

# The Viewing-Response Gap: Sequential Behavioral Trace Evidence for Social Scaffolding Failure in Technology-Enhanced Learning

Madjid Sadallah, Marie Lefevre

▶ **To cite this version:**

Madjid Sadallah, Marie Lefevre. The Viewing-Response Gap: Sequential Behavioral Trace Evidence for Social Scaffolding Failure in Technology-Enhanced Learning. 2026. ⟨hal-05716764⟩

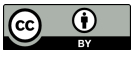
**HAL Id: hal-05716764**

**https://hal.science/hal-05716764v1**

Preprint submitted on 17 Aug 2026

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

# The Viewing–Response Gap: Sequential Behavioral Trace Evidence for Social Scaffolding Failure in Technology-Enhanced Learning

**Madjid Sadallah and Marie Lefevre**

*Université Claude Bernard Lyon 1, CNRS, INSA Lyon, LIRIS UMR5205, 69622 Villeurbanne, France*

## Abstract

Technology-enhanced learning platforms routinely embed social features yet frequently struggle to convert content engagement into social production. The prevailing explanation—learner motivation and cultural factors—has been untestable without behavioral data granular enough to distinguish unwillingness to interact from engagement without structural support for responding. This study provides sequential trace evidence for a structural alternative. Employing a within-participant observational contrast in which the same learners simultaneously received complete scaffolding for individual tasks and partial scaffolding for social production, we isolate a recommendation notification funnel collapse: explicit personalized invitations to engage with peer content generated high motivational activation but near-total production failure at the response stage—in the same participants who demonstrably adjusted individual task performance when scaffolding existed. The widespread pattern of social non-interaction is stable across measurement conditions and structurally consistent with documented patterns in learning management system discussion forums, massive open online courses, and collaborative annotation tools. The core conceptual contribution is a distinction between social affordance and social scaffolding—a distinction the behavioral evidence renders empirically consequential. Six provisional design principles are proposed, each derived from specific findings and operationalized with sufficient specificity to support empirical testing.

**Keywords**: learning analytics, behavioral trace analysis, social scaffolding, mobile learning, technology-enhanced learning, social production

## Introduction

A persistent pattern across technology-enhanced learning environments is the difficulty of converting content engagement into social production. Learners frequently access peer-generated content without responding, producing interaction spaces that resemble what has been termed a "parallel monologue" in the telecollaboration literature (e.g., O'Dowd, 2009). This pattern has been documented across learning management system discussion forums, massive open online courses, mobile learning applications,

virtual exchange platforms, and collaborative annotation tools (Ware & Kramsch, 2005; Dooly, 2011; Hrastinski, 2008; Seaton et al., 2014; Gyasi et al., 2021). While the present study examines this phenomenon in an intercultural mobile learning context, the mechanism it addresses may extend more broadly to digital learning environments in which social interaction is implemented primarily as a platform affordance rather than as a scaffolded production process.

We situate this investigation at the descriptive-characterization stage of a broader evidence trajectory linking observation, mechanism specification, and causal testing (cf. Shavelson & Towne, 2002; McKenney & Reeves, 2019). Our aim is therefore not to establish causal explanation but to specify the conditions for testing it.

Prevailing explanations for the parallel monologue have largely emphasized motivational, affective, or cultural mechanisms. Studies have pointed to insufficient intercultural motivation, communicative or linguistic anxiety, and divergent "cultures of use" as key explanatory factors (O'Dowd & Ritter, 2006; Liaw, 2006; Ware, 2005; Kramsch & Thorne, 2002). While these mechanisms are relevant, their explanatory sufficiency remains uncertain if structural conditions governing the transition from content consumption to response production are not examined simultaneously. Relatively little attention has been devoted to whether digital learning environments provide scaffolding for social interaction comparable in specificity and granularity to that typically provided for individual task completion.

Distinguishing between motivational absence ("learners did not wish to interact") and structural absence ("learners were not scaffolded to respond") requires behavioral evidence that self-report instruments cannot capture. Sequential behavioral trace data, by contrast, preserve the temporal relationship between viewing, prompting, and responding actions, allowing these mechanisms to be analytically separated. The present study therefore relies on fine-grained event-log data to examine the behavioral conditions under which social interaction fails to emerge.

The empirical context is MOBILES, a location-based mobile learning application designed to support intercultural competence development. The system was instrumented with a knowledge-tracing-based behavioral event-logging module (Settouti et al., 2009) that captured 490,578 behavioral events over the five-week active use period. In addition, the application incorporated a personalized scaffolding recommendation system grounded in an annotation-quality evaluation framework (Sadallah and Lefevre 2025a). This architecture created a diagnostic comparison within the same participant population:

learners who exhibited minimal social interaction simultaneously demonstrated measurable behavioral adaptation in individual documentation tasks when explicit scaffolding was present.

The analysis focuses on the recommendation notification funnel collapse—a behaviorally grounded construct operationalized through sequential examination of interaction-targeting notification events. By tracing the progression from notification exposure to content viewing and potential response, the construct identifies where breakdown occurs in the interaction process, helps distinguish structural from motivational interpretations of the parallel monologue, and grounds design implications for scaffolding social interaction in technology-enhanced learning environments.

Three research questions structure the empirical inquiry. RQ1 (Baseline condition) asks to what extent participants demonstrated behavioral responsiveness to individual scaffolding recommendations—establishing whether social non-interaction can be attributed to general scaffolding unresponsiveness. RQ2 examines the quantitative profile of social interaction produced during the study period. RQ3 asks at what stage behavioral collapse occurs within the recommendation notification funnel, and whether this pattern is stable across participants and temporal windows.

The article first reviews the theoretical framework, then presents MOBILES as the diagnostic instrument and details the methodological approach. It next reports the empirical results, discusses their theoretical and design implications, and closes with concluding remarks.

## Theoretical Framework and Literature Review

The present study addresses a tension in the e-learning and digital media literature between two conflicting bodies of evidence. On one side, theorizations of networked platform affordances have established that digital environments structurally reconfigure the conditions for social learning: Ünlüsoy et al. (2022) identify visibility, scalability, flexibility, and persistence as transformative affordances of networked digital platforms. On the other side, the Community of Inquiry framework (Garrison et al., 1999) and its subsequent social presence revision (Kreijns et al., 2014) establish that social presence is a prerequisite for genuine collaborative learning online, and that its absence in low-structure asynchronous environments favors passive consumption over dialogue. The present study provides sequential behavioral trace evidence for the empirical consequences of the gap between these two theoretical bodies: platforms that exploit networked affordances without providing the social presence architecture

that converts those affordances into dialogue are associated with the behavioral signature documented here—high content consumption accompanied by near-zero verbal production.

**The Parallel Monologue**

The parallel monologue was identified as a dominant pattern in early telecollaboration research (Belz, 2003; Belz & Müller-Hartmann, 2003; Thorne, 2003) and has shown itself resistant to technological remediation. O'Dowd and Ritter (2006) identified failure modes including incompatible communicative expectations, task design insufficiency, and learner motivation. Ware (2005) showed how cultural norms suppressed dialogue; Dooly (2011) found virtual exchange did not reliably produce critical perspective-taking; Gyasi et al. (2021), in a systematic review of K-12 contexts, found that social interaction consistently fell short of the reciprocity required for knowledge co-construction.

A clarification is necessary. The literature does not uniformly present a "pure motivational" account—Kreijns et al. (2003), O'Dowd and Ritter (2006), and Garrison et al. (1999) each identify structural and design factors. However, principal explanatory weight has been attributed to learner-level mechanisms because behavioral data at the required sequential resolution were unavailable. Activity logs without sequential resolution cannot distinguish "did not want to interact" from "encountered no structural support after engaging." The argument advanced here is not that prior motivational accounts were incorrect, but that they were invoked as sufficient in contexts where a prior structural condition—the asymmetric underdevelopment of social scaffolding—remained unexamined.

**The Differential Cost of Social Production Acts and Its Design Implications**

Face theory (Goffman, 1967; Brown & Levinson, 1987) identifies public commentary on another's personal cultural experience as inherently face-threatening, requiring the commenter to adopt a stance, risk cross-cultural misinterpretation, and expose identity—risks substantially amplified in second-language contexts (Horwitz et al., 1986; Ware, 2005). Fogg's (2009) Behavior Model identifies the convergence of motivation, ability, and a trigger as necessary for behavioral enactment. The "ability" condition is contextual: a public act requiring second-language production under face-threat demands more structural cost-reduction than a private act. The absence of any mechanism reducing intercultural commentary production cost—through response templates, guided prompts, or graduated scaffolds—constitutes the absence of the ability-enabling condition Fogg's model identifies as necessary even when motivation is present.

An objection warrants consideration. Comparing scaffolding responsiveness across individual documentation and social interaction behaviors is not a symmetric comparison: the acts differ in cost, public exposure, and face-threat risk. We accept this non-equivalence as a limitation of the within-participant contrast. The comparison demonstrates that behavioral responsiveness to scaffolding was present in this population, making motivational absence the least parsimonious account of the Stage 3 collapse. The cost-difference objection predicts a lower but non-negligible Stage 3 rate; the observed 4.7%—barely above the 1.0% organic baseline despite 86.7% Stage 2 engagement—is inconsistent with a cost-only account, though it does not rule out cost factors operating in parallel.

**The Viewing-Response Gap as a Behavioral Diagnostic**

A well-documented intention-behavior gap (see Webb & Sheeran, 2006, for a meta-analytic review) provides a framework for understanding why motivation may be present yet behavior absent. The present study operationalizes the more conservative construct of the viewing-response gap: the behavioral signature of a user who demonstrably engaged with a peer's content—navigating to it in response to an explicit recommendation, evidenced by a post-view event in the log—but produced no verbal response. This construct makes no claim about internal intention states; it claims only that engagement occurred and production did not, thereby excluding the interpretation that social non-interaction reflects simple motivational absence. The construct is detectable only through fine-grained sequential trace data.

**Pedagogical Scaffolding: The Individual—Social Asymmetry**

Pedagogical scaffolding, grounded in Vygotsky's (1978) Zone of Proximal Development and operationalized through Reiser's (2004) framework, encompasses procedural, metacognitive, conceptual, and strategic scaffolding types. A theoretically significant asymmetry, noted by Kreijns et al. (2003) and persistent in subsequent design practice, concerns the differential development of individual versus social scaffolding. The majority of scaffolding research targets individual cognition; social scaffolding—mechanisms specifically designed to support the transition from individual knowledge construction to collaborative discourse—remains substantially under-theorized and under-implemented. The practical distinction is between social affordance, understood as providing a comment interface, and social scaffolding, understood as monitoring view-without-response sequences, detecting collapse, and delivering targeted production support. MOBILES provided the former without the latter—a pattern

consistent with the broader design literature on social affordance in mobile learning environments (Kreijns et al., 2003).

### Community of Inquiry, Social Presence, and the Prerequisites of Online Dialogue

The Community of Inquiry model (Garrison et al., 1999) identifies social presence, cognitive presence, and teaching presence as jointly required for effective collaborative online learning. Social Presence Theory (Gunawardena, 1995) predicts that asynchronous, low-structure environments exhibit low spontaneous interaction unless counteracted through deliberate social presence design. Kreijns et al. (2014) provide an important revision: distinguishing perceived social presence as a psychological state from actual social presence as enacted behavior, they demonstrate that the gap between them constitutes the design problem e-learning environments have struggled to close. The funnel collapse documented in the present study represents, in Kreijns et al.'s (2014) terms, the behavioral signature of high perceived social presence activated by notification, evidenced by 86.7% Stage 2 engagement, and collapsed actual social presence production, evidenced by 4.7% Stage 3 response. This failure is also interpretable through the Community of Inquiry's third element, teaching presence—defined as the design and facilitation of the learning experience (Garrison et al., 1999). The viewing-response gap constitutes a teaching presence failure at the design level: the scaffolding architecture teaching presence is intended to provide was absent at the production stage.

### Fine-Grained Trace-Based Learning Analytics

Learning analytics (Siemens, 2013) and trace-based approaches (Settouti et al., 2009) enable researchers to examine not only what learners report but what they actually do in response to explicit prompts. Sequential behavioral resolution at millisecond granularity enables detection of the viewing-response gap. Prior learning analytics studies of mobile intercultural learning have documented behavioral frequencies without the sequential resolution required to examine the behavioral pathway between an explicit social prompt and its non-execution. The present study demonstrates that fine-grained event-log instrumentation enables this resolution, yielding the recommendation notification funnel collapse—a pattern invisible to coarser-grained approaches. A recent review of log-based measurement identifies generalizability and transferability as the central methodological challenges for single-deployment log studies (Hershkovitz & Alexandron, 2025)—challenges explicitly acknowledged in the Limitations section and addressed through event-level rather than participant-level inferential claims. Without

sequential resolution at this granularity, the structural and motivational accounts remain empirically indistinguishable.

## MOBILES as a Diagnostic Instrument

MOBILES is a cross-platform location-based mobile application for intercultural competence development, deployed on both iOS and Android operating systems. The application serves in the present study as a diagnostic instrument for examining whether behavioral trace data can distinguish structural from motivational accounts of the parallel monologue. Two design features create the theoretical test conditions. First, the application included intentionally designed social interaction mechanisms such that the motivational account would predict at least some improvement: users received not merely the affordance but an explicit invitation to interact. Second, its personalized scaffolding recommendation system, grounded in an annotation quality evaluation framework documented prior to the study (Sadallah and Lefevre 2025; Sadallah and Lefevre 2025a), monitored individual behavioral indicators and delivered targeted interventions.

MOBILES supports two primary content types. Annotations are multimodal records comprising free text, geo-referenced location, photographs, emoticons, semantic tags, and icon categories. Tiered sharing options—public, group, and private—accommodate varying comfort levels. Tours are curated sequences of annotations forming thematic itineraries, supported through GPS tracking, manual path tracing, or retrospective compilation.

The asymmetry between the fully scaffolded individual documentation environment shown in Figure 1, panel (b), and the absence of equivalent production scaffolding at the point of peer content viewing in panel (a) is central to our structural argument.

MOBILES integrates a knowledge-tracing-based behavioral event-logging module (Settouti et al., 2009) providing continuous, non-intrusive logging of all user interactions at millisecond resolution. The system captures five behavioral dimensions relevant to this study: content production and editing (annotation, tour creation, modification, sharing); social interaction (comments, emoji reactions, favorites); notification engagement (receipt, opening, dismissal); spatial navigation (GPS position, map filter use, location search); and session activity (login, logout, synchronization). Each event is time-

stamped and geolocated where relevant, enabling reconstruction of complete behavioral sequences—including the three-stage notification-to-viewing-to-response pathway central to RQ3.

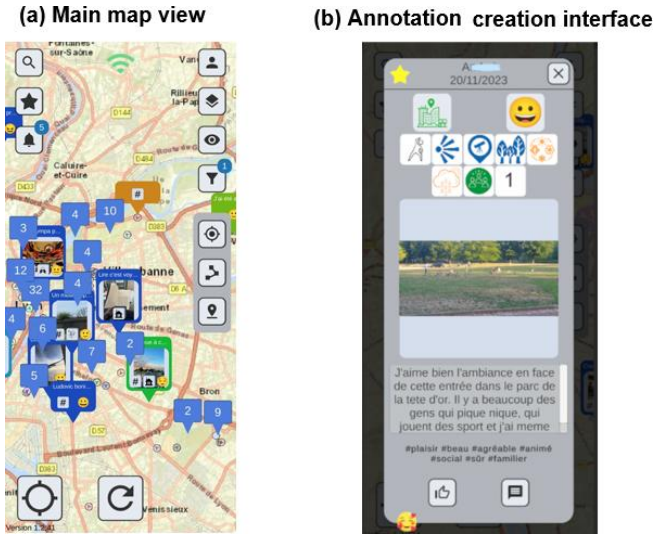



**Figure 1** MOBILES application interface illustrating the asymmetry between peer content access (a) and the individually scaffolded production environment (b).

The application's scaffolding architecture spans four types. Procedural scaffolding operates through an interactive map with filters that reduce cognitive load. Metacognitive scaffolding functions through annotation creation prompts that invite multimodal reflection and a recommendation notification system that delivers thematic prompts. Conceptual scaffolding operates through tour creation that supports synthesis of individual experiences. Strategic scaffolding functions through a personalized recommendation system (Sadallah and Lefevre 2025) that suggests relevant peer content based on behavioral patterns. Critically, this architecture was complete for individual behaviors but stopped one step short for social production behaviors: it encompassed monitoring, detection, and prompt delivery, but not production-cost reduction. This asymmetry—between social affordance and social scaffolding—constitutes the within-participant observational contrast the study exploits.

An important distinction applies to the notification system. General thematic notifications were weekly, non-personalized metacognitive prompts. Interaction-targeting recommendation notifications were personalized prompts explicitly inviting social engagement with a specific named peer's content, triggered by the system's detection of insufficient recent social interaction. Of 434 total notifications

delivered — comprising 210 interaction-targeting recommendation notifications, 219 individual behavior-targeting recommendations, and 5 weekly general thematic notifications — 210 (48.4%) were interaction-targeting. It is exclusively this last category that constitutes the funnel analyzed in RQ3. By restricting the funnel analysis to notifications that were explicitly personalized and interaction-targeting, we examine conditions under which motivational activation is maximized by design—making the Stage 3 collapse a structurally conservative test: if social production fails even when motivation is deliberately primed, the structural insufficiency account is strengthened.

## Methodology

### Research Design

The present study employs quantitative behavioral trace analysis as the primary method, supplemented by qualitative survey data serving a corroborative-illustrative function. All participants provided informed written consent prior to data collection, covering behavioral event logging. The same participants received, simultaneously, a complete scaffolding architecture for individual behaviors and a partial scaffolding architecture for social behaviors. The logic is analogous to component isolation in optimization research (Collins et al., 2005). The differential behavioral outcomes—robust individual scaffolding responsiveness accompanied by near-total social production collapse—constitute a within-participant, within-period, within-platform contrast; confounds associated with participant characteristics are substantially reduced by construction. This within-participant logic is the primary inferential basis for the central argument: differential outcomes cannot be attributed to between-participant variance in motivation, language proficiency, or technology confidence.

The research team includes members involved in MOBILES' design. Procedural safeguards were implemented to mitigate confirmation bias. These safeguards included automated logging without researcher involvement, pre-specified analyses, an independently accessible annotation quality framework (Sadallah and Lefevre 2025; Sadallah and Lefevre 2025a), two-researcher qualitative coding with consensus, and the central finding—near-total absence of social interaction—constituting a negative result relative to design ambitions.

### Participants and Context

Thirty-one international students consented to participate, comprising 15 male and 16 female participants with a mean age of 25.2 years, a median of 24, and a range of 19 to 43 years. Eighteen

participants were under 25 years of age, nine were aged 25 to 29, and four were over 29. The sample represented 15 nationalities: 18 from African countries, including participants from North Africa and Sub-Saharan Africa; 4 from the Americas; 7 from Asia, including participants from Southeast, East, West, and Pacific Asia; and 2 from Europe. Participants were enrolled across undergraduate (n = 14), postgraduate (n = 8), and first-year (n = 9) programmes at multiple universities in Lyon, France.

Volunteer recruitment with a performance-contingent financial reward selects for higher technology engagement and intercultural motivation than the general student population (Shadish et al., 2002). The 95.3% Stage 2-to-3 collapse was nonetheless observed in this motivationally and financially motivated sample: a purely motivational account becomes progressively less parsimonious when social non-interaction persists even in a sample selected for engagement and rewarded for regular participation.

Participant-level inferential claims with n = 31 are treated as preliminary and descriptive. Behavioral-sequence-level claims operate at the event level, encompassing 210 notification events, and are not reducible to 31 data points: the per-participant zero-count, with 25 of 31 participants producing zero Stage 3 events, is a descriptive count fact, not an inferential statistic.

**Procedure**

Data collection unfolded in three phases over a six-week period in Spring 2024.

**Phase 1: Onboarding (Week 0, March 18–22).** All participants attended a 90-minute in-person onboarding session covering study rationale, ethical procedures, and participant rights, followed by a 45-minute hands-on tutorial in which each participant installed MOBILES and completed a guided walkthrough of all core features, with particular emphasis on the social interaction interface and its intended pedagogical function.

**Phase 2: Active use (Weeks 1–5, March 25–May 2).** Participants used MOBILES autonomously, integrating the application into their daily movements across Lyon with no minimum usage requirement. The event-logging module recorded all interactions continuously in the background. The personalized recommendation system delivered individual scaffolding interventions autonomously; interaction-targeting notifications were delivered when the system detected insufficient recent social engagement. Researcher-designed contact was intentionally limited to preserve naturalistic behavior, consisting only of five automated weekly in-app thematic notifications (non-personalized, distinct from the interaction-

targeting notifications analyzed under RQ3) and one mid-study technical support message delivered in Week 3.

**Phase 3: Reflection (Post-study, May 5–12).** Participants completed an online reflection survey within seven days following the active use period. Survey responses were collected only after trace data extraction to avoid influencing naturalistic in-app behavior.

**Data Sources and Instruments**

**Behavioral event-log data.** All 490,578 user interactions were captured via the knowledge-tracing-based event-logging module. Key derived metrics included total daily interactions, annotation composition variables aligned with (Sadallah and Lefevre 2025), spatial coordinates, notification engagement events, and all social interaction events. Sessions were identified using a 30-minute inactivity threshold, a standard convention in mobile application log analysis; mean session duration was 12.3 minutes. For the recommendation notification funnel, three-step behavioral sequences were examined. First, a notification-open event, recorded as *selectNotification*, was captured. Second, a post-view event, recorded as *openPost*, on the referred annotation occurring within 15 minutes was captured. Third, a comment or reaction event, recorded as *addMessage* or *addEmoji* respectively, on that annotation occurring within 24 hours was captured. The 15-minute Stage 2 window was set to exceed the mean session duration of 12.3 minutes, ensuring that content views occurring within a typical session are captured. The 24-hour Stage 3 window spans one full daily usage cycle. Both windows were pre-specified before data extraction. Sensitivity analyses using alternative windows (10 and 20 minutes for Stage 2; 8 and 48 hours for Stage 3) produced comparable collapse rates, confirming robustness.

**Participant feedback survey.** The online survey comprised five sections: Technology Acceptance (Davis, 1989); Experience Documentation; City Exploration; Social Connection and Interaction; and Reflexivity. All quantitative items used 5-point Likert scales. The survey was pilot-tested for clarity and cross-cultural appropriateness. The survey component plays an illustrative role: open-response quotations are reported where they converge with quantitative patterns and are not advanced as findings of an independent qualitative analysis. To ensure that quotations selected for reporting are representative of the inductively identified thematic range rather than selectively illustrative, systematic inter-rater coding was conducted, yielding Cohen's $\kappa = .81$, indicating almost perfect agreement according to Landis and Koch (1977), with discrepancies resolved through discussion.

Of 31 participants, 28 returned at least one survey section, representing 90.3%, and 25 completed all five sections. Open-response items were optional within each section while Likert-scale items were compulsory, yielding 118 open responses across eight items, with a mean of 4.2 responses per respondent. Quotations reported in the text were selected to represent the thematic range most directly relevant to the viewing-response experience, principally perceived production barriers, spatial peer influence, and scaffolding awareness.

### Analysis Procedures

Linear regression examined temporal engagement trends; residual autocorrelation was assessed via the Durbin-Watson statistic. K-means clustering with k = 3, determined through elbow and silhouette methods with 50 random restarts, identified descriptive engagement profile typologies. Haversine distances (a standard great-circle formula for computing geographic distance between GPS coordinates) examined spatial proximity between new annotation creation locations and previously viewed peer annotation locations. Welch's t-tests examined textual richness by comment receipt. Wilson 95% confidence intervals were computed for all funnel conversion rates; for Stage 3, bootstrap 95% confidence intervals were additionally computed to correct for within-participant clustering—multiple notifications to the same participant violate the independence assumption of Wilson confidence intervals, as notification events are nested within participants. Bootstrap resampling (1,000 participant-level iterations) yielded the 2.5th–97.5th percentile CI. Recommendation impact was assessed using 7-day pre- and post-intervention windows; all effects are treated as preliminary pending trend-controlled analysis. Daily interaction counts were normalized to the maximum daily interactions observed during the study period to facilitate interpretation of the regression coefficient.

Quantitative trace findings constitute the primary evidence base. Survey quotations are integrated as corroborative illustrations where they converge with quantitative patterns, and explicitly flagged when they could support alternative interpretations.

## Results

### Individual Scaffolding Responsiveness: Establishing the Baseline

#### *Temporal Engagement Trajectory*

Over the five-week study period, total daily interactions increased significantly, with $\beta$ = 0.42 normalized units per day (normalized to the maximum daily interactions observed), $R^2$ = 0.68, $p$ < .001, and a Durbin-Watson statistic of $d$ = 1.83, indicating no problematic serial autocorrelation. Users created 497 annotations in total, with a mean of 16.0 per user; content production accelerated markedly in Weeks 3 and 4, coinciding with peak scaffolding recommendation density.

Participants' open-ended responses reflected this growing engagement. One participant noted, "At first I only posted when I saw something unusual, but then I started looking for things to document" (P14, Cameroonian). Another described a shift in perception: "The app changed how I walk—I pay attention differently now" (P22, Vietnamese).

The juxtaposition of Figure 2 panels (a) and (b) illustrates the divergence between sustained individual documentation activity and declining social production over the same period.

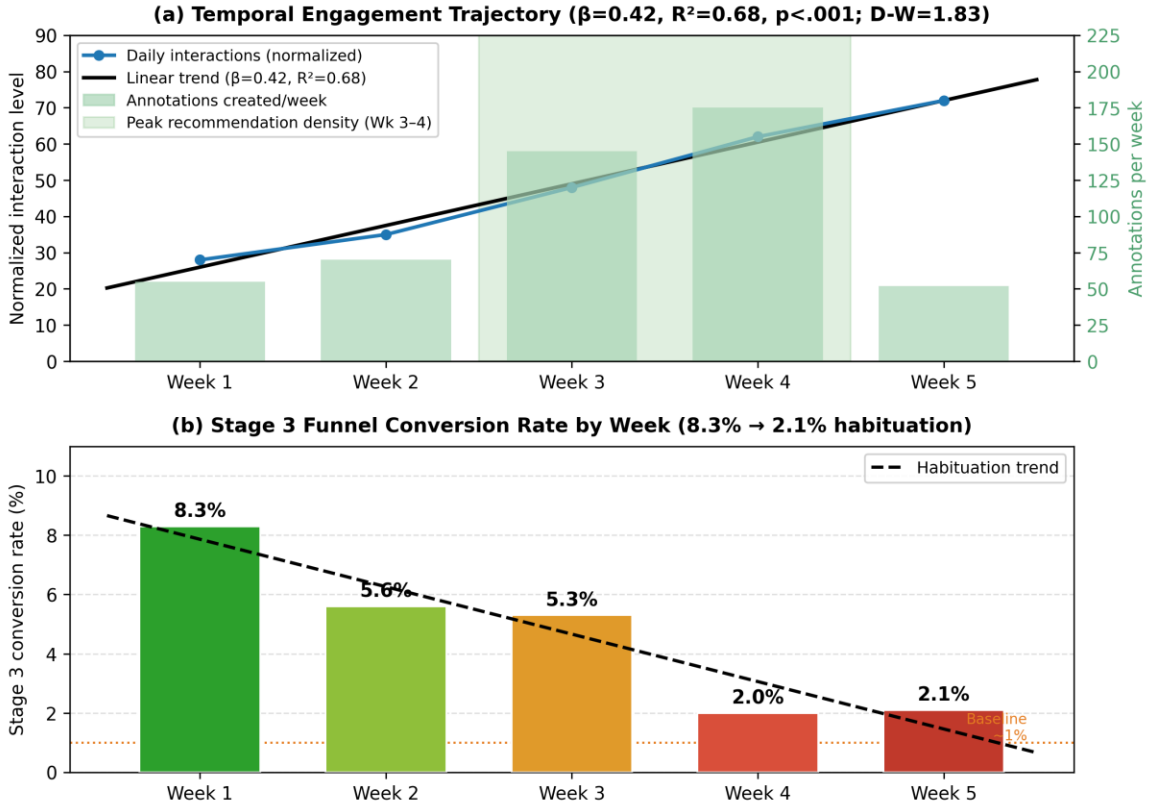



**Figure 2** Temporal dynamics: (a) daily interaction trajectory (β = 0.42, $R^2$ = 0.68) with weekly annotation production; (b) weekly Stage 3 conversion rate using Stage 1 (notification-open events) as denominator (8.3% Week 1 → 2.1% Week 5; see Table 6 note for relationship to primary 4.7% estimate).

### *Engagement Profile Analysis*

K-means clustering with k = 3, determined through elbow and silhouette methods, yielded three descriptive profile typologies. Cluster sizes, particularly Profile A with n = 4, preclude statistical between-cluster inference; these are presented as descriptive typologies only (Table 1).

**Table 1** Engagement Profile Typologies

| **Profile** | ***n*** | **Annotations** | **Sessions** | **Features/session** | **Active Days** |
|---|---|---|---|---|---|
| | | ***M*** | ***M*** | ***M*** | ***M*** |
| A – Intensive Documenters | 4 | 41.2 | 22.3 | 6.1 | 29 |
| B – Consistent Contributors | 18 | 16.8 | 12.7 | 4.3 | 22 |
| C – Episodic Explorers | 9 | 8.1 | 6.2 | 3.1 | 14 |

*Note.* Social interaction rate across all three profiles was negligible (0–3 comments per user, with 6 of 31 participants producing any text comment at all; these six correspond to the non-zero Stage 3 producers reported in Table 5, though Stage 3 additionally includes emoji reactions). Cluster sizes (especially Profile A, $n = 4$) preclude statistical between-cluster inference; these are descriptive typologies only.

Cross-profile uniformity of Stage 3 non-production constitutes a structural indicator. Social non-interaction was cross-profile: Intensive Documenters, who produced over 40 annotations each and were active 29 of 35 days, representing the most scaffolding-responsive participants, produced fewer than four comments each—rates no higher than those of Episodic Explorers. Uniform social non-production across all engagement levels is inconsistent with a motivational trait account and consistent with a structural one.

### *Scaffolding Recommendation Impact on Individual Behavior*

The recommendation system delivered 219 individual behavior-targeting recommendations (distinct from the 210 interaction-targeting notifications analyzed under RQ3), with behavioral changes detailed in Table 2. These effects establish that this participant group was demonstrably responsive to scaffolding interventions during this study period—the necessary baseline for our central argument. All estimates are preliminary, based on a 7-day pre/post window without formal trend control (overall engagement trend: $\beta = 0.42$, $R^2 = 0.68$, $p < .001$). The image use effect (+300% absolute change) is treated as directionally reliable given the magnitude of the shift, while textual expression, sensory icon diversity, and memory icon diversity are treated as preliminary estimates only. Survey data converge with this interpretation: 78% of participants reported feeling motivated by recommendations to diversify content.

One participant explained, "When I got a suggestion to add more photos, I started taking pictures of things I'd just been walking past" (P7, Moroccan). Figure 3 presents the within-sample behavioral asymmetry that constitutes our core finding.

**Table 2** Individual Scaffolding Recommendation Impact on Behavioral Indicators

| Recommendation Type | Baseline | Post-Recommend. | Change | Reliability Note |
|---|---|---|---|---|
| Textual expression | 18.3 words/post | 21.2 words/post | +16% | Preliminary—7-day window, no trend control |
| Image use | 12% of posts | 48% of posts | +300% | Directionally reliable; magnitude rules out trend confounding. |
| Sensory icon diversity | 0.31 types/post | 0.49 types/post | +59.38% | Preliminary. |
| Activity icon diversity | 0.52 types/post | 0.54 types/post | +3.85% | Non-informative without trend control; not interpreted. |
| Memory icon diversity | 0.28 types/post | 0.39 types/post | +39% | Preliminary. |

*Note.* All estimates use a 7-day pre/post window without formal trend control.

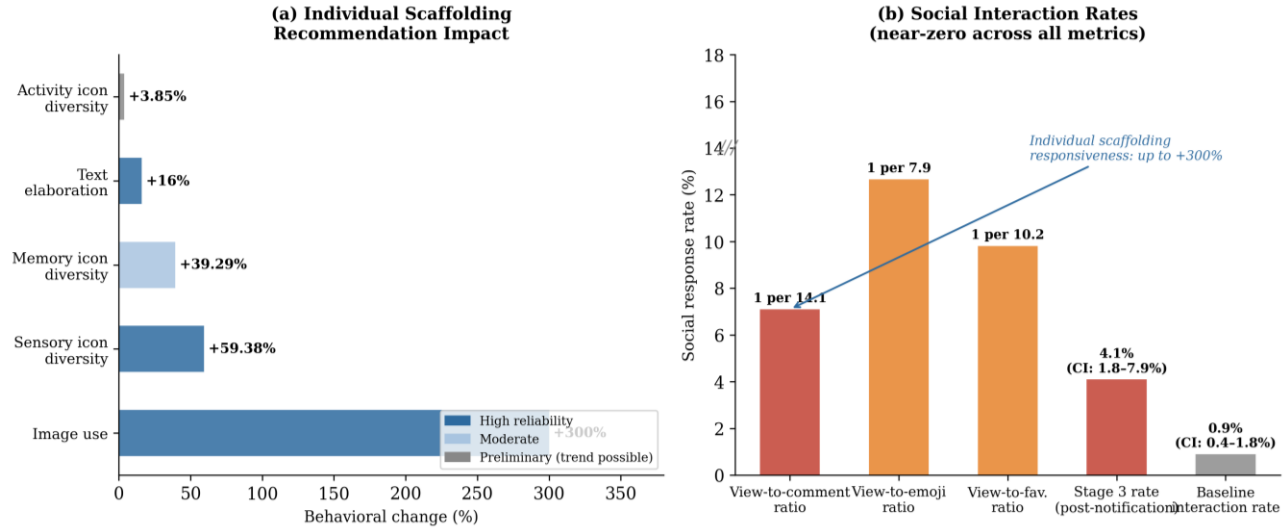



**Figure 3** Within-participant behavioral asymmetry: (a) impact of individual scaffolding recommendations (five dimensions); (b) social interaction rates (comments, reactions, favorites).

## The Nature and Behavioral Signature of Social Non-Interaction

### *Quantitative Portrait of Social Interaction*

Table 3 presents the quantitative portrait of social interaction observed during the study period. The 1:14.1 view-to-comment ratio is notable in combination with RQ3 evidence that these views followed explicit interaction invitations, and that the same participants demonstrably responded to non-social scaffolding during the same period. The 63 comments (mean length: 12.5 words, range: 2–47) were predominantly phatic; none engaged comparatively with a peer's cultural interpretation as defined by Deardorff (2006). Of the 175 total social interaction events (63 comments + 112 emoji reactions), 8 occurred within the funnel pathway. The remaining 167 were produced either more than 24 hours after a peer view, via navigation not linked to a traceable *openPost* event, or through browsing without a prior notification.

**Table 3** Quantitative Portrait of Social Interaction

| **Metric** | **Total** | **Per User *M*** | **Ratio to Views** |
|---|---|---|---|
| Annotations shared publicly | 497 | 16.0 | — |
| Peer views of shared annotations | 890 | 28.7 | — |
| Comments (addMessage) | 63 | 2.03 | 1 per 14.1 views |
| Emoji reactions (addEmoji) | 112 | 3.61 | 1 per 7.9 views |
| Favorites (addFavorite) | 87 | 2.81 | 1 per 10.2 views |

*Note.* View:comment ratio of 1:14.1 is computed from 890 total peer views and 63 total comments. Favorites (*addFavorite*) are included as engagement indicators but excluded from the 175-event total and from Stage 3 funnel counts, which are restricted to direct peer-addressed responses.

Participants' open-ended responses highlighted perceived barriers. One participant explained, "I read what others posted but I didn't know what to say that would be interesting" (P3, Ivorian). Another cited linguistic insecurity: "I felt I should comment but I was not sure my French was good enough" (P19, Colombian). A third described a different kind of response: "Reading what others saw made me want to go there but not necessarily to write back to them" (P27, Thai).

### *Analysis of Commented Versus Non-Commented Annotations*

Commented annotations, numbering 23, were significantly more textually elaborate than non-commented ones, numbering 474. The mean word count for commented annotations was 42.3 words,

compared to 18.7 words for non-commented annotations. Welch's t-test yielded $t(28.4) = 4.12$, $p = .0003$, with Cohen's $d = 1.01$ and a 95% confidence interval of 0.46 to 1.55. The extreme asymmetry in group sizes ($n = 23$ vs. $n = 474$) is a behavioral phenomenon, not a statistical artifact. The wide confidence interval reflects the small commented-annotation group; this finding should be treated as directionally informative only. It is consistent with a threshold hypothesis: verbal social response may require a minimum level of elaboration that gives a potential commenter something substantive to engage with. This hypothesis requires replication in a design with balanced group sizes. Given the extreme asymmetry (n = 23 vs. n = 474), the effective sample for power calculation is constrained by the smaller group; nonetheless, post-hoc power for the observed d = 1.01 at $\alpha = .05$ exceeds .99, confirming the test is not underpowered. The caveat is interpretive rather than statistical: the comparison is post-hoc, the threshold is theory-free, and the result requires confirmatory replication with an a priori design. The Shapiro-Wilk test confirmed approximate normality in the commented-annotation group, with $W = 0.93$, $p = .12$; a non-parametric Mann-Whitney $U$ test yielded a consistent result ($U = 4{,}891$, $p = .0004$), confirming the finding is not dependent on the normality assumption.

### *Spatial Proximity as Suggestive Evidence*

Two confounds constrain interpretation of spatial proximity data: urban geography bias, since student-relevant zones in Lyon are spatially concentrated, and temporal co-presence, since participants exploring similar areas at similar times may generate proximity patterns independently of peer content viewing. Neither can be resolved with the current data. With this framing, 25.1% of new annotations were created within one kilometer of a peer annotation viewed in the preceding 48 hours (see Figure 4).

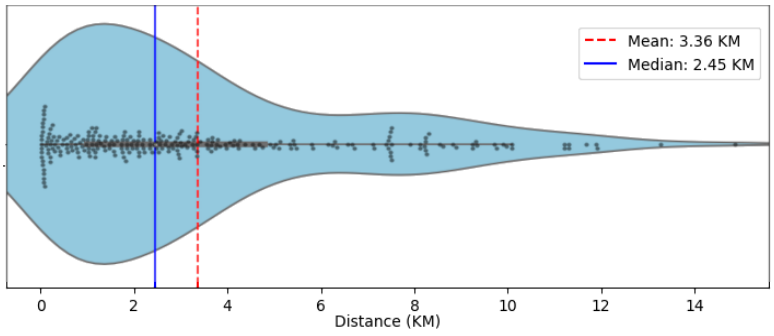



**Figure 4** Spatial proximity (km) between created annotations and previously viewed peer annotations. Mean = 3.36 km (red dashed), median = 2.45 km (blue solid). 25.1% within 1 km.

## The Recommendation Notification Funnel Collapse

### *Funnel Architecture, Collapse Point, and Per-Participant Distribution*

The recommendation notification funnel operationalizes a three-stage behavioral sequence for interaction-targeting recommendation notifications exclusively: 210 personalized invitations to engage with a specific named peer's annotation.

Figure 5 presents the recommendation notification funnel collapse across 210 interaction-targeting notifications. Stage 1 (notification opened) occurred for 93.3% of notifications (Wilson CI: 89.1–96.2%). Stage 2 (content viewed within 15 minutes) occurred for 86.7% of opened notifications (Wilson CI: 81.4–90.8%). Stage 3 (social interaction within 24 hours) occurred for 4.7% of viewed notifications (bootstrap CI: 1.2–8.4%), representing a 95.3% attrition at the production stage. Using Stage 1 (*n* = 196 notification-openers) as denominator instead yields 4.1% (bootstrap CI: 1.0–7.6%; Wilson CI: 1.8–7.9%). Bootstrap confidence intervals correct for within-participant clustering, as notification events are nested within participants. Panel (b) provides a funnel visualization with trapezoids proportional to Stage 3 volume. Table 4 details the stage-by-stage conversion rates with confidence intervals. Stage 3 uses Stage 2 as denominator (n = 170); the organic baseline applies the same 24-hour window for comparability.

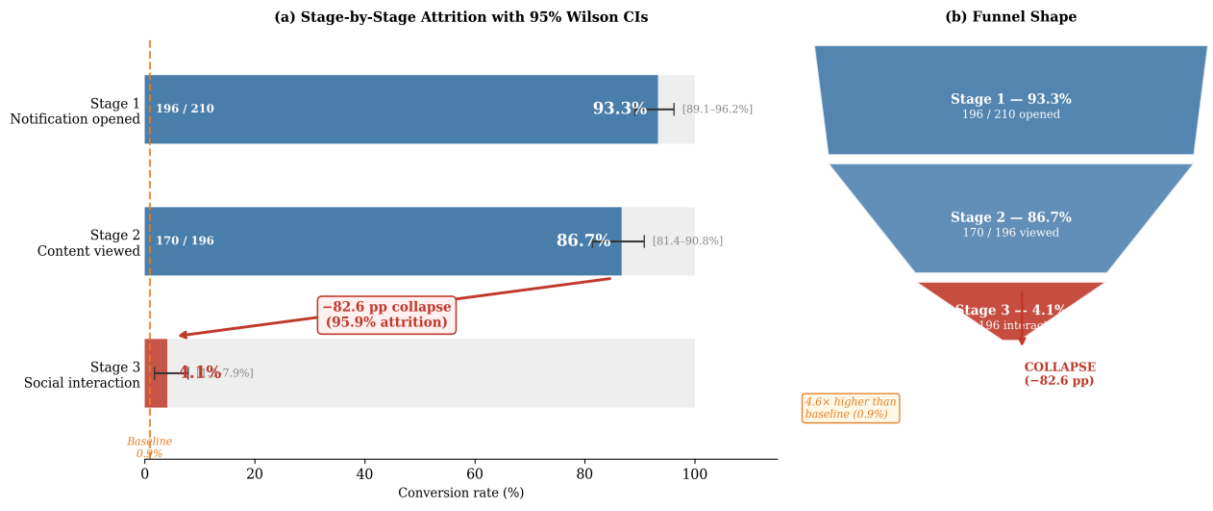



**Figure 5** Recommendation notification funnel collapse (n = 210): (a) stage-by-stage attrition with 95% CIs and baseline (1.0%); (b) funnel visualization with proportional trapezoids.

**Table 4** Recommendation Notification Funnel: Stage-by-Stage Conversion Rates (n = 210)

| Stage | Description | N / Total | Rate (95% CI) | Interpretation |
|---|---|---|---|---|
| Stage 1 | Notification opened (selectNotification) | 196 / 210 | 93.3%; Wilson: 89.1–96.2% | Near-universal awareness |
| Stage 2 | Referred content viewed (openPost ≤ 15 min) | 170 / 196 | 86.7%; Wilson: 81.4–90.8% | High engagement |
| ▼ Stage 3 | Social interaction produced (addMessage / addEmoji ≤ 24 h) | 8 / 170 | 4.7%; Bootstrap CI: 1.2–8.4% (primary); Wilson CI: 2.1–9.1% | COLLAPSE: −82.0 pp (95.3% attrition) |

*Note.* Organic baseline: 7/720 = 1.0% (Wilson CI: 0.5–2.0%). Figure 5 reports Stage 1-denominator figures (95.9% attrition, −82.6 pp); Table 4 uses Stage 2 as primary denominator.

The per-participant analysis is straightforward: 25 of 31 participants (80.6%) produced zero Stage 3 social interactions following any notified content view. This is a descriptive count fact; it does not require inferential statistics. Among the six non-zero producers, individual Stage 2→3 rates ranged from 14.3% to 25.0% (14.3%, 16.7%, 19.0%, 20.0%, 22.2%, and 25.0%; M = 19.7%, SD = 3.6% computed on unrounded rates prior to display rounding), producing the eight total Stage 3 events. Table 5 presents the full per-participant distribution.

**Table 5** Per-Participant Stage 2→3 Conversion Rate Distribution

| Participant group | n | Notifications M (SD) | Stage 2→3 rate |
|---|---|---|---|
| Zero Stage 3 producers | 25 | 6.4 (4.2) | 0% (all) |
| Non-zero Stage 3 producers | 6 | 8.3 (3.8) | 14.3–25.0%; M = 19.7%, SD = 3.6% |
| *All participants* | *31* | *6.8 (4.1)* | *Mdn = 0%; Aggregate = 4.7%* |

*Note.* Aggregate rate in the All participants row uses Stage 2 (*n* = 170) as denominator.

User-level distribution across engagement profiles further supports the structural interpretation. The six non-zero Stage 3 producers were distributed across all three engagement profiles: one from Profile A (*Intensive Documenters*), four from Profile B (*Consistent Contributors*), and one from Profile C (*Episodic Explorers*). This distribution indicates that social non-production was not confined to low-engagement participants and is inconsistent with a stable trait account.

### *The Natural Reference Condition and Its Interpretive Limits*

The 720 peer content views generated without a prior interaction-targeting notification—comprising spontaneous map-based browsing and views triggered by general thematic notifications—constituted a within-dataset natural reference condition and produced seven social interactions within 24 hours (1.0%; Wilson CI: 0.5–2.0%). The 170 views generated following an explicit interaction-targeting notification produced eight social interactions (4.7%). This constitutes a 4.7-fold difference on the same platform, with the same participants, during the same period. This difference conflates a notification effect with a selection effect, since post-notification viewers were users who also chose to open and act on the notification. Nonetheless, even attributing all of the difference to motivational activation, 95.3% of motivationally activated users still did not produce social interaction; the Stage 2-to-3 collapse calls for explanation regardless of how the difference is partitioned.

### *The Viewing-Response Gap as Structural Signature*

Three alternative configurations illuminate the significance of the observed pattern. Alternative A (*motivational absence*) predicts a collapsed Stage 1; the observed 93.3% is the opposite. Alternative B (*motivational presence with structural support*) predicts a non-collapsed Stage 3; the observed 4.7% represents near-total failure. Alternative C (*motivational engagement without production scaffolding*) predicts high Stage 1 and Stage 2 accompanied by collapsed Stage 3—the pattern observed: motivational engagement was activated, as evidenced by 93.3% Stage 1 and 86.7% Stage 2, without structural reduction of production cost to an actionable threshold at Stage 3. Of the three configurations, only the observed pattern is consistent with all available evidence simultaneously: Stage 1 rules out disengagement; Stage 2 rules out avoidance of peer content; Stage 3 collapse localizes the failure specifically at production. A structural account is not merely compatible with this configuration—it is the most parsimonious account consistent with all three observations simultaneously. Motivational accounts, taken alone, predict either collapsed Stage 1 or uniformly collapsed Stage 3 across all participants regardless of scaffolding availability; neither prediction corresponds to what is observed.

Participants' qualitative accounts corroborate this configuration. One participant explained, "I read [peer's] annotation about the market—it was interesting—but I didn't know how to start a conversation about it" (P12, Congolese). Another articulated the perceived triviality of available responses: "What would I even say? Just 'nice'? That feels useless" (P3, Ivorian). A third described normative uncertainty:

"I wanted to ask him about the mosque he described but I wasn't sure if that was appropriate" (P8, French).

### *Temporal Dynamics and Window Robustness*

The weekly rates rest on small event counts and should be interpreted with caution. Table 6 presents the weekly Stage 1 and Stage 3 event counts and conversion rates.

**Table 6** Weekly Stage 1 and Stage 3 Event Counts and Conversion Rates

| Week | Stage 1 events (*n*) | Stage 3 events (*n*) | Stage 3 rate (%) | Note |
|---|---|---|---|---|
| 1 | 24 | 2 | 8.3% | 2 events; very wide effective CI |
| 2 | 36 | 2 | 5.6% | 2 events |
| 3 | 38 | 2 | 5.3% | 2 events |
| 4 | 50 | 1 | 2.0% | 1 event |
| 5 | 48 | 1 | 2.1% | 1 event |
| *Total* | *196* | *8* | *4.1%* | |

*Note.* Weekly rates and aggregate use Stage 1 (notification-open events) as denominator. The aggregate 4.1% (8/196) therefore differs from the primary Stage 3 rate of 4.7% (8/170) reported in Table 4, which uses Stage 2 as denominator.

Stage 3 conversion rates declined from Week 1 to Week 5, though weekly estimates carry wide uncertainty (95% CIs spanning approximately ±10–15 percentage points) and should not be treated as precise trend estimates. The declining trend interpretation rests on divergence with individual documentation behavior over the same period (Figure 2, panel a), not on the precision of individual weekly estimates. Two interpretations are compatible with this declining trend: novelty-driven inflation of early rates, and habituation to the notification prompt as the study progressed. The novelty-dissipation account predicts declining engagement across behavioral dimensions; however, the data show declining social production but stable or increasing individual documentation. This within-sample divergence—declining social production alongside growing individual production by the same participants—is inconsistent with uniform novelty dissipation and lends additional support to the structural account.

Window robustness was assessed by recalculating the Stage 2-to-3 collapse rate using alternative temporal windows. Using a Stage 2 window of 10 minutes and a Stage 3 window of 8 hours yielded a collapse rate of 94.8%. Using a Stage 2 window of 20 minutes and a Stage 3 window of 48 hours

yielded a collapse rate of 93.6%. These compare to the primary estimate of 95.3% using the pre-specified windows of 15 minutes and 24 hours. The collapse is stable across all parameterizations tested, ranging from 93.6% to 95.3%.

## Discussion

### The Within-Participant Asymmetric-Scaffolding Contrast and Its Inferential Scope

The study's core finding is the within-sample behavioral asymmetry we established: the same participants who produced near-zero social interaction demonstrably adjusted individual documentation behavior when scaffolding existed for that behavior. This contrast is observational rather than experimental: the scaffolding asymmetry was a pre-existing structural feature of the platform, not a condition randomly assigned. The inferential claims are therefore bounded accordingly.

The comparison is not symmetric. Individual recommendations targeted low-cost private behaviors through a complete scaffolding architecture, whereas interaction notifications merely prompted activity without supporting production within the interaction surface. The comparison is therefore between a complete architecture and a partial one that stops one step short.

The inference is specific: where the design provided a complete scaffolding architecture including execution-cost reduction, behavior changed; where it provided only a prompt, motivational engagement was activated but production did not follow. A version of MOBILES implementing complete social scaffolding architecture—monitor, detect, prompt, and reduce production cost—constitutes the natural comparative study required to test this falsifiable design claim. It is simultaneously a direct empirical test of the teaching presence hypothesis we advanced: whether deliberate production-stage scaffolding converts the Stage 2 perceived social presence activated by notification into enacted Stage 3 behavior.

### Social Affordance, Social Scaffolding, and the Structural Deficit

The argument that the parallel monologue reflects a design problem as much as a learner problem was articulated by Kreijns et al. (2003) over two decades ago. What the present study contributes is a different kind of evidence—behavioral trace signatures sufficiently fine-grained to partially distinguish the two competing accounts. The widespread pattern of social non-interaction documented under RQ2, comprising a 1:14.1 view-to-comment ratio and zero Stage 3 production in 80.6% of participants, is

most parsimoniously explained by the structural insufficiency identified under RQ3: 95.3% Stage 2-to-3 attrition despite 86.7% motivational activation.

The key distinction lies between *social affordance*—the provision of a comment interface, sharing features, and peer notifications—and *social scaffolding*—the monitoring of the behavioral pathway from viewing to production, the identification of collapse points, and the delivery of targeted support that reduces production costs at the moment of friction. MOBILES implemented the former but not the latter. One pattern warrants attention: Table 3 records 112 organic emoji reactions, yet Stage 3—which includes emoji reactions—yielded only 8 events across 170 post-notification views. This is consistent with the structural account: organic emoji use occurred in low-stakes browsing contexts, whereas post-notification viewing may have raised the perceived cost of even minimal responses by creating an obligation to react substantively. This interpretation is speculative; it nonetheless points to a notification design risk—that explicit invitations may paradoxically suppress the low-threshold responses they are intended to elicit.

Three intersecting factors explain why this distinction has been consistently overlooked. First, a prevalent theoretical tendency has located explanation in learners rather than in designed environments. Second, behavioral trace data at the required sequential resolution have only recently become available at scale. Third, adding a comment box requires far less investment than building a behavioral monitoring and scaffolding system. The combined effect has been a field that identified the right construct—the parallel monologue— while lacking the behavioral data needed to distinguish learner-level from design-level causes.

The mechanism documented here may characterize a broader class of technology-enhanced learning environments where social interaction is implemented primarily as an affordance rather than a scaffolded production architecture. A surface resemblance is observable across several widely deployed environments. In learning management system discussion forums, the frequently observed "lurking" problem—high enrollment accompanied by low participation—presents a similar pattern to the viewing-response gap, though whether the underlying mechanism is the same remains untested (Hrastinski, 2008). In massive open online courses, social participation features are associated with low engagement despite high content consumption (Seaton et al., 2014; Reich, 2015). In collaborative annotation tools, analogous low engagement rates have been reported (Miller et al., 2018). This surface pattern appears across cultural and regional contexts with varying mobile social learning adoption rates (Ooi et al.,

2018), suggesting it is not regionally confined. Whether these resemblances reflect the same structural mechanism—affordance without production scaffolding—or distinct context-specific causes is an empirical question; the present study proposes a diagnostic approach, grounded in sequential event-log analysis, for investigating it.

**Provisional Design Principles Grounded in Behavioral Evidence**

The following six design principles emerge from specific findings. They are provisional, grounded in a single observational study with a small and self-selected sample, and require empirical validation before treatment as established guidelines. The specific numerical parameters cited within each principle (e.g., window durations, word thresholds, weekly sequences) are illustrative operationalizations, not empirically derived optima; they are intended to anchor falsifiable tests, not to prescribe deployment values.

**Principle 1: Scaffold social production rather than merely social affordance.** This principle derives from the Stage 2-to-3 collapse rate of 95.3%. The operationalization involves detecting a post-view event on a peer annotation followed by no comment or reaction event within a configurable window (suggested: 8–12 minutes, calibrated to session length distribution), and injecting an in-situ prompt directly within the annotation viewing surface offering three content-sensitive response stems based on annotation category. For sensory annotations, stems might include "This reminds me of…", "In my country, this would be…", or "I noticed something similar at…". Such stems aim to lower production cost below the face-threat threshold while preserving authorial voice. Testable claim: a version of MOBILES implementing this principle will raise Stage 3 conversion above 4.7%.

**Principle 2: Build the bridge from spatial awareness to verbal production.** This principle derives from the spatial concentration pattern showing 25.1% of new annotations created within one kilometer of a recently viewed peer annotation, though a geographically calibrated null model is required for comparative interpretation. The operationalization involves monitoring GPS location of new annotation creation, detecting when a new annotation is created within 500 meters to one kilometer of a peer annotation viewed in the preceding 48 hours, and surfacing the connection explicitly to the user.

**Principle 3: Treat individual elaboration quality and social interaction quality as linked.** This principle derives from the finding that commented annotations were significantly more elaborate ($M$ = 42.3 words vs. 18.7 words, Cohen's $d$ = 1.01). The operationalization involves gating social sharing

behind a minimum elaboration threshold (suggested: 25–30 words combined with at least one image or one semantic tag), with below-threshold content saved privately by default accompanied by a prompt to enrich before sharing. A design risk requires acknowledgment: this threshold exceeds the mean non-commented annotation length of 18.7 words, meaning a majority of current annotations would be gated; pilot testing with progressive thresholds is advisable before full deployment.

**Principle 4: Design for the moment of production friction.** This principle derives from the observation of 86.7% Stage 2 engagement collapsing to 4.7% Stage 3 production. The operationalization involves detecting the tap event opening the comment entry field and, at this specific moment, surfacing a context-aware micro-scaffold within the text entry area itself—placeholder text or a floating tooltip rather than a modal interruption—that disappears as soon as the learner begins typing.

**Principle 5: Reduce linguistic exposure risk for second-language users.** This principle derives from qualitative accounts implicating linguistic insecurity as a Stage 2-to-3 barrier (P3, P19). The operationalization involves implementing a graduated response pathway presented as a visual ladder within the comment interface, with the lowest-cost options pre-expanded by default. The pathway would include emoji reaction, reaction stem (pre-written sentence with one fill-in-blank slot), voice memo, and free text in ascending order of production cost and linguistic exposure.

**Principle 6: Monitor escalation rather than merely presence.** This principle derives from the declining temporal Stage 3 conversion rates (Figure 2, panel b; Table 6). The operationalization involves tracking each learner's social response trajectory across weeks and applying an escalating scaffold sequence: Weeks 1–2 employ the graduated pathway from Principle 5. From Week 3, social comparison nudges are added cumulatively. From Week 4, dyadic pairing of learners who have viewed each other's content without responding is introduced. From Week 5 onward, structured peer dialogue tasks are added if Stage 3 conversion remains below 10%.

**Implications for practitioners without platform access.** Each principle has an instructor-level analogue requiring no event-log infrastructure: response stem templates as graded micro-tasks (Principle 1), minimum elaboration requirements before peer sharing (Principle 3), sentence starters for second-language learners (Principle 5), and an escalating weekly peer dialogue sequence (Principle 6).

## Limitations

**Absence of a comparison condition.** No condition with complete social scaffolding was implemented. The study identifies and characterizes the viewing-response gap but does not experimentally demonstrate that closing the structural gap would close the behavioral one. The comparative study—same population, same platform, one condition with partial scaffolding and one with complete social scaffolding architecture— constitutes the natural next step for testing the causal claim.

**Statistical non-independence of notification events.** Wilson confidence intervals assume event independence, an assumption violated by within-participant clustering with a mean of 6.8 notifications per participant. Bootstrap confidence intervals corrected for clustering are reported as the primary estimate. The correction yields a marginally wider Stage 3 interval—bootstrap 1.2% to 8.4% compared to Wilson 2.1% to 9.1%—confirming robustness.

**Selection bias in the natural reference comparison.** The 4.7-fold difference between post-notification views (4.7%) and organic views (1.0%) conflates a notification effect with a selection effect. However, the 95.3% Stage 2-to-3 collapse is interpretable regardless of how this difference is partitioned.

**Sample size and level of analysis.** Thirty-one participants limits the robustness of participant-level inferential claims. Behavioral-sequence-level claims operate at the event level (210 notification events and 497 annotation instances) and are less constrained by participant-level sample size, but remain single-deployment observational frequencies in a self-selected sample.

**Theoretical scope of the viewing-response gap.** The behavioral configuration—high Stage 2 accompanied by collapsed Stage 3—is consistent with, but not exclusive to, structural insufficiency. Face-threat deterrence, linguistic inadequacy, and perceived low communicative value could independently or jointly produce the same pattern. The evidence supports a structural account as the most parsimonious explanation but does not definitively rule out these alternative mechanisms operating in parallel.

**Spatial analysis limitations.** The spatial proximity finding of 25.1% within one kilometer is subject to urban geography bias and temporal co-presence confounds that the current data cannot resolve. Any interpretation as evidence of non-verbal peer engagement remains speculative; a geographically calibrated null model is required for empirical grounding. Future research could construct such models and replicate across different urban contexts.

**Weekly rate instability.** Stage 3 conversion rates by week (Table 6) are based on one to two events per week, yielding highly unstable estimates. The declining trend is interpreted through comparison with individual documentation behavior trends (Figure 2, panel a), rather than on the precision of individual weekly estimates.

**Researcher contact during the active use period.** Five automated weekly in-app general thematic notifications and one mid-study technical support message were delivered. Both constitute researcher-designed stimuli whose potential influence on general engagement trends cannot be fully disentangled from autonomous behavioral trajectories. Future designs could isolate general notification effects from personalized scaffolding effects through separate control conditions.

**Additional methodological constraints.** The 7-day recommendation analysis does not control for the general engagement trend; all effects are preliminary. The event-logging system does not capture abandoned comment drafts (unsubmitted text), which means the viewing-response gap inference may underestimate production attempts that were initiated but not completed; future instrumentation capturing draft activity would strengthen this inference. The researcher-designer dual role carries inherent confirmation bias risk, partially mitigated by pre-specified analyses, automated logging, and the negative-result character of the central finding.

**Absence of learning outcome measurement.** This study documents behavioral patterns but does not measure learning outcomes. Future work could couple trace analysis with validated pre- and post-intervention competence measures. The present contribution is explicitly to behavioral mechanics, not to learning outcomes.

**Onboarding as an uncontrolled scaffolding event.** The 45-minute in-person tutorial covering the social interaction interface and its intended pedagogical function constitutes a form of social scaffolding that preceded the active use period. Its potential priming effect on social interaction behavior cannot be separated from the platform's ongoing scaffolding architecture. Whether the onboarding session functioned as a front-loaded social scaffold cannot be determined without a condition in which it omits the social interaction component.

### Future Research Directions

Three lines of inquiry follow from the present study. First, replication across different technology-enhanced learning platforms would help establish whether the recommendation notification funnel collapse represents a general phenomenon; learning management system discussion forums, massive open online course peer review modules, and collaborative annotation tools offer natural settings, provided they are instrumented with event logging of sufficient granularity. Second, experimental manipulation of social scaffolding would be needed to test the causal claim that adding production-stage support reduces collapse; a comparative study contrasting a version of MOBILES with complete social scaffolding against the current partial-scaffolding version would provide direct evidence, and the proposed design principles offer concrete operationalizations for such an experiment. Third, the spatial dimension of peer engagement warrants investigation: geographically calibrated null models accounting for participant activity distribution and urban points of interest would enable rigorous testing of whether proximity effects reflect genuine social influence—and, if confirmed, motivate extension of the Community of Inquiry framework to non-verbal social presence in location-based learning.

## Conclusion

The parallel monologue has persisted as a theoretically troubling and practically underaddressed feature of technology-enhanced learning for over two decades. The present study provides the behavioral trace evidence needed to partially characterize its mechanisms—locating the collapse point between explicit social prompt and social production, and distinguishing this structural account from motivational explanations.

Three findings converge to constitute the study's argument. First, participants demonstrated measurable responsiveness to individual scaffolding recommendations across multiple behavioral dimensions, suggesting that social non-interaction is unlikely to reflect general disengagement or scaffolding unresponsiveness. Second, despite 890 documented peer views, only 63 comments were produced, yielding a 1:14.1 view-to-comment ratio consistent with widespread passive consumption. Third, of 210 interaction-targeting notifications, 93.3% were opened, 86.7% led to peer content viewing, and 4.7% of those post-notification views resulted in social interaction. Per-participant analysis confirms this collapse was widespread: 80.6% of participants individually produced zero Stage 3 events, uniformly across all engagement profiles.

This collapse is most parsimoniously explained by structural insufficiency: the social scaffolding architecture was insufficient to convert notification-activated motivational engagement into social production. We explicitly do not claim that structural design caused the parallel monologue. The behavioral evidence makes the insufficiency account more parsimonious than pure motivational disengagement, and specifies where intervention would need to be targeted.

The core conceptual distinction—between social affordance and social scaffolding—frames a design agenda for intercultural mobile learning: not adding social features but instrumenting the behavioral pathway from viewing to production—detecting collapse and intervening at the point of breakdown. More broadly, the study illustrates a methodological implication: if self-report data cannot distinguish "I did not want to interact" from "I engaged with a peer's content but nothing scaffolded the transition to response," then two decades of motivational explanations may have been asking the right questions of the wrong data.

## ACKNOWLEDGEMENTS

This work is supported by the MOBILES project funded by the Agence Nationale de la Recherche (ANR), France (grant ANR-20-CE38-0009).

## References

Belz, J. A. (2003). Linguistic perspectives on the development of intercultural competence in telecollaboration. *Language Learning & Technology*, *7*(2), 68–117.

Belz, J. A., & Müller-Hartmann, A. (2003). Teachers as intercultural learners: Negotiating German-American telecollaboration along the institutional fault line. *The Modern Language Journal*, *87*(1), 71–89.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.

Collins, L. M., Murphy, S. A., Nair, V. N., & Strecher, V. J. (2005). A strategy for optimizing and evaluating behavioral interventions. *Annals of Behavioral Medicine*, *30*(1), 65–73. https://doi.org/10.1207/s15324796abm3001_8

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319–340. https://doi.org/10.2307/249008

Deardorff, D. K. (2006). Identification and assessment of intercultural competence as a student outcome of internationalization. *Journal of Studies in International Education*, *10*(3), 241–266. https://doi.org/10.1177/1028315306287002

Dooly, M. (2011). Divergent perceptions of telecollaborative language learning tasks. *Language Learning & Technology*, *15*(2), 69–91.

Fogg, B. J. (2009). A behavior model for persuasive design. In *Proceedings of the 4th International Conference on Persuasive Technology* (pp. 1–7). ACM. https://doi.org/10.1145/1541948.1541999

Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, *2*(2–3), 87–105. https://doi.org/10.1016/S1096-7516(99)00016-1

Goffman, E. (1967). *Interaction ritual: Essays on face-to-face behavior.* Anchor Books.

Gunawardena, C. N. (1995). Social presence theory and implications for interaction and collaborative learning in computer conferences. *International Journal of Educational Telecommunications*, *1*(2–3), 147–166.

Gyasi, J. F., Ouyang, J., & Appiah-Kubi, P. (2021). Remote cross-cultural collaborative learning in K-12: A systematic review. *International Journal of Web-Based Learning and Teaching Technologies*, *16*(3), 1–17. https://doi.org/10.4018/IJWLTT.20210501.oa4

Hershkovitz, A., & Alexandron, G. (2025). A review of main issues regarding validity, reliability, generalizability, transferability, and applicability of log-based measurement of computer-assisted learning. *Research and Practice in Technology Enhanced Learning*, *20*, 020. https://doi.org/10.58459/rptel.2025.20020

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, *70*(2), 125–132.

Hrastinski, S. (2008). Asynchronous and synchronous e-learning. *Educause Quarterly*, *31*(4), 51–55.

Kramsch, C., & Thorne, S. L. (2002). Foreign language learning as global communicative practice. In D. Block & D. Cameron (Eds.), *Globalization and language teaching* (pp. 83–100). Routledge.

Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments. *Computers in Human Behavior*, *19*(3), 335–353. https://doi.org/10.1016/S0747-5632(02)00057-2

Kreijns, K., Van Acker, F., Vermeulen, M., & Van Buuren, H. (2014). Community of Inquiry: Social presence revisited. *E-Learning and Digital Media*, *11*(1), 5–18. https://doi.org/10.2304/elea.2014.11.1.5

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Liaw, M. L. (2006). E-learning and the development of intercultural competence. *Language Learning & Technology*, *10*(3), 49–64.

McKenney, S., & Reeves, T. C. (2019). *Conducting educational design research* (2nd ed.). Routledge.

Miller, K., Lukoff, B., King, G., & Mazur, E. (2018). Use of a social annotation platform for pre-class reading assignments in a flipped introductory physics course. *Frontiers in Education*, *3*, 8. https://doi.org/10.3389/feduc.2018.00008

O'Dowd, R. (2009). The many faces of telecollaboration. *Language Learning & Technology*, *13*(2), 2–7.

O'Dowd, R., & Ritter, M. (2006). Understanding and working with "failed communication" in telecollaborative exchanges. *CALICO Journal*, *23*(3), 623–642. https://doi.org/10.11139/cj.23.3.623-642

Ooi, K. B., Hew, J. J., & Lee, V. H. (2018). Could the mobile and social perspectives of mobile social learning platforms motivate learners to learn continuously? *Computers & Education*, *120*, 127–145. https://doi.org/10.1016/j.compedu.2018.01.015

Reich, J. (2015). Rebooting MOOC research. *Science*, *347*(6217), 34–35. https://doi.org/10.1126/science.aaa5357

Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences*, *13*(3), 273–304. https://doi.org/10.1207/s15327809jls1303_2

Sadallah, M., & Lefevre, M. (2025). L'application MOBILES: favoriser le partage d'expériences urbaines entre étudiants internationaux. In 12ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain (pp. 582-588).

Sadallah, M., & Lefevre, M. (2025a). Personalized Recommender System for Improving Urban Exploration and Experience Documentation of International Students. In CSEDU (2) (pp. 923-930). https://doi.org/10.5220/0013202800000393

Seaton, D. T., Bergner, Y., Chuang, I., Mitros, P., & Pritchard, D. E. (2014). Who does what in a massive open online course? *Communications of the ACM*, *57*(4), 58–65. https://doi.org/10.1145/2500876

Settouti, L. S., Prié, Y., Mille, A., & Marty, J. C. (2009). A trace-based system for technology-enhanced learning systems personalisation. In *Proceedings of the 9th IEEE International Conference on Advanced Learning Technologies* (pp. 112–116). IEEE. https://doi.org/10.1109/ICALT.2009.18

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Houghton Mifflin.

Shavelson, R. J., & Towne, L. (Eds.). (2002). *Scientific research in education*. National Academy Press.

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, *57*(10), 1380–1400. https://doi.org/10.1177/0002764213498851

Thorne, S. L. (2003). Artifacts and cultures-of-use in intercultural communication. *Language Learning & Technology*, *7*(2), 38–67.

Ünlüsoy, A., Leander, K., & de Haan, M. (2022). Rethinking sociocultural notions of learning in the digital era: Understanding the affordances of networked platforms. *E-Learning and Digital Media*, *19*(4), 356–373.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Harvard University Press.

Ware, P. D. (2005). "Missed" communication in online communication: Tensions in a German-American telecollaboration. *Language Learning & Technology*, *9*(2), 64–89.

Ware, P. D., & Kramsch, C. (2005). Toward an intercultural stance: Teaching German and English through telecollaboration. *The Modern Language Journal*, *89*(2), 190–205.

Webb, T. L., & Sheeran, P. (2006). Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychological Bulletin*, *132*(2), 249–268. https://doi.org/10.1037/0033-2909.132.2.249