Supplementary Information for

When Threads Fall Quiet:

Tracing Idea Histories Across cMOOC Discussions

Target journal: International Journal of Computer-Supported Collaborative Learning

This anonymous file accompanies the double-blind manuscript. Verified author and institutional metadata will be added in the submission system or publication version as required by the journal.

1 Supplementary Methods and Supporting Analyses

1.1 Sparse relation-enumeration sensitivity analysis

The primary RQ2 analysis counts each source idea once when the combined list contains at least one human-coded substantive development. The sensitivity analysis uses a different unit: every positive source-to-later-span pair. The combined list contained 72 of the 131 coded positive relations for sources with $R_{72} = 0$. A stratified sample of 228 relations outside the list contained nine positives across eight source ideas.

The nine positives comprised seven `ELABORATE`, one transfer, and one new-evidence relation, with three in each delay band. Their score-band distribution was two in P00–80, one in P80–95, four in P95–99, and two in P99–100. Inverse-probability weights summed to 83,339.5; the largest was 33,109.5, and the three largest supplied 88.3% of the total. A small number of observations from sparsely sampled strata therefore drive much of the weighted estimate.

We treat this as a sensitivity analysis of source-to-later-span pairs. The unit is neither a distinct later episode nor a source idea, so it cannot estimate developments per idea. Its wide interval and concentrated weights make it a boundary on exhaustive pair enumeration rather than a central

source-idea result. It does not independently validate, correct, or break the source-level 30/36 visibility result. Table S1 reports the weighted quantities under this unit.

Table S1: Sparse pair-level relation-enumeration sensitivity analysis for the 2024 combined candidate list

Measure	Result and unit
Coded-reference-set relation coverage	131 positive source-idea–later-span relations; 72/131 (55.0%); 95% CI [35.9%, 83.3%]
Stratified sample outside the combined pool	228 sampled relations; 9 human-coded positive relations
Estimated positive-pair density outside the pool	1,859,339 structural relations outside the pool; 4.48%; 95% CI [0.72%, 8.90%]
Design-weighted pair-enumeration coverage	Estimated positive source-idea–later-span pairs; 0.086%; 95% CI [0.043%, 0.532%]

Note. The 72/131 interval used 20,000 source-idea-cluster bootstrap resamples. Estimates outside the combined list used fixed stratum-specific inverse-probability weights and a stratified finite-population-corrected bootstrap with 20,000 replicates. Only nine sampled relations were positive. Their largest three weights contributed 88.3% of the weighted positive total. The unit is the source-idea–to–later-span pair, not a distinct development event or source idea. These values are not comparable with the 30/36 source-level visibility result: they describe pair-enumeration sensitivity and neither validate nor correct source-level visibility. They are not used to estimate incidence.

1.2 Temporal eligibility and first-stage source sampling

The three offerings ran from 1 November to 27 December, 1 November to 22 December, and 31 October to 26 December 2024. Discussion participation was required, and all discussions were conducted in Chinese.

All timestamps used for eligibility decisions were interpreted in the Asia/Shanghai timezone. The early-response interval included both endpoints. A resolved cross-author learner reply counted as early when $0 \leq t - t_0 \leq 72$ hours. Later relations required $t > t_0 + 72$ hours and no later than the course-specific administrative endpoint. All structurally eligible relations were considered through that endpoint. The 14-, 30-, and 60-day cut points reported delay or observation availability.

Records were included only after the required structural keys and creation times had been resolved. A focal contribution was excluded from the $R_{72} = 0$ source frame when an early reply came from a participant whose role could not be resolved. Posts and comments could supply later spans. These rules were applied before candidates were ranked.

After the 60 one-course-study sources and records failing the structural rules had been excluded, the sampling frame contained 71,359 learner comments. The sample allocated 40 contributions to

each course and 10 to each equal-duration course–time quartile. Within each course–quartile cell, every nonempty R_{72} stratum first received one place. The remaining places were allocated by largest remainder in proportion to the finite population available in each stratum. The seed was `20260731-DCR-RQ1-FIRST-STAGE-V1`. A fixed SHA-256 ordering of de-identified contribution keys determined the without-replacement draw within each stratum.

Candidate source spans were split at sentence-ending punctuation and line breaks and required at least 12 Chinese characters. These deterministic rules produced 170 candidate spans. The source-screening rules were fixed before reviewers examined contribution content, and reviewers worked independently of the provisional internal screen. Independent human review classified 73 spans as ineligible, three as context dependent, 94 as eligible, and none as unresolved. The three context-dependent spans were joined to adjacent eligible spans during boundary adjudication. All eligible ideas in a sampled contribution entered the analysis, and contributions with no retained idea were not replaced. The final register contained 94 source ideas from 54 of the 120 contributions: 91 standalone ideas and three merged ideas. Figure S1 shows the analysis flow from sampling through source screening and human coding.

1.3 Coding and adjudication provenance

The formal relation-coding workbooks displayed a blind pair identifier and the two de-identified spans. They hid R_{72} status, retrieval channel, score, rank, sampling stratum, inclusion probability, delay, internal diagnostic judgments, and the other coder's labels. Candidate order varied between coders. Before adjudication, each coder had completed 3,567 judgments. They agreed on 2,241 relation classes (62.83%; Cohen's $\kappa = .357$). Among the 1,326 class disagreements, 1,230 concerned class 0 versus class 1, and 96 crossed the binary development boundary. The coders agreed on 3,471 binary judgments (97.31%; $\kappa = .576$). The binary table contained 69 both-positive, 69 coder-A-only positive, 27 coder-B-only positive, and 3,402 both-negative judgments. Positive specific agreement was 59.0%, and negative specific agreement was 98.6%.

Among the 2,241 class agreements, 69 relations were coded as substantive development by both coders but received different primary form labels. Complete agreement on relation class and, when positive, primary form therefore accounted for 2,172 labels. Three adjudication batches supplied the remaining 1,395 labels: 1,326 class disagreements plus 69 form disagreements. All final relation

labels were 0, 1, or 2; no relation retained `U`.

1.4 Strict-consensus label-sensitivity analysis

The conservative sensitivity set retained a relation only when coder A and coder B had both independently assigned label 2 before adjudication. This criterion was applied to the 57 $R_{72} = 0$ source ideas and was not treated as a superior gold standard. Across the full formal coding set, 69 relations met the criterion and all 69 received a final positive label; the $R_{72} = 0$ branch contained 26 such relations across 15 source ideas. Lexical Top-5 included 7/26 strict-consensus relations and surfaced 6/15 source ideas; the three-list union included 14/26 and surfaced 10/15; the expanded fused list included 17/26 and surfaced 12/15. The corresponding primary-label results were 14/36, 24/36, and 30/36. The analysis therefore tests sensitivity to label strictness while preserving the distinction between the magnitude of the positive set and the relative visibility pattern.

1.5 Conditional random-ranking baseline

For the 33 positive source ideas in the bounded one-course set, the observed numbers of positive candidates were $k = 1$ for 20 sources, $k = 2$ for 10, $k = 3$ for 2, $k = 4$ for 1, and $k = 5$ for none. Under the conditional random-ranking null, the k_i positive positions were uniformly permuted within each source's five fixed candidates. The Top-1 and Top-3 hit probabilities were $p_1(k_i) = k_i/5$ and $p_3(k_i) = 1 - \binom{5-k_i}{3}/\binom{5}{3}$, respectively. Exact Poisson–binomial dynamic programming gave a Top-1 expectation of 10.0/33 and $P(T \geq 10) = .5702$, and a Top-3 expectation of 24.0/33 and $P(T \geq 25) = .4250$. These probabilities concern internal ordering within the fixed presentation sets only.

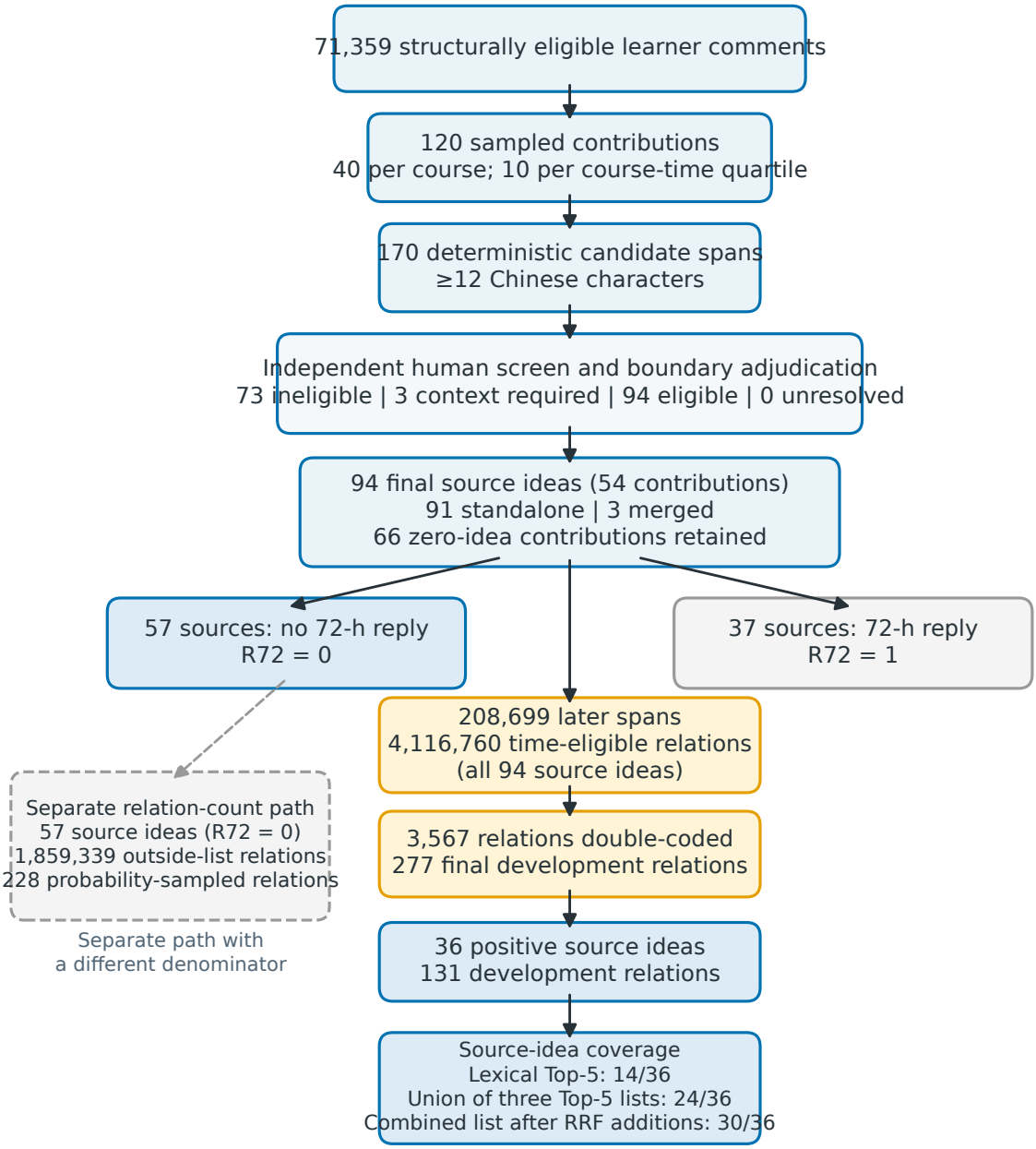


Figure S1: Analysis flow from the 2024 structural register to the human-coded set. Across all 94 source ideas, the search frame contained 208,699 later spans and 4,116,760 time-eligible source-to-span pairs; search and audit procedures selected 3,567 relations for formal human coding. The outside-list sensitivity analysis follows a separate path for the 57 ideas with $R_{72} = 0$ and contains 1,859,339 relations.

The constructed $R_{72} = 0$ branch contained 1,894 relations and 131 positives across 36/57 ideas; the $R_{72} = 1$ branch contained 1,673 relations and 146 positives across 27/37 ideas. Unequal candidate exposure makes these branch descriptions rather than estimates of a response-state effect.

A transcription check compared 54 source-evidence entries with the retained source record. Fifty matched, and four were corrected before the human-coded set was fixed. This check documented the match between the coded evidence and the retained source record.

Table S2 shows how the one-course development forms map to the 2024 relation codes. In the 2024 guide, `ELABORATE` includes explanatory or mechanistic elaboration. It also includes added conditions, causal chains, actionable processes, and new reasoning steps. The manuscript reports this category as explanatory or mechanistic elaboration while retaining the original code and denominator.

Table S2: Crosswalk between one-course development forms and 2024 type codes

One-course form	2024 code	Common meaning	Scope of the 2024 code
Explanation or mechanism	`ELABORATE`	Explanatory or mechanistic elaboration	Adds explanation, mechanism, condition, causal chain, actionable process, or a new reasoning step
New evidence	`EVIDENCE`	Evidential development	Adds data, a testable observation, a concrete case, or a counterexample that bears on the source proposition
Substantive revision	`REVISE`	Revision	Corrects, qualifies, challenges, or states a boundary condition for the source proposition
Synthesis	`SYNTHESIZE`	Synthesis	Organizes the source proposition with a distinct idea or body of evidence into a more complete account
Transfer	`TRANSFER`	Transfer	Preserves the source relation while applying it to a new field, population, tool, activity, or practical setting

Note. The two coding schemes are conceptually related but not measurement-equivalent. The crosswalk supports interpretive transparency while each coding guide and its type counts retain their original denominator; it is not a direct frequency-comparison scale. The value of zero for `SYNTHESIZE` indicates that no synthesis relation was observed in the current coded sets.

Tables S3 and S4 report the two coding guides in full. They remain separate because source-

span eligibility and later relation development are different judgment tasks even though both use numeric codes.

Table S3: Operational codebook for candidate source-span screening

Code	Name	Operational decision
2	Eligible	Contains a proposition, explanation, mechanism, evidence, substantive question, revision, synthesis, transfer, or reasoned action suggestion that can be judged independently.
1	Context required	May contain substantive content but must be joined to an adjacent span or reviewed with its immediate context before the idea can be judged reliably.
0	Ineligible	Greeting, thanks, attendance marker, activity arrangement, pure agreement, repetition without new content, fragment that does not form a proposition, or personal-identity statement without a substantive claim.
`U`	Unresolved	De-identification leaves too little information for a reliable decision.

Note. Candidate spans were split at sentence-ending punctuation and line breaks and required at least 12 Chinese characters. Adjacent dependent spans could be merged. In the final human freeze, the 170 candidate spans received 73 code-0, three code-1, 94 code-2, and zero `U` decisions. The three context-dependent spans were incorporated into merged ideas, yielding 91 standalone ideas and three merged ideas.

Table S4: Operational codebook for source-idea–to–later-span relations

Code or form	Operational decision
0: Unrelated	Shares only a broad topic, keyword, or setting, or follows a different object and reasoning path; the later span does not continue the source proposition.
1: Related nondevelopment	Continues the source proposition but only repeats, paraphrases, agrees, or lists similar content without adding a new reasoning step.
2: Substantive development	Continues the source proposition and adds explanation or mechanism, evidence, revision, synthesis, or a transfer that preserves the reasoning relation.
`U`: Unable to judge	De-identification, missing context, or damaged text prevents a reliable decision; this code is not used for uncertainty between 1 and 2.
`ELABORATE`	Adds an explanation, mechanism, condition, causal chain, actionable process, or new reasoning step. Greater length or more items of the same kind is not sufficient.
`EVIDENCE`	Adds data, a testable observation, a concrete instance, or a counterexample that supports, tests, or challenges the source proposition.
`REVISE`	Corrects, limits, challenges, or identifies a condition or boundary for the source proposition.
`SYNTHESIZE`	Organizes the source proposition with a distinct idea or body of evidence into a more complete framework; simple juxtaposition is not sufficient.
`TRANSFER`	Preserves the source reasoning relation while applying it to a different discipline, population, tool, activity, or practical setting.

Note. A code-2 relation received one primary form and could receive a secondary form. Neither completed formal relation analysis retained a final `U` label. No forum passage is reproduced in this codebook.

1.6 Retrieval implementation

The lexical model used a character analyzer with 2–5-grams, `min_df=2`, sublinear term frequency, and L2 normalization. It was fitted jointly to the 94 fixed source-idea texts and the complete later-span corpus. Candidates were scored with cosine similarity.

The local-semantic model used the simplified-Chinese sentence embedding in macOS Natural Language. Participant text remained on the research device; vectors were computed on device, and no forum text was sent to an external embedding service. Valid 640-dimensional vectors were L2-normalized before cosine comparison. An unavailable source or later-span vector was recorded as missing for this channel. Within each channel, candidates were ordered by descending score and then by ascending de-identified unique key. Non-finite semantic scores followed finite scores and contributed nothing to rank fusion.

Concept expansion extended the lexical representation with terms derived from local Chinese tokenization. We removed punctuation, numeric-only tokens, functional words, and terms found in more than 20% of later spans. Each later span contributed no more than 30 retained terms with the highest inverse-document-frequency values. A source term required a document frequency of at least five. A term pair had to occur together in at least three later spans. Positive normalized pointwise mutual information selected up to three neighbors for each source term. Original terms received weight 1, and expansion terms received weight 0.5. Inverse-document-frequency weighting and candidate-norm normalization were then applied.

For the main union, each search contributed its prespecified Top-5 list. The common depth was fixed before relation labels were available. Reciprocal rank fusion (RRF) used

$$\mathrm{RRF}(j) = \sum_{c} \frac{1}{60 + \mathrm{rank}_{cj}} \tag{S1}$$

and then added previously unselected relations from its Top-10 list. This step enlarged the joint list beyond the three five-candidate lists. Retrieval channel, score, and rank were generated before the human labels were summarized.

1.7 Expansion of the human-coded set

The fixed expansion seed was `20260731-DCR-DEEP-ANCHOR-V1`. Before inspecting relation text or scores, we selected one contribution from each of 12 course–time-quartile cells. Selection used an alternating R_{72} target and a fallback when the target stratum was empty. All eligible source ideas in the selected contributions entered the expansion search. For each idea, the search retained each channel's Top-50, the RRF Top-100, and one candidate outside the combined list from every nonempty consecutive seven-day block. This procedure enlarged the human-coded set but did not receive a probability weight.

1.8 Sampling outside the pool and uncertainty

Relations outside the combined list were divided into strata by course, delay band, maximum channel-score percentile band, and dominant channel. The score bands were P99–100, P95–99, P80–95, and P00–80. From each nonempty stratum h, we sampled $n_h = \min(4, N_h)$ relations without replacement. The inclusion probability was $\pi_h = n_h/N_h$. Relations were labeled only after the sample was fixed.

Let $Y_k = 1$ denote a human-coded positive relation in the sample outside the pool. The positive total outside the combined pool was estimated by

$$\widehat{T}_{\text{out}} = \sum_h \sum_{k \in s_h} \frac{Y_k}{\pi_h}. \tag{S2}$$

The estimated positive-pair density outside the pool was this estimated total divided by the known number of structural relations outside the pool, N_{out}. This is a relation-level quantity for the outside-pool pair denominator, not an idea-level incidence or prediction measure. Design-weighted pair-enumeration coverage was

$$\widehat{C}_{\text{pair}} = \frac{T_{\text{main}}}{T_{\text{main}} + \widehat{T}_{\text{out}}}, \tag{S3}$$

where T_{main} is the complete count of positive relations in the combined pool. Intervals used 20,000 stratified, finite-population-corrected bootstrap replicates with random seed `20260802`. The analysis ran with Python 3.12.13, openpyxl 3.1.5, and matplotlib 3.11.0 on macOS 15.7.7 (ARM64).

1.9 Supporting summaries for the main-text evidence hierarchy

The main manuscript foregrounds the timing, branching, and source-level visibility results. Tables S5–S7 retain the supporting coding, lexical, form-specific, and relation-level quantities under their original units and denominators. Figure S2 gives two bounded review-depth descriptions: an internal curve for the already assembled one-course lexical presentations and an equal-depth comparison of the three separate Top-5 views in the 2024 audit.

Table S5: Supporting coding and lexical diagnostics for the bounded one-course fixed candidate set

Diagnostic	Result and unit
Three-class coder agreement	300 relations; 265/300 (88.3%); Cohen's $\kappa = .776$
Development vs. nondevelopment agreement	300 relations; 275/300 (91.7%); Cohen's $\kappa = .706$
Final binary development decisions	300 relations; 50 developments across 33/60 source ideas
Development forms	50 positive relations; explanation/mechanism 19; transfer 15; new evidence 14; substantive revision 2; synthesis 0
Resolved three-class distribution	290 relations; development 50; related nondevelopment 190; unrelated 50
Median term frequency–inverse document frequency (TF–IDF) similarity	290 relations; development .0836; related nondevelopment .1026; unrelated .0759
TF–IDF discrimination	50 developments vs. 190 related nondevelopments; area under the receiver operating characteristic curve (AUC) .391; source-cluster bootstrap 95% confidence interval (CI) [.312, .472]
Highest-similarity quartile	73 relations; development 7; related nondevelopment 61; unrelated 5
Positive-count distribution	33 positive source ideas; $k = 1$: 20; $k = 2$: 10; $k = 3$: 2; $k = 4$: 1; $k = 5$: 0
Top-1 ordering	Observed 10/33; conditional random expectation 10.0/33; exact $P(T \geq 10) = .5702$
Top-3 ordering	Observed 25/33; conditional random expectation 24.0/33; exact $P(T \geq 25) = .4250$

Note. All presented relations received a final binary development decision. Ten nonpositive relations retained an unresolved unrelated-versus-related-nondevelopment distinction, leaving $n = 290$ for three-class analyses. The AUC used 5,000 source-cluster bootstrap resamples. The conditional random-ranking calculation uniformly permuted each source's observed positive positions within its five fixed candidates and used exact Poisson–binomial dynamic programming. These are diagnostics inside an already assembled presentation set.

Table S6: Relation-level delay by epistemic form in the bounded one-course fixed candidate set

Epistemic form	n	Median days [IQR]	By day 14	Days 15–30	After day 30
All observed forms	50	33.69 [13.87, 42.87]	13	5	32
Explanation or mechanism	19	32.67 [16.35, 36.87]	4	4	11
Transfer	15	40.84 [30.19, 44.76]	3	1	11
New evidence	14	31.44 [8.23, 34.07]	6	0	8
Substantive revision	2	39.62 [35.19, 44.05]	0	0	2
Synthesis	0	—	0	0	0

Note. Delay is measured from the end of the 72-hour reply window to the later contribution. IQR denotes interquartile range. Values describe positive relations found inside the fixed 60-day candidate window. No between-form test was conducted; the row with two revisions and the zero synthesis row are descriptive counts, not grounds for ordering the forms.

Table S7: Supporting label-sensitivity and relation-level visibility diagnostics for the 2024 audit

Diagnostic	Result and unit
Strict-consensus positive set	26 relations across 15/57 $R_{72} = 0$ source ideas
Strict-consensus relation sequence	Lexical Top-5 7/26; three-view union 14/26; expanded fused list 17/26
Strict-consensus source sequence	Lexical Top-5 6/15; three-view union 10/15; expanded fused list 12/15
Adjudicated relation visibility by view	Lexical Top-5 25/131; on-device semantic Top-5 16/131; related-term Top-5 14/131; three-view union 52/131; expanded fused list 72/131
Expanded-list explanation/mechanism relations	52/91
Expanded-list transfer relations	12/16
Expanded-list new-evidence relations	2/4
Expanded-list substantive-revision relations	6/20
Expanded-list synthesis relations	0/0; no synthesis relation in the human-coded set

Note. Strict consensus required both independent pre-adjudication labels to be 2 and is a conservative label-sensitivity set. The 131-relation denominator counts adjudicated positive source-idea–later-span relations for the 57 $R_{72} = 0$ source ideas. Relation-level visibility and source-level visibility answer different design questions. Form-specific rows apply only to the complete expanded list; unavailable channel-by-form joint cells were not reconstructed.

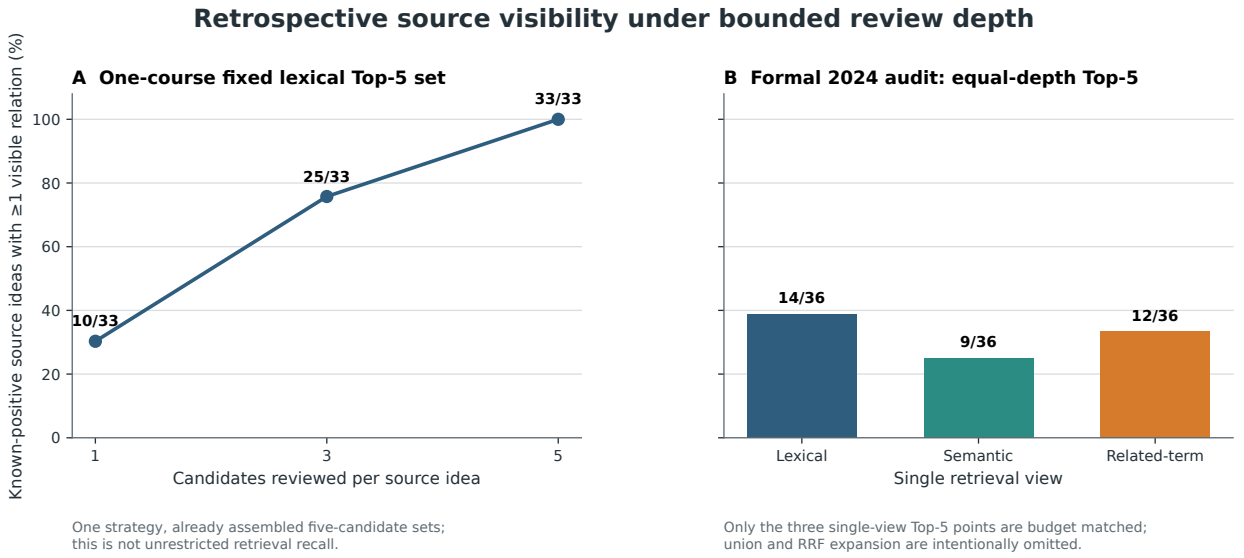


Figure S2: Retrospective source visibility under bounded review depth. Panel A describes the already assembled one-course lexical Top-5 presentations: at depths 1, 3, and 5, at least one confirmed relation was visible for 10/33, 25/33, and 33/33 positive source ideas. Panel B compares the three individually budget-matched Top-5 views in the fixed 2024 reference set: lexical 14/36, on-device semantic 9/36, and related-term 12/36. The three-view union and expanded fused list are excluded from Panel B because their unique review costs are not available in the safe aggregate. Neither panel estimates unrestricted recall or prospective performance.