	450	340	340
Required responses	4	4	4	3
Case body words on matched route	651	657	657	613
Opening view text in the model	597	610	610	378
Rotating main case pairs	3	3	3	Common case
Gradeable report	100	100	100	100

What ran

The matched suite used 41 common assertion categories and two checks tailored to the shape of each release. It included starting Class, staged reading, exact text preservation, draft restoration and the gradeable report. The two archived interim builds also passed it. Separate historical replay found all 24 opening routes intact on each older file. Version 4 uses one common case, so that route count serves a different purpose.

Version 4 also passed 53 release-specific checks, covering exact filing floors, source reveals, the sealed first claim, separate modes, data and figure labels, and returning to a draft after a failed save. The older backups remained usable through their original releases. An independent two-page conversion of the print output contained the three exact responses, nine record IDs, the rubric and the end marker.

Release identities and limits

Exact file	SHA-256 prefix
Version 1.1	7d2afade6f5d
Version 2 first review build	c8c90d555208
Version 2	266d2260012e
Version 3 first review build	b3014c9078af
Version 3	9a113a7b9f1e
On_the_Record_V4.html	83bcf61b9f52

The checks ran in a simulated document and event model. The PDF proof used captured print content. Native browser reflow, screen reader interaction, the print dialog and classroom performance require live testing.