

Twelve fresh reviews. Stronger reasoning.

Technical editing, teaching, political science, student use and responsible public demonstration.

Main finding: the activity is understandable and offers defensible research work. Remaining problems involved methodological alignment, state-dependent instructions, source provenance and reading burden. The revision strengthens those areas while preserving the evidence, writing and comparison demands.

Every reviewer started in a fresh agent context without earlier feedback, code or design notes. Five used the visible frozen game; seven cold-read exported player screens. The text packet expanded disclosures and contained no practice-answer feedback. Faculty consulted permitted companion documents only after their cold understanding. These are AI role simulations, not human participants.

ID	Perspective	Review scope / observed outcome
R1	Technical/copy editor	Cold read: Russia/Malaysia and recurring instructions.
R2	Research-teaching specialist	Cold read: Malaysia/Malawi and guide alignment.
R3	Electoral-authoritarianism scholar	Cold read and primary-source spot checks: Russia/Serbia.
R4	Comparative methods	Browser: substantive Malawi/Malaysia record filed and previewed.
R5	Expert undergraduate	Browser: all stages; deliberately weak final writing reached completion.
R6	UX/accessibility reviewer	Browser: stages 1-5 verified; final-stage attempt limited by protocol failure.
R7	Law/democratization scholar	Cold read: legal claims, provenance and sequence.
R8	Novice Gen Z student	Browser: substantive Serbia/Russia record filed.
R9	Working Gen Z student	Cold read: terminology and likely misunderstandings.
R10	Educational creator lens	Browser: evidence selection, challenge and task inspection; no final filing.
R11	E-learning/course-design specialist	Cold read: scaffolding, assessment and access risks.
R12	Second technical editor	Cold read: form instructions and completion consistency.

Two browser reviewers filed substantive final writing; one filed deliberately weak final writing. The other two did focused/partial reviews. Do not convert these different scopes into a completion rate. All twelve full reviews are preserved in the kit.

Clearer prompts. Greater rigor.

The changes make the analytical work explicit, rather than remove it.

Problem found independently	Response implemented
A later step could pass as a rival explanation.	The rival addresses the same observation or link. Accounts may both contribute; a later step alone does not test a rival.
A generic request for more evidence could satisfy the prompt.	The short request now asks for different expected patterns and an actor/period. A fictional example models discrimination without supplying a country answer.
Malaysia/Malawi could be labelled different outcomes despite both ending in turnover.	Comparison labels are neutral. Students name the event and period, compare an affected option, and state a limiting difference. A failed analogy is legitimate.
A universal before/after prompt could invite an invented baseline.	Describe an option enabled, limited or left open. State where the earlier condition is inferred. Actions, rules and held advantages are all legitimate starting points.
A Serbia starter put non-turnover in the required option field.	The starter now names the pending Constitutional Court route. A rule/allocation route is also available; a result alone remains insufficient as a mechanism.
Complication handling appeared late in paragraph sign-off.	The opening and drafting instructions now state it. The same 60-85-word paragraph must use both supporting IDs, address the complication and give a claim-specific limit.
Institutional summaries and legal allegations shared an imprecise label.	Malaysia source badges identify IPU election summaries and the legal announcement. The allegation status stays explicit in the headline and source notes.
Some document caveats were remote from the actual claim.	Card-specific limits now foreground legal mediation, implementation, exposure, voter transfer or remedy timing as appropriate. Full provenance and numerical discrepancies remain available.

Other preserved distinctions include reported support versus unrestricted preferences, monitored content versus audience impact, declared returns versus a corrected count, seat conversion versus biased competition, and opening a contest versus choosing its winner.

Make the next action truthful.

Fewer misleading signals, with the same authored reasoning requirements.

Revision guidance matches the actual situation

No revision warning appears above an empty first paragraph. An older narrowing is labelled as earlier wording, with the current field also shown. It no longer impersonates the latest change after an evidence replacement. The student still revises the paragraph; the app does not write it.

Controls state what they do

Hearing actions say Record my reason, Apply this revised step, or Replace this supporting card. Pressed-state semantics identify the selected decision. Card, source, question and missing-item controls have contextual accessible names. Required fields and optional role labels are clearer. A decision reason now needs a valid main-case card, table or source ID.

A reliable next action and a clear hand-in boundary

Guidance follows the first actual missing requirement, including preparation, practice, review and saving. Finishing marks a local record for human discussion and explicitly says reasoning is not assessed. It does not submit anything to a teacher. The readable report and editable resume file have plain labels; optional application/briefing work follows the primary finish/save area.

A saved-record bug found and fixed in the informed retest

An earlier saved completion could coexist with missing current items. The retester observed complete guidance beside an empty rival and unchecked review. The fix retains the earlier timestamp but classifies the current record as draft until its current requirements are met. The corrected next action, on-screen status and report preview were all verified in the browser. Valid completed imports remain complete.

Engagement through a visible evidence dispute

A new entry link reaches case selection quickly. Casual mode progressively discloses the full task without skipping requirements. The existing editorial collage, open evidence, three-space attention constraint and source-dependence warning remain. The creator review includes an accurate short-demo outline, clearly identified as one move inside a full activity.

The creator lens is a simulated editorial perspective, not an influencer endorsement, TikTok audience study or forecast of engagement. No content was posted.

A complete record is not a good argument.

The adversarial student route makes the need for human review concrete.

Demand or safeguard	Preserved / strengthened
Evidence	All 24 historical card texts preserved. Two distinct supporting cards and a separate complication remain required.
Authored explanation	One current 60-85-word paragraph, both selected supporting IDs, the complication, an affected option and a relevant limit.
Challenges and revision	Two source-use practice attempts and two reasoned decisions. Keep, narrow and replace remain legitimate; every retry is retained.
Alternative and inquiry	A competing account of the same observation and a request specifying potentially different evidence patterns.
Comparison	One 30-45-word second-case explanation, valid card ID, named event/period and a process difference that tests or bounds the claim.
Sources	Original records, locators, caveats, shared-source warnings and the unexplained official-total discrepancy are retained.
Assessment	No semantic auto-grading, no historical winner changed, no points for a political conclusion, no reward for speed or changing one's mind.
Support	A student-visible fictional example and human criteria; no country paragraph supplied for copying. Accommodations remain a teaching responsibility.

What the weak-writing control test demonstrated

R5 entered a generic main explanation and comparison that met the word/ID checks, then deliberately gave generous self-confirmations. It reached the declared completion state. This is a real limitation of structural checking, consistent with the stated design. The revised status makes that boundary harder to mistake; adding a hidden prose score would not resolve it reliably.

Read the argument, not the green state

A human should locate the distinct observations, the actor-to-option warrant, the limit that narrows that claim, a meaningful complication, the competing account and the comparison's relevance. A correct fact with a generic caveat can still be a weak explanation. Participation credit is not demonstrated mastery.

What this review establishes

A revised game, documented checks and explicit limits.

Check	Scope and result
Automated regressions	198 new + 128 existing = 326 assertions passed. Node VM event/state and HTML-output checks, not browser learning evidence.
Fidelity	24 historical card texts and all 12 active/12 legacy answer keys match the frozen baseline. Word limits, analytical gates and old attempt histories remain.
Informed browser retest	Verified empty-paragraph guidance, contextual names, pressed states and revised comparison prompts. Found and verified the saved-completion fix. This was targeted QA, not another blind full playthrough.
Source review	R3 spot-checked primary Levada, ODIHR, HRW, CINS and Serbian-government material. No blocking numerical falsehood was established. R7 checked the supplied legal/source documentation; original Malawi judgments remain uninspected.
Documents and package	Guide, this review and source-use clarification were rendered and visually checked. Full first- and second-round reports, response register and retest notes are included.

Suggestions retained as future options

A one-card preview route, a focused single-question layout, new allocation exercises and a new individual pair assessment may be useful, but would add routes or tasks requiring fresh testing. They were not silently substituted for the full research activity. Physical-phone and full screen-reader testing still need appropriate devices/users.

What remains unverified

No human students, faculty or influencers participated. Browser protocol failures limited two initial scopes and some final edits; reports name those limits. No full new replay, cloud-download completion, human timing, retention, accessibility certification or learning effect is claimed. The 55-minute schedule is reconciled across the game and guide, but remains a planning estimate.

The useful next test

Pilot with actual students of varied preparation. Record stage times and confusion, inspect whether revisions reach prose, and rate claim-specific limits, rivals, discriminating requests and comparison scope. Separate completion, participation and analytical performance. Use a delayed new-case task for retention; causal learning claims need an appropriate comparison design.