

Ten fresh perspectives. Concrete revisions.

Five simulated novice students and five simulated faculty used the frozen first version before changes.

Main finding: all ten completed a case and comparison and reached a filed report. They found strong source discipline, attractive visuals and recoverable navigation. They also exposed vocabulary friction, redundant claim fields, easy answer cues and a gap between changing the board and revising the paragraph.

Each reviewer started without design notes, source code, hidden answers or other reviewers' comments. Student roles represented uneven ordinary preparation rather than exceptional performance. Faculty could read the facilitator guide only after recording their first understanding and playing. Separate save spaces prevented reviewers from overwriting one another.

ID	Perspective	Main / comparison	Observed outcome
S1	Willing, low confidence	Russia / Malawi	Filed; 72-word account, 39-word comparison
S2	Skimming, impatient	Malawi / Russia	Filed after recovering three missing items
S3	Writing-reluctant pair	Malaysia / Malawi	One shared record filed; 64 / 37 words
S4	Confident, confirming facts	Serbia / Russia	Filed; weak current claims remained beside a narrower paragraph
S5	Slow reading, limited terms	Malawi / Malaysia	Filed; sampled keyboard actions worked
F1	Comparative politics	Russia / Malaysia	Filed; challenged assessment alignment
F2	Research methods	Malawi / Serbia	Filed; source and causal critiques
F3	Learning design	Malaysia / Malawi	Filed; affected-actor scaffold critique
F4	Inclusive course design	Serbia / Malaysia	Filed; reproduced keyboard orientation issue
F5	Assessment and game design	Malawi / Serbia	Filed; challenged mastery inferences

These are AI role simulations and expert walkthroughs. Ten completions are not a human usability sample, a validated completion rate, a timing result or evidence of learning gains. The first-use reviews remain unchanged; informed retests are separate.

What they said, and the response

Full comments, action logs and sample explanations are included in First_Use_Reviews in the kit.

ID	First-use comment or finding	Response implemented
S1	"Two anchors" could mean two countries; the hearing asked for explanations without an obvious place to record them.	Main case and supporting evidence now have different names. A short rival explanation and evidence request are explicit.
S2	Applying a narrowing felt like finishing the revision, although the paragraph stayed unchanged.	Before/after board text, an explicit unchanged-paragraph notice, a focused review action and a final reconciliation check.
S3	"MY-D" looked like another card; "civil forfeiture" and party labels required decoding.	Country actor keys, card/source explanations, legal terminology and a searchable glossary. The allegation question has one clearer source focus.
S4	The complication became another confirming fact; old causal claims and attachments survived beside new writing.	Prompt asks what becomes less certain. Changed-step attachments clear; students review all current claims and the selected supporting evidence.
S5	Three controls per chain step, another revision choice, and another limitation phrase felt repetitive.	One current paragraph, compact optional starters/attachments, no required duplicate limitation phrase or overall revision selector; notes wrap.
F1	Choosing the cautious answer is easier than producing a rival explanation; too many claim objects.	Closer alternatives plus an authored rival and evidence request. One retained opening and one current written explanation.
F2	Two card IDs are not necessarily independent corroboration; source limits should fit the claim.	Shared-source notice and source-role prompts; full provenance stays available. Human review still judges the causal link.
F3	The options field could repeat PH rather than identify another actor whose choices changed.	Explicit affected-person/group prompt and a nearby neutral example; original political evidence is preserved.
F4	Reopen scrolled to a card but keyboard focus stayed at the stage heading. Missing-item errors were generic.	Requested-card focus, return-to-task controls and direct finishing links to the relevant field.
F5	Completion is submission, not mastery. Plausible rivals and human calibration examples are needed.	Practice is labelled as practice; revised question bank and a five-page guide with weak/strong fictional examples and a human rubric.

Less decoding. More visible reasoning.

The four cases, historical results, precise source caveats and comparison requirement remain.

Plain language and orientation

Main case, supporting evidence, evidence board and Finish & save replace ambiguous overlapping labels. Actor keys explain the cast. The glossary defines denominator, majority versus largest share, ticket, source role and other terms. It is deduplicated, sorted and searchable; case notes are labelled separately.

One explanation, with a visible revision handoff

The opening is retained as a first version; the current paragraph is shared across stages. A board revision shows before/after wording and sends the learner back to review the paragraph. The app does not rewrite student prose. Earlier planning notes and old question rounds remain recoverable rather than masquerading as current writing.

A clearer thinking requirement

A short rival explanation and a specific evidence request are visible work for a reader to assess. A complication can be a rival, source limit or missing step; it need not contradict everything. The main paragraph must cite the two currently selected supporting cards. Self-checks ask the student to reconcile the board, evidence and paragraph.

More discriminating practice questions

Twelve active questions now use closer mistakes about populations, units, timing, source status and necessary versus sufficient explanations. Two faculty-role reviewers independently checked the revised bank. Their corrections removed an ambiguous subgroup claim and several invented-event distractors. Old questions retain their original wording in saved histories.

Reading and recovery behave consistently

Reading counts update immediately. Opening the same source through the hearing counts as source access. Notes wrap. Reopen focuses the requested card and provides a return route. Finishing checks identify the missing item and take the learner to it. A shared evidence request now survives first comparison selection and later country changes.

What we kept and what we deferred

Reviewers consistently valued the calm collage design, visible known outcomes, limited board, editable connections and source limitations. We kept them. We did not add timers, points for changing views, invented electoral effects or a machine interpretation grade. More competitive audience dynamics are an optional future design question after the classroom pilot, not another required phase now.

The revisions improve instructions and make reasoning inspectable. They do not guarantee that a selected complication is logically effective or that an explanation is well supported. A human must read the report and discuss the claim.

What was retested

Initial blind reviews and later informed checks are different forms of evidence.

Check	Result and scope
Student S2 retest	Finished the revised Malawi/Russia route. Verified before/after handoff, unchanged prose, live counts, rival/request fields and self-check recovery.
Faculty F4 retest	Verified requested-card keyboard focus, next-Tab destination, return to the hearing, actionable finishing links and source access through hearing drawers.
Student S3 retest	Verified terms, party key, focused legal question and revision handoff. Confirmed the request reset. Final completion was not retested because the QA route became unavailable.
Content verification	Two informed faculty-role checks reviewed all 12 revised active questions against the original evidence. Corrected ambiguity and grounded the closer alternatives.
Logic and migration	128 automated state/content assertions across four cases: source fidelity, chosen-card citations, review flags, filing, replacement, escaping, old questions/attempts and evidence-request preservation.
Document review	The revised guide and this report were rendered and visually checked. Full first-use comments are preserved in the kit.

A real regression found during retest

S2 noticed a request disappearing after choosing a comparison. Code inspection confirmed that country selection recreated the comparison object without the newly shared request. The revised handler preserves it. Event tests cover first selection, later switching and continued editing. A final rendered-browser check also confirmed preservation on first selection and a later country switch. The original retest note keeps its initial, cautious description.

What this does not establish

No human students or external faculty participated. Browser automation time is not learner time. Protocol timeouts recovered with reload and Resume; they were not classified as application failures. Sampled keyboard behavior is not a full accessibility or physical-phone audit. In-game download completion remains unconfirmed in the cloud browser; report generation and preview are checked, with print/copy alternatives available.

The next useful pilot

Run the 55-minute plan with an actual class. Record phase-finish times, misunderstood terms, whether revisions reach the prose, and the quality of the rival/evidence request. Review initial and revised work with the five human criteria in the guide. A later new-case task can probe retention; estimating learning effects requires a suitable comparison design.