Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

When Threads Fall Quiet: Tracing Idea Histories Across cMOOC Discussions

Yulin Tian

congqi79010@163.com

Beijing Normal University

Research Article

Keywords:

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10695452/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

When Threads Fall Quiet:

Tracing Idea Histories Across cMOOC Discussions

Abstract

A forum thread can fall silent while an idea articulated within it remains open to development. Knowledge Building treats ideas as improvable public objects whose histories may extend beyond their source threads. We examined whether source ideas whose originating contributions received no qualifying cross-author reply within an inclusive 72-hour window were later developed by another participant in an independent thread of the same course. The study combined a longitudinal analysis in one course with a stratified human review across three 2024 connectivist massive open online courses (cMOOCs). The broader review drew on a frame of 71,359 learner comments and assessed 3,567 possible cross-thread relations. In the one-course analysis, confirmed developments were often first observed weeks after the early reply window had closed, and some idea histories branched across multiple later authors and independent threads. Within the broader review's fixed set of source ideas with human-confirmed development, combining complementary archive views made more idea histories available for human review than lexical matching alone. Together, the findings show that local reply activity can end while an idea continues and branches elsewhere in the community. We call the corresponding design principle *connective visibility*: organizing persistent archives around longer idea histories so that people can compare dispersed developments and judge whether they advance the idea.

Keywords: cross-thread idea development; branching idea histories; knowledge building; cMOOCs; asynchronous discussion; connective visibility

1 Introduction

Online forums make collaboration visible through threads and replies. When replies stop, a thread appears locally complete, and older exchanges can recede behind newer activity (Hewitt, 2005;

Weng et al., 2012). Yet the idea expressed in that thread may not have reached its endpoint. The archive may later contain another participant's explanation, revision, evidence, or application of the same recognizable proposition in a different discussion. Knowledge Building treats ideas as public objects that communities can revisit, connect, and improve over time (Scardamalia and Bereiter, 2006). A thread therefore organizes interaction, but it does not necessarily bound the collective life of an idea.

This creates a mismatch between what the platform represents and what collaborative knowledge development asks us to follow. Threads and reply links record who responded, where, and when, making the local interaction history directly inspectable. The educational question is different: what does later discourse add to the proposition? When related developments are stored in separate threads, the archive does not readily present them as parts of the same developing idea. The platform's interaction structure can thus be mistaken for the boundary of the knowledge object. This mismatch is especially consequential in cMOOCs, where a connectivist design distributes participation across people, resources, times, and online spaces (Siemens, 2005; Wang et al., 2018).

Research on temporal and distributed interaction provides one part of the account. Uptake analysis follows how an earlier contribution becomes relevant in later activity (Suthers et al., 2010), while event-centred approaches preserve the order through which collaborative processes unfold (Reimann, 2009). Studies of mediated dialogue also show how collaborative work can connect activity across times, spaces, and social configurations (Beal and Steier, 2024). Together, these approaches make relations and sequence analytically visible. What remains unresolved is whether a proposition whose source thread received no early reply later acquires substantive development elsewhere in an ordinary course archive.

Knowledge Building research provides a second part of the account by treating ideas as objects for sustained collective work. Promising-idea judgments help a community decide which ideas deserve further investment (Chen et al., 2015), while Idea Thread Mapper and cross-community meta-spaces reconnect contributions dispersed across groups and discussions (Zhang et al., 2020; Yuan et al., 2022). More recently, knowledge-synthesis work has shown how new artifacts can gather and reorganize earlier contributions to support continued discourse (Zhu et al., 2026). These studies demonstrate the value of deliberately creating spaces in which distributed ideas can be revisited. Our study turns to a complementary problem: reconstructing later developments that already exist

across the archive's native threads after the source proposition received no early cross-author reply.

To address this problem, we distinguish an interaction history from an idea history. The interaction history records the visible reply state of the source thread. The idea history follows what later discourse adds to a recognizable source proposition, even when the later contribution appears in another thread. We define later cross-thread idea development as a contribution that adds explanation or mechanism, new evidence, substantive revision, synthesis, or transfer in an independent thread of the same course after the source contribution received no direct reply from another participant during an inclusive 72-hour window. An idea history may contain a single later continuation, or it may branch when the source proposition connects to later developments by different authors in different threads. This proposition-centred view makes it possible to examine both how long an idea remains open to development and whether that development becomes distributed across the community.

These distinctions lead to two linked questions. We ask first whether early-unanswered ideas receive later cross-thread development and how its timing, form, lexical position, and branching structure can be characterized. We then ask how complementary archive representations change which confirmed developments become available for human review, including under a stricter coding rule. The second question turns the empirical phenomenon into a design problem. We use *connective visibility* to describe the capacity of an archive representation to bring possible cross-thread relations into human comparison while leaving the judgment of idea development to people.

To answer these questions, we analyzed archived learner discussions from three 2024 cMOOC offerings. A one-course longitudinal analysis followed early-unanswered source ideas, and a broader human-coded review examined how complementary representations changed which idea histories entered human review across all three courses. Within the 60-day observation window, development was first observed weeks after the 72-hour reply window closed and across several locations in the community record. Of the 33 source ideas with confirmed later development, 13 had confirmed later relations involving at least two distinct later authors across at least two independent threads. The broader review showed that complementary representations brought more of these idea histories into view than lexical matching alone. Together, the findings distinguish interaction history from idea history, document branching as a form of collective idea development, and motivate *connective visibility* as a design principle for persistent collaborative archives.

2 Theoretical Framework and Related Work

2.1 Ideas as public objects of development

Knowledge Building treats ideas as public, improvable objects rather than as contributions whose significance ends when they are posted or answered (Scardamalia and Bereiter, 2006). This view shifts attention from message persistence to development of the idea it makes available. A recognizable proposition can be revisited after its original exchange and advanced when later discourse adds an explanation or mechanism, new evidence, a substantive revision, a synthesis, or transfer to a new context. Continuity does not require repeated wording; enough propositional continuity must remain for a human coder to specify the addition. These forms do not imply that every addition constitutes knowledge creation. The distinction matters because sharing information, constructing understanding, and creating communal knowledge involve different qualities of discourse (van Aalst, 2009). Repetition, agreement, and topical continuation may sustain participation without changing what a proposition can explain or support. Related distinctions appear in analyses of elaboration, integration, qualification, and dialogic movement in asynchronous discourse (Fu et al., 2016; Xiao, 2026). Work on promising ideas further shows why communities keep selected ideas available for continuing inquiry rather than treating initial uptake as their endpoint (Chen et al., 2015; Lee and Tan, 2017). In this study, the epistemic distinction provides the coding criterion: what the later contribution adds to the source proposition.

2.2 From interaction history to idea history

Knowledge Building gives a reason to follow an idea beyond its posting event, whereas a persistent forum makes collaborative activity visible through replies in local threads. These structures support related but nonidentical histories. Explicit reply links organize an interaction history: who replied, where, and when. An idea history is a source-anchored reconstruction of how later discourse substantively develops a recognizable proposition expressed in a source contribution. Because that discourse may occur outside the source thread, the two histories need not share temporal or structural boundaries. Event-centred analysis makes this distinction observable by treating order as part of the phenomenon rather than as background metadata (Reimann, 2009). The source event and its early reply state must be fixed before eligible later events are considered;

otherwise an eventual continuation can be mistaken for an immediate response. In this study, the inclusive 72-hour interval seals that early reply state; it does not define how long an idea remains active. Distributed-interaction research likewise cautions against equating one technical link structure with collaborative activity across actors, settings, times, and media (Suthers et al., 2010). A confirmed source–continuation relation records a content relation in this ordered archive without establishing encounter or influence. This distinction makes it possible to ask when, in what form, and through what structure proposition development appears after the local interaction history has been established.

2.3 Distributed and branching idea work

Once idea history is separated from interaction history, continuity need not be confined to one conversational route. Collaborative work can reconnect with a proposition through a different participant, discussion, or moment, and these later additions may follow more than one path. Research on opportunistic collaboration shows that productive knowledge work can be organized through relations that are not limited to a fixed group arrangement (Siqin et al., 2015); cross-community Knowledge Building similarly demonstrates how ideas can become objects of work across initially separate social spaces (Yuan et al., 2022). Other studies show participants taking up shared material across time and space or entering knowledge work through emerging opportunities rather than through a single planned sequence (Beal and Steier, 2024; Chen et al., 2026). In a source-anchored reconstruction, an idea history may follow a single observed route or branch across several routes. It branches when at least two confirmed later relations involve at least two distinct later authors and at least two distinct later threads, each independent of the source thread. This definition describes the source-centred structural multiplicity of later additions; it does not identify a transmission process or causal pathway. Branching is therefore a structural property of an idea history: it asks not only whether an idea was developed later, but whether its development was distributed across distinct contributors and discussion locations.

2.4 Representational visibility and human judgment

The separation between interaction and idea histories creates a representational problem. Native thread views make local response activity easy to follow, but they do not ordinarily gather later

contributions around the source proposition those contributions develop. As new replies accumulate, active discussions become more prominent and older threads recede from collective attention (Hewitt, 2005). This tendency is consequential in a cMOOC, where people, resources, and contributions are deliberately distributed across a network rather than organized within a stable cohort or a single discussion path (Siemens, 2005; Wang et al., 2018). The archive can thus preserve a source contribution while representing its later developments in separate interaction containers. This mismatch concerns how the archive organizes relations rather than what participants noticed, read, or valued. A thread-centred view answers where local interaction occurred. It cannot by itself show where the source proposition was subsequently elaborated, revised, supported, synthesized, or transferred. For an idea history to become inspectable, candidate continuities distributed across threads must be brought into a common field of comparison.

Knowledge Building analytics provide design precedents for this reorganization. Principle-based analytic tools have been used to support sustained idea improvement (Hong et al., 2015), while Idea Thread Mapper represents the collective course of inquiry through connected idea episodes rather than isolated postings (Zhang et al., 2018). Cross-community spaces extend that representational move across initially separate communities (Zhang et al., 2020; Yuan et al., 2022), and synthesis tools create additional occasions for participants to compare and reorganize dispersed contributions (Zhu et al., 2026). These designs motivate *connective visibility*, a representational property concerning whether possible source–continuation relations are made available at the source level for human comparison. Connective visibility concerns availability for judgment, not a score of idea quality, participant uptake, or learning. Retrieval can assemble plausible candidates, and complementary views can expose different portions of a distributed record, but a candidate becomes a confirmed development only when people compare its content with the source proposition and its original context. This distinction also applies to graph and visual analytics: reorganizing collaborative traces can support interpretation, but it does not automate the educational judgment those traces require (Saqr and López-Pernas, 2021; Tong et al., 2024; van Leeuwen et al., 2019).

Prior work therefore establishes that ideas remain improvable, collaboration extends across distributed times and spaces, and representations reorganize dispersed traces for inquiry. These foundations expose a specific analytic problem: threaded environments organize observable activity through local reply structures, while a recognizable proposition may develop beyond them. An

interaction history and its idea history therefore need not share an endpoint. If interaction and idea histories can diverge, the first analytic task is to reconstruct and characterize the latter beyond the source thread. If such histories are distributed across native threads, a second task is to examine how archive representations make their possible developmental relations available for human comparison. These two analytic tasks organize the study. Knowledge Building supplies the epistemic criterion for substantive development; event order separates the sealed early state from eligible later contributions; source-centred reconstruction describes whether development follows one route or branches; and connective visibility describes how representations make relations available for judgment. RQ1 concerns the occurrence, duration, forms, ordering, and branching structure of reconstructed idea histories. RQ2 concerns how representations change source-level availability within a fixed human-coded reference set. Together, the questions shift the analytic object from thread-local reply records to source-anchored idea histories while leaving epistemic judgment with people.

3 Research Questions

Beyond the early reply state, proposition development raises two connected questions about the histories that can be reconstructed and the representations that make them visible.

RQ1. Within the bounded one-course fixed candidate set, what share of source ideas with $R_{72} = 0$ showed confirmed later cross-thread development, when and in what epistemic forms did it appear, and to what extent did the resulting idea histories branch across distinct later authors and threads?

RQ2. Within the fixed 2024 audit of $R_{72} = 0$ source ideas, how did complementary archive representations change source-level visibility of confirmed later development?

4 Methods

4.1 Study context and linked analytical design

The study used archived learner-discussion records from three 2024 cMOOC offerings about digital education and teaching, conducted between 31 October and 27 December 2024. Discussion participation was required, and all discussions were conducted in Chinese. Platform records supplied

contribution timestamps, author roles, thread locations, and explicit reply links for ordering events. We analyzed de-identified records with locally processed text. No learner text is reproduced in the manuscript or Online Resource, and all reported findings are aggregate coded results.

Two linked analyses addressed different parts of the argument. The bounded one-course analysis reconstructed later idea histories within a fixed candidate set and described their occurrence, timing, development forms, lexical ordering, and branching. The broader 2024 audit asked which known-positive source ideas became available for human review under complementary archive representations. The first analysis therefore characterized observed development inside a bounded presentation set; the second compared source-level visibility inside a fixed human-coded reference set. Both analyses established the source's early reply state at a prespecified 72-hour cutoff before considering later contributions, and both reserved the substantive-development decision for human coders. Table 1 makes the two inferential targets and their supporting analyses explicit.

Table 1: Linked analytical designs, units, and inference boundaries.

Analysis / RQ	Frame and source construction	Candidate generation and human review	Unit / output	Inference boundary
Bounded one-course / RQ1	Deterministic 60-source set; $R_{72} = 0$; fixed 60-day observation window	Lexical Top-5 per source; 300 fixed relations; two independent coders and adjudication	Source histories and confirmed relations; occurrence, timing, form, ordering, branching	Observation inside a nonprobability presentation set; no prevalence, prediction, exposure, or influence claim
2024 visibility audit / RQ2	Stratified 120-contribution sample; 94 retained source ideas; primary scope of 57 with $R_{72} = 0$	Three Top-5 views, union, and expanded fused list; 3,567 deduplicated relations human-coded before summaries	Source-level availability in a fixed 36-idea positive reference set	Retrospective source-level coverage under unequal review budgets; retrieval organizes human review, not learning or participation outcomes
Supporting analyses / both	Same frozen sources, candidates, and labels as the linked analyses	Strict pre-adjudication consensus, relation enumeration, lexical ordering, and stratified outside-list audit	Label sensitivity, relation-level visibility, diagnostic ordering, and coverage estimates	Supporting interpretation only; each analysis retains its own denominator and estimand

Note. $R_{72} = 0$ indicates no qualifying cross-author reply in the inclusive 72-hour interval. Retrieval assembled candidates for comparison; human coding determined whether a later contribution substantively developed the source proposition. Full sampling, coding, retrieval, and sensitivity procedures are reported in Online Resource 1.

4.2 Temporal eligibility, analytic units, and human judgment

Human review identified a source idea in a focal contribution posted at time t_0. A qualifying early response was an explicit learner-to-learner reply from an author distinct from the originator. The inclusive interval counted that response when $0 \leq t - t_0 \leq 72$ hours; $R_{72} = 0$ denotes zero

qualifying cross-author replies in the interval. The 72-hour cutoff was fixed before later discourse was examined. An eligible later contribution came from the same course and another author, occupied a thread independent of the source thread, appeared strictly after 72 hours, and occurred on or before the course endpoint. Thus, the early reply state and later-development search were temporally ordered rather than defined from the same events.

The analysis kept three units distinct. A *source idea* was the focal conceptual unit extracted from a sampled contribution. A *candidate relation* paired one source idea with one temporally eligible later span. A *confirmed development relation* was a candidate relation judged to add substantive reasoning to the source proposition. A positive source idea had at least one such confirmed relation. Source-level summaries counted an idea once even when several confirmed relations were present; relation-level summaries counted the individual source-idea–later-span pairs. This separation allowed the study to describe whether an idea history was observed, when its first confirmed development was observed, what later relations added, and whether more than one developmental route was present without treating these quantities as interchangeable denominators.

The relation-coding instrument distinguished three levels of connection. *Unrelated* contributions contained only broad-topic, keyword, or setting overlap. *Related nondevelopment* contributions preserved the proposition at its current reasoning level. *Substantive development* extended the proposition through explanation or mechanism, new evidence, substantive revision, synthesis, or transfer. This Knowledge Building criterion required observable epistemic work; repetition, agreement, and topical continuation were insufficient. Coders evaluated the de-identified source and candidate spans in their supplied context. Retrieval channel, score, rank, sampling information, internal diagnostic judgments, and the other coder's label were hidden during formal coding. Independent labels were retained for agreement analyses, and disagreements relevant to the final relation set were resolved through relation-level adjudication. Online Resource 1, Tables S3 and S4, gives the source-span and relation codebooks.

For each positive source idea s in the bounded study, we counted confirmed relations (r_s), distinct later contributions (c_s), distinct later authors (a_s), and distinct later threads (h_s). The branching indicator was

$$B_s = \mathbf{1}(a_s \geq 2 \ \wedge \ h_s \geq 2). \tag{1}$$

Because both conditions require at least two confirmed relations, branching involved multiple later authors and independent later threads around the same source proposition. Branching was defined entirely from confirmed relations between the source proposition and its later developments, using multiplicity across later authors and threads within the candidate set. This source-centred structure treated the source proposition as the organizing unit for confirmed later developments.

4.3 Bounded one-course analysis

The bounded set was constructed deterministically from eligible learner comments ordered by posting time. We retained the first 60 source ideas after a one-source-per-thread cap and exact normalized-idea deduplication; their posting times ranged from 0.37 to 5.87 days after course start. All 60 met the structural rules, had $R_{72} = 0$, and remained observable throughout a prespecified 60-day window. For every source, character-based term frequency–inverse document frequency (TF–IDF) ranked contributions from other authors in independent threads. The five highest-ranked eligible candidates were fixed before relation coding, producing a nonprobability presentation set of 300 source-idea–later-span relations. RQ1 therefore reconstructs observed idea histories within the five prespecified candidates presented for each source, making the bounded presentation set the empirical frame for occurrence, timing, form, ordering, and branching.

Two coders independently judged all 300 relations. They agreed on 265/300 three-class labels (88.3%; Cohen's $\kappa = .776$) and on 275/300 development-versus-nondevelopment decisions (91.7%; $\kappa = .706$). Coder agreement or relation-level adjudication supplied the final development-versus-nondevelopment decisions. Three-class labels were retained only for relations with a resolved three-class assignment. RQ1 used source ideas as the denominator for observed occurrence, first-development timing, and branching, and confirmed relations as the denominator for development forms and relation-level delay. For a positive source, the first-development delay was the elapsed time from t_0 to its earliest confirmed later relation. Relation-level delay was measured from the end of the defined reply window to the later contribution. Prespecified timing summaries used days through 14, days 15–30, and after day 30. The lexical score was retained only as an ordering diagnostic inside the already assembled five-candidate sets. Online Resource 1, Tables S5 and S6, reports the resolved class distribution, form counts, delay by form, and complete lexical diagnostics.

4.4 2024 visibility audit

The broader audit began with a stratified sample of 120 contributions from 71,359 learner comments: 40 contributions per course and 10 in each of the 12 equal-duration course–time-quartile cells. Independent human screening segmented and evaluated the sampled contributions and retained 94 source ideas from 54 contributions. Of these ideas, 57 had $R_{72} = 0$ and formed the primary search scope; the remaining 37 supplied a descriptive branch. Because contributions were the sampling unit, the 57/94 split describes the retained source ideas as conceptual units within this audit sample and is reported descriptively at that level.

For each source idea, three representations searched the same eligible archive. Character n-gram TF–IDF supplied a lexical view, local related-term expansion supplied an expanded-term view, and an on-device simplified-Chinese model supplied a semantic-similarity view. Each produced a pre-specified Top-5 list before formal labels were summarized. The three-view union and a reciprocal rank fusion (RRF) expansion (Cormack et al., 2009) formed two larger review sets. The multi-view pool, preselected-source expansion, and stratified outside-list audit were deduplicated before coding, yielding 3,567 unique source-idea–later-span relations. All candidate generation and semantic computation were conducted locally; participant text remained on the research device.

Two coders independently evaluated all 3,567 de-identified relations under the blinding procedure described above. They agreed on 3,471/3,567 binary development judgments (97.31%; Cohen's $\kappa = .576$); complete three-class agreement and adjudication provenance are reported in Online Resource 1. Final labels came from coder agreement or relation-level adjudication, never from retrieval scores. The fixed reference set for RQ2 comprised the 36 of 57 $R_{72} = 0$ source ideas with at least one confirmed substantive-development relation anywhere in the human-coded set. Source-level visibility counted each of these 36 ideas once when a specified review set contained at least one confirmed relation. Source-level visibility therefore measured whether a known-positive idea history was available for human comparison under a given representation. The resulting estimand was retrospective source-level availability within this fixed human-coded reference set.

4.5 Scope, sensitivity, and provenance

The principal RQ2 comparison used final adjudicated labels. A conservative sensitivity analysis repeated the source-level sequence using only relations that both coders had independently assigned label 2 before adjudication. Relation enumeration answered a separate supporting question: how many of the 131 confirmed positive relations associated with the 57 $R_{72} = 0$ source ideas entered a review set. Relation enumeration used relation-level coverage as a separate supporting quantity while retaining the primary source-level visibility estimand. The stratified outside-list analysis estimated relation coverage beyond the assembled lists under its own sampling design. Together, the sensitivity, relation-level, review-budget, and coverage analyses provide diagnostic context for the primary source-level visibility estimand. Online Resource 1 reports them under their distinct denominators and inferential roles.

5 Results

5.1 Idea histories beyond the early reply window (RQ1)

Later cross-thread development was confirmed for 33 of the 60 source ideas with $R_{72} = 0$ in the bounded one-course fixed candidate set (55.0%). The 300 presented candidates contained 50 confirmed substantive-development relations across these 33 source ideas.

The first confirmed developments were observed well beyond the initial response window. Across the 33 positive source ideas, the median delay to first observed development was 30.57 days (interquartile range [IQR] 8.42–36.11). Eleven were first observed by day 14, five during days 15–30, and 17/33 were first observed after day 30 (Figure 1). The observed idea histories therefore extended over weeks within the fixed 60-day window.

Of the 33 positive source ideas, 20 had one confirmed later relation and 13 had between two and four. For all 13 sources with multiple confirmed relations, the later developments involved at least two distinct authors and at least two independent threads. Thus, 13/33 positive source ideas (39.4%) met the multi-author, multi-thread branching criterion (Figure 2). The bounded set therefore included idea histories that were structurally distributed as well as temporally extended. These timing, multiplicity, and branching results are summarized in Table 2.

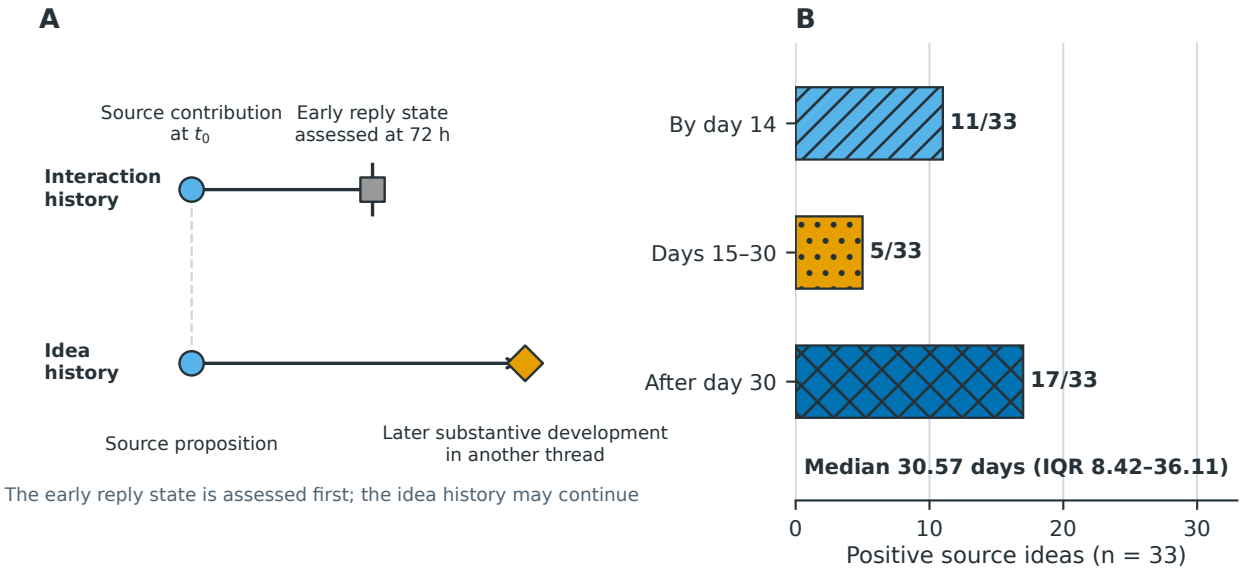


Figure 1: Two analytical records and the timing of confirmed development. Panel A distinguishes the early reply state assessed at 72 hours from the longer reconstructed idea history. The lanes and connecting lines encode analytical order; exposure and causal pathways fall outside the figure. Panel B places the 33 positive source ideas into three first-development bands: 11 by day 14, five during days 15–30, and 17 after day 30. The delay distribution describes positives found within the fixed 60-day search window.

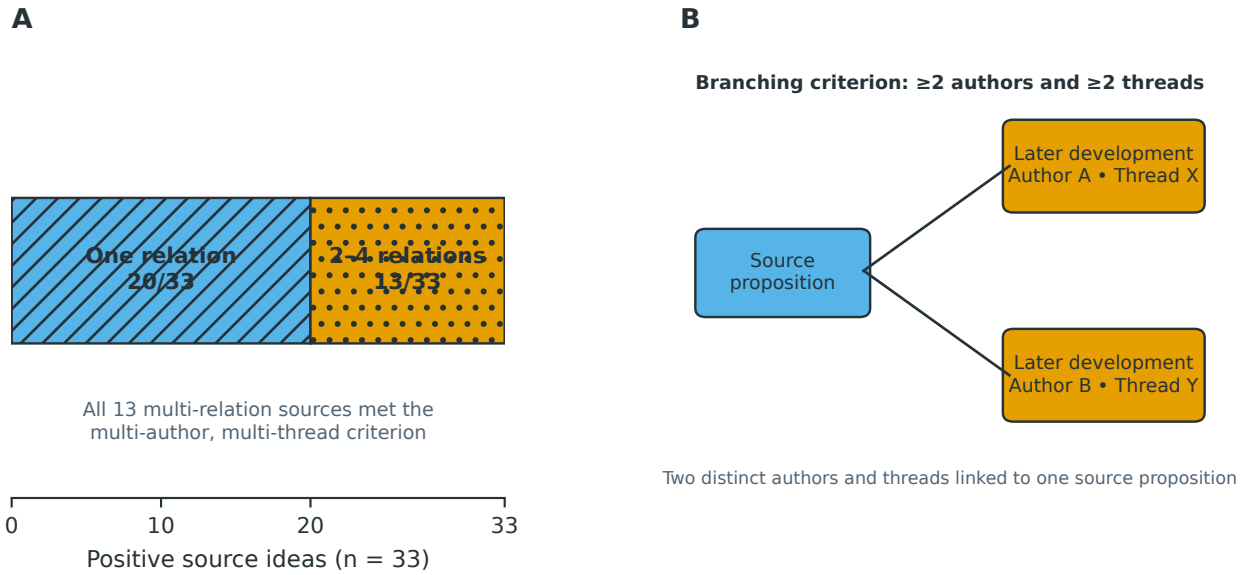


Figure 2: Multiplicity and the branching criterion in the bounded one-course study. Panel A divides the 33 positive source ideas into 20 with one confirmed later relation and 13 with two to four relations. Each of those 13 involved at least two distinct later authors across at least two independent threads. Panel B depicts the source-centred analytic relation defined in Methods. The counts describe the fixed Top-5 candidate set.

Table 2: Longer and branching idea histories in the bounded one-course study.

Observation	Result and unit
Fixed candidate set	60 source ideas with $R_{72} = 0$; 300 lexical Top-5 candidate relations
Confirmed later development	50 substantive-development relations across 33/60 source ideas
First confirmed development	33 positive source ideas; median 30.57 days; IQR 8.42–36.11; 17/33 after day 30
Post-window relation-level delay	50 positive relations; median 33.69 days; IQR 13.87–42.87
Source-level multiplicity	20/33 positive source ideas with one confirmed relation; 13/33 with two to four
Multi-author, multi-thread branching	13/33 positive source ideas; at least two distinct later authors and at least two distinct later threads
Maximum observed multiplicity	Four confirmed relations; four later contributions; three later authors; four later threads

Note. The one-course study used a nonprobability fixed Top-5 presentation set and a fixed 60-day observation window. Source-level first-development delay was measured from source posting; post-window relation-level delay was measured from the end of the inclusive reply window to the later contribution. Timing is conditional on positives found within the observation window. Branching and multiplicity describe confirmed source-to-later-contribution relations inside the five presented candidates; exhaustive trajectory census and relations among later contributors fall outside this table's scope. Coding agreement, epistemic forms, lexical diagnostics, and the conditional random-ranking calculation are reported in Online Resource 1, Tables S5 and S6.

Confirmed development took several epistemic forms: explanation or mechanism accounted for 19 of 50 positive relations, transfer for 15, new evidence for 14, and substantive revision for two. No confirmed relation was coded as synthesis. The 2024 audit used a related but separately specified coding guide; its form counts and the crosswalk between the two instruments are reported in Online Resource 1, Tables S2 and S7.

In this fixed presentation set, lexical similarity did not order substantive development ahead of topical continuation. When substantive development was contrasted with related nondevelopment, the area under the receiver operating characteristic curve (AUC) was .391 (source-cluster bootstrap 95% confidence interval [CI] [.312, .472]); lexical Top-1 and Top-3 included a positive for 10/33 and 25/33 source ideas, compared with conditional random expectations of 10.0/33 and 24.0/33. These diagnostics reinforce the distinction between candidate presentation and human judgment of what later discourse added to the proposition. Full counts and exact probabilities are reported in Online Resource 1, Table S5.

5.2 Connective visibility across archive representations (RQ2)

Among the 57 $R_{72} = 0$ source ideas in the 2024 audit, 36 had at least one confirmed later development in the adjudicated relation set. These 36 source ideas constituted the fixed known-positive reference set for the visibility comparison. Lexical Top-5 made a development relation available for 14/36 source ideas (38.9%). The union of lexical, on-device semantic, and related-term Top-5 lists increased source-level visibility to 24/36 (66.7%), and the expanded fused list increased it to 30/36 (83.3%; Figure 3). Complementary representations therefore brought at least one confirmed later development into human review for more source ideas in the fixed reference set.

The individual on-device semantic and related-term Top-5 lists made developments visible for 9/36 (25.0%) and 12/36 (33.3%) source ideas. Before joint-ranking expansion, six positive ideas were unique to lexical retrieval, five to on-device semantic retrieval, and five to related-term retrieval. Each Top-5 view surfaced positive source ideas absent from the other two views.

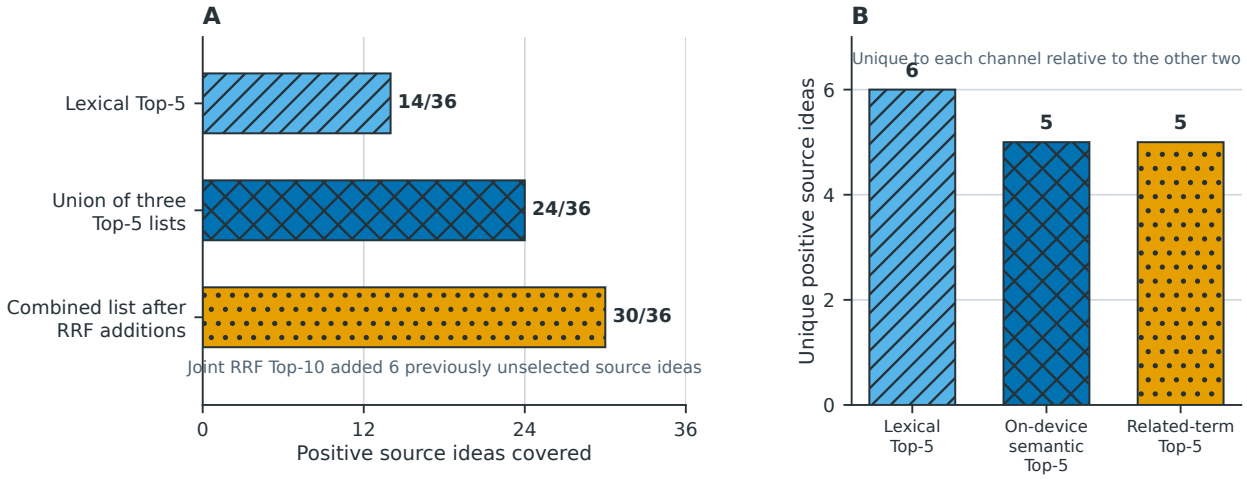


Figure 3: Connective visibility across the 2024 archive representations. Panel A shows source-level visibility within the fixed set of 36 source ideas with a confirmed positive relation: lexical Top-5 surfaced 14, the three-view Top-5 union surfaced 24, and the expanded fused list surfaced 30/36. Panel B shows positive ideas unique to one Top-5 view before enlargement: six lexical, five on-device semantic, and five related-term. Candidate presentation made relations available for comparison; human coding determined development.

Table 3: Source-level visibility of confirmed later development across the 2024 archive representations.

Positive-label set	Lexical Top-5	Three-view Top-5 union	Expanded fused list
Adjudicated final labels	14/36 (38.9%)	24/36 (66.7%)	30/36 (83.3%)
Strict two-coder consensus	6/15 (40.0%)	10/15 (66.7%)	12/15 (80.0%)

Note. The primary reference set comprised the 36 of 57 $R_{72} = 0$ source ideas with at least one final positive relation in the 3,567-relation human-coded set. A source was counted once when a review set contained at least one confirmed positive relation. Strict consensus required both coders to have independently assigned label 2 before adjudication; it is a conservative label-sensitivity set, while adjudicated labels remain primary. The union and expanded list contain different numbers of unique candidates per source and therefore represent unequal review budgets. Relation-level and form-specific results are reported in Online Resource 1, Tables S1 and S7.

A stricter pre-adjudication consensus rule reduced the positive set from 36/57 to 15/57 source ideas. Within this stricter set, lexical Top-5 surfaced 6/15 positive source ideas, the three-view union surfaced 10/15, and the expanded fused list surfaced 12/15. The corresponding primary adjudicated sequence was 14/36, 24/36, and 30/36. The size of the positive set therefore depended on label strictness, while the progression in source-level visibility remained similar (Table 3).

Source-level visibility counts whether at least one confirmed relation for a known-positive source idea entered review. Online Resource 1, Table S1, separately examines the more demanding relation-level enumeration task.

6 Discussion

6.1 The reply clock did not bound the proposition-development clock

The first finding is that the locally visible response period and the later development of a proposition followed different temporal courses. In the bounded one-course analysis, the median time to the first confirmed development observed in the fixed candidate set was 30.57 days, and for 17/33 positive source ideas the first confirmed development was observed more than 30 days after the source contribution was posted. Each source contribution had received no qualifying cross-author reply in the inclusive 72-hour interval. Yet later discourse by another author in an independent thread explained, supported, revised, or transferred its recognizable proposition. The early thread state therefore described the absence of a local reply, whereas the reconstructed idea history followed the proposition's later development in the observed course record.

This distinction changes how thread silence is interpreted. Research on why asynchronous threads die explains how local activity loses prominence as attention moves to other discussions (Hewitt, 2005). Reply-centred research can then ask which message characteristics are associated with receiving responses and sustaining an online discussion (Chen et al., 2020). Those questions remain important because replies organize participation in the visible interaction space. The present analysis follows the proposition beyond the sealed early reply state, asking whether later discourse elsewhere in the course record adds to it. Such development may appear in an independent thread and in different wording, so a quiet source thread can coexist with an extending idea history: one describes the local interaction container, while the other describes later epistemic work connected

to its content.

Event order makes the distinction analytically clear. Event-centred process analysis treats the sequence and eligibility of events as part of the phenomenon (Reimann, 2009). By determining the early reply state first and admitting only later, cross-thread contributions afterward, this temporal ordering keeps the early reply state and later development analytically distinct. It also extends temporal computer-supported collaborative learning (CSCL) work from the timing of participation and relational configurations to the timing of proposition development. Sequence analyses show that collaborative activity changes across ordered events (Saqr and López-Pernas, 2023), and temporal-network analyses show that relations can reorganize across longer intervals (Saqr et al., 2026). Here, the relevant relation is epistemic: what a later contribution adds to an earlier proposition. The result demonstrates that this proposition-level relation can become visible on a much longer horizon than the reply relation recorded by the source thread.

The educational implication concerns the object whose history is being traced. An interaction history organized by reply links is appropriate for questions about direct response, responsiveness, and the organization of a local exchange. An idea history is needed when the question is how later discourse in the persistent course archive develops a proposition. Specifying the object therefore changes what counts as continuity across the archive. Distributed-interaction theory already directs attention across actors, settings, times, and media (Suthers et al., 2010); the present finding identifies the proposition as the organizing object for that reconstruction. In a persistent learning environment, thread closure can mark a local interaction endpoint while the reconstructed idea history continues.

6.2 Some longer idea histories branched

The second finding adds structure to this longer horizon. Of the 33 source ideas with at least one confirmed later development, 20 had one confirmed relation in the fixed five-candidate presentation set, whereas 13 had two to four. Every one of those 13 involved at least two distinct later authors across at least two independent later threads and therefore met the source-centred branching criterion. For the sources with multiple confirmed relations, later development was distributed across separate contributors and discussion locations rather than contained in a single continuing conversational route.

From a Knowledge Building perspective, this structure makes visible a consequence of treating ideas as public, improvable objects (Scardamalia and Bereiter, 2006). An idea that becomes a communal object of work can be developed beyond the participant who first articulated it and the thread in which it appeared. A source-centred view can therefore place together two forms of continuity across threads: a singular continuation, in which one observed later route adds to the proposition, and a branching history, in which distinct routes add to the same source proposition. Each branch was required to satisfy the same epistemic criterion; branching then describes how confirmed developments are distributed across contributors and discussion locations.

This retrospective reconstruction complements work that deliberately builds infrastructures for ideas to travel and grow. Idea Thread Mapper organizes inquiry around idea trajectories rather than isolated posts (Zhang et al., 2018), and cross-community meta-spaces make locally developed ideas available for work across community boundaries (Yuan et al., 2022). Process-oriented research also shows how opportunities for collaboration can emerge as participants move through a knowledge-building environment (Chen et al., 2026). These studies examine designed arrangements through which participants can pursue connected knowledge work. This retrospective reconstruction shows why a related analytic layer is useful even when activity was originally recorded in ordinary forum threads: the native archive may already contain dispersed additions to the same source proposition, and a source-centred layer can represent their common source as the organizing relation.

The contribution of branching is both conceptual and descriptive: it gives idea history a source-centred structural form and distinguishes distributed development from simple longevity. Longevity records how far an observed history extends in time, whereas branching records whether confirmed development follows one route or several routes across discussion locations. The evidential claim is confined to coded archive relations: substantive additions recognizable as developments of the same source proposition appeared through distinct authors and independent threads. The archive therefore contained a distributed history of substantive additions to the idea across discussion locations. Questions of exposure, transmission, and causation remain separate from this reconstruction.

6.3 Complementary representations widened human comparison

The third finding addresses how such dispersed histories can become inspectable. Within the fixed set of 36 source ideas that had at least one human-confirmed later development in the 2024

audit, lexical Top-5 made a confirmed relation available for 14/36 ideas. Combining the lexical, semantic, and related-term Top-5 views raised source-level visibility to 24/36, and the expanded fused list raised it to 30/36. Each individual view contributed positive source ideas absent from the other two before expansion. The sequence therefore shows that complementary representations opened different portions of the known-positive reference set to review. Because the union and expanded lists required larger review budgets, this sequence is interpreted as retrospective source-level coverage under unequal review budgets.

The lexical ordering diagnostic showed weak alignment with substantive development inside the fixed presentation sets: AUC was .391, while Top-1 and Top-3 performance (10/33 and 25/33) was close to the corresponding conditional random expectations (10.0/33 and 24.0/33). Lexical overlap can find a plausible neighborhood in a large archive, but proposition development may preserve few surface terms while adding a mechanism, evidence, qualification, or transfer. Conversely, a lexically close contribution may restate the same reasoning. This weak alignment explains why widening the comparison field matters and clarifies the division of labor built into the study: computational views organize candidates; people decide what the later discourse adds.

Prior CSCL designs point to several ways of organizing that comparison. Alternative forum structures can support dialogue that is less fragmented by conventional threads (Smith and Sherry, 2025). Idea-thread and cross-community spaces can reorganize discourse around evolving problems and shared progress (Zhang et al., 2020; Yuan et al., 2022), while synthesis artifacts invite participants to compare and integrate dispersed contributions (Zhu et al., 2026). The present study extends this design logic to retrospective archives. A proposition-centred layer could sit above the native thread structure and gather candidate continuations around a source idea while preserving their authors, original threads, timing, and local context. The thread view would continue to show where interaction occurred; the added layer would show where the proposition may have been developed elsewhere.

Connective visibility refers here to representational availability for epistemic comparison. Graph and visual analytics can expose patterns that are difficult to see in a chronological interface (Saqr and López-Pernas, 2021; Tong et al., 2024), while dashboards can arrange collaborative information for interpretation (van Leeuwen et al., 2019). In the present study, archive representations bring candidate relations into view, and human comparison of the source proposition and its context

determines whether those relations constitute development. This division of labor preserves two distinct functions: representations organize dispersed candidates into a common comparison field, and human reviewers determine their epistemic relation to the source. Complementary views therefore widen the comparison field while retaining human epistemic judgment.

Taken together, the three findings supply one coherent account of persistent asynchronous collaboration. The 17/33 timing result shows that first development was measured from source posting and extended beyond the sealed early reply state. The 13/33 branching result shows that some of those longer histories were distributed across authors and independent threads. The 14/36 to 24/36 to 30/36 visibility sequence shows how complementary representations brought more of a fixed known-positive set into human comparison. Persistent archives therefore support two related but nonidentical records: threads preserve local interaction histories, while proposition-centred reconstruction can reveal longer and branching histories of idea development. A design for connective visibility could make those dispersed histories more inspectable while leaving epistemic judgment with people.

7 Conclusion

An early reply state cannot, by itself, define the temporal or structural boundaries of an idea's subsequent development in asynchronous collaborative learning. Among the positive source ideas, the first confirmed development was observed more than 30 days after source posting for 17/33, well beyond the early reply state assessed at 72 hours, and the confirmed later developments for 13/33 formed multi-author, multi-thread branches. Later substantive development was therefore observed well after the early response window and through independent discussion paths elsewhere in the course record. The interaction history of a source thread and the reconstructed idea history of its proposition are related, but they are not the same analytic object.

Archive representation also shapes which of these histories enter human review. In the fixed human-coded reference set, source-level visibility across the lexical Top-5, three-view union, and expanded fused review sets was 14/36, 24/36, and 30/36, respectively. Connective visibility therefore has a specific design role: retrieval can bring a wider range of candidate relations spanning authors, threads, and time into human comparison, while people determine what the later discourse

adds to the proposition. Thread views preserve local interaction context; proposition-centred views can reconnect developments dispersed across that context. Analyzing asynchronous collaboration therefore requires considering both what happened in a thread and what later discourse added to its proposition in the observed record.

Data Availability

The participant-level course discussion records are not publicly available because their textual content and relational structure create privacy and re-identification risks. Aggregate coded results and non-identifying methodological details are provided in Online Resource 1. Access to the participant-level records is restricted by the governing consent and data-protection conditions. The algorithmic code supporting the reported analyses and figures can be shared independently of those records and will be released in a public repository upon publication.

References

Beal, C., & Steier, R. (2024). Dialogues across time and space in a video-based collaborative learning environment. *International Journal of Computer-Supported Collaborative Learning, 19*(2), 157–186. https://doi.org/10.1007/s11412-024-09420-9

Chen, B., Scardamalia, M., & Bereiter, C. (2015). Advancing knowledge-building discourse through judgments of promising ideas. *International Journal of Computer-Supported Collaborative Learning, 10*(4), 345–366. https://doi.org/10.1007/s11412-015-9225-z

Chen, G., Lo, C. K., & Hu, L. (2020). Sustaining online academic discussions: Identifying the characteristics of messages that receive responses. *Computers & Education, 156*, Article 103938. https://doi.org/10.1016/j.compedu.2020.103938

Chen, Y., Yi, S., & Zhang, Y. (2026). How students engage in opportunistic collaboration: A process-oriented study in a knowledge-building community. *International Journal of Computer-Supported Collaborative Learning*. Advance online publication. https://doi.org/10.1007/s11412-026-09491-w

Cormack, G. V., Clarke, C. L. A., & Büttcher, S. (2009). Reciprocal rank fusion outperforms Condorcet and individual rank learning methods. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 758–759). Association for Computing Machinery. https://doi.org/10.1145/1571941.1572114

Fu, E. L. F., van Aalst, J., & Chan, C. K. K. (2016). Toward a classification of discourse patterns in asynchronous online discussions. *International Journal of Computer-Supported Collaborative Learning, 11*(4), 441–478. https://doi.org/10.1007/s11412-016-9245-3

Hewitt, J. (2005). Toward an understanding of how threads die in asynchronous computer conferences. *Journal of the Learning Sciences, 14*(4), 567–589. https://doi.org/10.1207/s15327809jls1404_4

Hong, H.-Y., Scardamalia, M., Messina, R., & Teo, C. L. (2015). Fostering sustained idea improvement with principle-based knowledge building analytic tools. *Computers & Education, 89*, 91–102. https://doi.org/10.1016/j.compedu.2015.08.012

Lee, A. V. Y., & Tan, S. C. (2017). Promising ideas for collective advancement of communal knowledge using temporal analytics and cluster analysis. *Journal of Learning Analytics, 4*(3), 76–101. https://doi.org/10.18608/jla.2017.43.5

Reimann, P. (2009). Time is precious: Variable- and event-centred approaches to process analysis in CSCL research. *International Journal of Computer-Supported Collaborative Learning, 4*(3), 239–257. https://doi.org/10.1007/s11412-009-9070-z

Saqr, M., & López-Pernas, S. (2021). Modelling diffusion in computer-supported collaborative learning: A large scale learning analytics study. *International Journal of Computer-Supported Collaborative Learning, 16*(4), 441–483. https://doi.org/10.1007/s11412-021-09356-4

Saqr, M., & López-Pernas, S. (2023). The temporal dynamics of online problem-based learning: Why and when sequence matters. *International Journal of Computer-Supported Collaborative Learning, 18*(1), 11–37. https://doi.org/10.1007/s11412-023-09385-1

Saqr, M., López-Pernas, S., & Törmänen, T. (2026). A temporal network approach to reveal the longitudinal dynamics of CSCL group regulation and productive collaboration. *International Journal of Computer-Supported Collaborative Learning, 21*(2), 237–270. https://doi.org/10.1007/s11412-025-09464-5

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 97–116). Cambridge University Press. https://doi.org/10.1017/CBO9780511816833.008

Siemens, G. (2005). Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning, 2*(1). itdl.org/Journal/Jan_05/article01.htm

Siqin, T., van Aalst, J., & Chu, S. K. W. (2015). Fixed group and opportunistic collaboration in a CSCL environment. *International Journal of Computer-Supported Collaborative Learning, 10*(2), 161–181. https://doi.org/10.1007/s11412-014-9206-7

Smith, G. G., & Sherry, M. B. (2025). Online discussion or authentic dialogue? How design affects discussions in two alternative types of online forums. *British Journal of Educational Technology, 56*(1), 103–127. https://doi.org/10.1111/bjet.13491

Suthers, D. D., Dwyer, N., Medina, R., & Vatrapu, R. (2010). A framework for conceptualizing, representing, and analyzing distributed interaction. *International Journal of Computer-Supported Collaborative Learning, 5*(1), 5–42. https://doi.org/10.1007/s11412-009-9081-9

Tong, Y., Yang, C., & Chen, G. (2024). A visual learning analytics approach for knowledge building: Impact on students' epistemic understanding of discourse, productive inquiry and domain knowledge. *British Journal of Educational Technology, 55*(3), 992–1019. https://doi.org/10.1111/bjet.13409

van Aalst, J. (2009). Distinguishing knowledge-sharing, knowledge-construction, and knowledge-creation discourses. *International Journal of Computer-Supported Collaborative Learning, 4*(3), 259–287. https://doi.org/10.1007/s11412-009-9069-5

van Leeuwen, A., Rummel, N., & van Gog, T. (2019). What information should CSCL teacher dashboards provide to help teachers interpret CSCL situations? *International Journal of Computer-Supported Collaborative Learning, 14*(3), 261–289. https://doi.org/10.1007/s11412-019-09299-x

Wang, Z., Anderson, T., & Chen, L. (2018). How learners participate in connectivist learning: An analysis of the interaction traces from a cMOOC. *The International Review of Research in Open and Distributed Learning, 19*(1). https://doi.org/10.19173/irrodl.v19i1.3269

Weng, L., Flammini, A., Vespignani, A., & Menczer, F. (2012). Competition among memes in a world with limited attention. *Scientific Reports, 2*, Article 335. https://doi.org/10.1038/srep00335

Xiao, W. (2026). A dialogic model for analyzing idea development in creativity-oriented text-based discussion forums. *International Journal of Computer-Supported Collaborative Learning*. Advance online publication. https://doi.org/10.1007/s11412-026-09472-z

Yuan, G., Zhang, J., & Chen, M.-H. (2022). Cross-community knowledge building with idea thread mapper. *International Journal of Computer-Supported Collaborative Learning, 17*(2), 293–326. https://doi.org/10.1007/s11412-022-09371-z

Zhang, J., Tao, D., Chen, M.-H., Sun, Y., Judson, D., & Naqvi, S. (2018). Co-organizing the collective journey of inquiry with idea thread mapper. *Journal of the Learning Sciences, 27*(3), 390–430. https://doi.org/10.1080/10508406.2018.1444992

Zhang, J., Yuan, G., & Bogouslavsky, M. (2020). Give student ideas a larger stage: Support cross-community interaction for knowledge building. *International Journal of Computer-Supported Collaborative Learning, 15*(4), 389–410. https://doi.org/10.1007/s11412-020-09332-4

Zhu, X., Chen, B., & DeLiema, D. (2026). Advancing collaborative discourse through knowledge synthesis. *International Journal of Computer-Supported Collaborative Learning, 21*(2), 205–235. https://doi.org/10.1007/s11412-025-09463-6

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- WhenThreadsFallQuietOnlineResource1.pdf