APPLING
International Applied Linguistics Research Conference

https://doi.org/10.29140/9781763711686-10

Revisiting the Case Study for the Age of AI: Multi-Case Qualitative Designs and Agency as a System Property

Jinru Luo, Rongrong Dong
Southern University of Science and Technology, China

Abstract

Generative AI in second language writing classrooms raises new research questions and, less obviously, strains applied linguistics' methodological habits. Studies of learner agency in human-AI interaction typically treat agency as something the individual learner possesses, coding learner moves and triangulating self-reports for convergence. Cukurova's (2026) reframing of agency as a system property, an emergent feature of the human-AI exchange rather than a trait inside either party, unsettles this practice: if the construct lives at the system level, triangulating individual-level indicators measures the wrong unit. We argue that the multiple-case study, revisited with a proper framework, is well suited to such system-level questions, and illustrate this with ten L2 writers completing an AI-assisted argumentative essay. Chat logs served as the primary data, with screencasts and stimulated-recall interviews chosen for system-level rather than convergent triangulation; 85 episodes were coded on a four-level interaction framework. Cross-case comparison showed AI tools with similar operating mechanisms producing markedly different system-level agency configurations: specification agency (Level 2) dominated (85.9%), reasoning agency (Level 3) was rare (8.2%) and confined to one instructional group, and co-regulative interaction was absent. Comparing agency-as-observed with agency-as-experienced exposed systematic divergence, a finding rather than an error. We close with design recommendations and argue the field now needs intervention studies testing whether agency can be developed.

E-mail: luojr@sustech.edu.cn, dongrr@sustech.edu.cn

In Y. Wang, J. Matthews, & B. Zou (Eds.), *Proceedings of the 2nd International Applied Linguistics Research Conference, 2026* (pp. 77–82). Castledown.

Keywords: multiple-case study, agency, system-level configuration, human-AI interaction framework

Background

The presence of generative AI in second language (L2) writing classrooms confronts applied linguistics with new research questions and, less obviously, with a challenge to its methodological assumptions. When a learner writes with an AI assistant, cognition, decision-making, and text production are distributed across a human-machine system, yet the field's dominant analytic habit remains individual-centred. Research on learner agency shows this strain especially clearly (Cukurova, 2026).

For situated, process-level questions, the case study has long been applied linguistics' method of choice. Over time it has grown from the single, holistic case toward multiple-case designs, where comparison across cases, not depth within a single one, becomes the analytic focus (Yin, 2018). Researchers have used such designs to study language socialisation, identity, and, more recently, technology-mediated learning, because they keep the contextual detail while still letting patterns be checked across bounded instances.

Learner agency is a natural match for these designs. Understood as the socioculturally mediated capacity to act (Ahearn, 2001), agency is relational and bound to its setting, and multi-case work has shown how differently it operates across learners, contexts, and tasks (Tao & Gao, 2026). Studies that carry this tradition into human-AI interaction usually draw on several data streams at once: users' AI chat logs, process recordings, and self-report from interviews or surveys (Deng et al., 2026; Lee et al., 2025). They also tend to share one assumption. Agency is treated as something the learner possesses: researchers code individual moves, gather self-reports, and triangulate these individual-level indicators for convergence.

It is this assumption that Cukurova (2026) calls into question. If agency is instead a system property, an emergent feature of the human-AI interaction rather than a trait sitting inside either party, then it is enacted through the dynamics of the exchange and the affordances of the system. And if the construct lives at the level of the system, triangulating individual-level indicators for agreement is measuring the wrong thing. Two further problems sit alongside this one. Most existing studies lean on self-report, and few track enacted agency through data one can actually trace. More worrying, the recent evidence is largely descriptive: it tells us that agency matters, not how it might be built. We still have no systematic account of how to cultivate learner agency for AI interaction, and almost no intervention studies asking whether agency

can be taught, or whether stronger agency with AI actually improves what students learn.

This paper argues that the multiple-case study, revisited with a proper framework, is well suited to address these system-level questions, and illustrates the argument with a study of L2 writers interacting with AI writing assistants.

Methodology

The illustrative study used an exploratory multiple-case qualitative design (Yin, 2018). Each case was one learner's full engagement with a single AI-assisted argumentative writing task. Ten STEM undergraduates (aged 19–21, intermediate English) at a Chinese university wrote a roughly 400-word argumentative essay in English within a one-hour session, arguing a position on renewable energy while integrating a culturally specific Chinese source text into the essay. Students chose their own AI tools, all with similar operating mechanisms. They came from two teaching backgrounds that happened to differ, a data-driven learning module (five cases) and an AI-prompt-based module (five cases), which gave us a ready basis for comparison across cases.

We chose three data sources with system-level analysis in mind, rather than for the usual convergent triangulation (Table 1). The learner-AI chat logs were the primary data, since they record how agency is co-constructed turn by turn inside the system. Screencast recordings captured what happened in the writing process outside the chat window. Post-task semi-structured interviews, using stimulated recall, gave us access to how learners experienced their own agency. We segmented the chat logs into 85 episodes and coded them with a four-level human-AI interaction framework[1] adapted from Cukurova (2026), treating the line between specifying what an essay should contain (Level 2) and directing how the argument should be reasoned (Level 3) as the threshold that mattered most. Analysis moved from within-case to cross-case. Triangulation here performed double duty: it corroborated the codes, but it also enabled setting agency-as-observed in the interaction trace against agency-as-experienced in interview, a comparison deemed necessary from the system-property view.

Representative Results

The multiple-case design proved indispensable. Comparing across cases showed that the AI tools with similar operating mechanisms produced quite different system-level agency configurations from one learner to the next, the

kind of variability that a single-case study, or a variable-centred quantitative design, would simply miss. In the aggregate, specification agency (Level 2) prevailed, accounting for 85.9% of episodes: students steered content, structure, and register while handing the core argumentative reasoning to the AI. Reasoning agency (Level 3) was uncommon (8.2% in total) and turned up only in the data-driven learning group. Co-regulative interaction (Level 4) did not appear at all (Table 2). The group contrast, reasoning-level configurations present in one group and absent in the other, is itself a product of cross-case comparison rather than of any single case.

Table 1. Three Data Sources and Their System-Level Analytic Function

Data source	What it records	System-level analytic function
Learner-AI chat logs (primary)	Turn-by-turn human-AI exchange	Captures co-construction of agency within the system; the unit for episode-level coding
Screencast recordings	On-screen writing process beyond the chat window	Reveals agency enacted outside the chat (independent drafting, source reading, external verification)
Post-task interviews (stimulated recall)	Learner's account anchored to the interaction trace	Accesses agency-as-experienced for deliberate comparison with agency-as-observed

Reading the three sources together produced things no single source could. Chat logs and screencasts side by side allow us to separate substantive epistemic strategies (checking AI output against primary sources) from operational ones (playing one AI off against another). The difference lies in what students hold the AI to account against, not merely in whether they scrutinise it. The most striking result came from placing agency-as-observed next to agency-as-experienced: several learners talked about themselves as (argumentatively) directive, while their traces showed them working at the level of specification throughout. A convergence-seeking, individual-level design would read this mismatch as measurement error. Within a system-property frame it becomes a finding in its own right, telling us something about the gap between how agency feels and how it is enacted.

Table 2. Distribution of Human-AI Interaction Levels by Instructional Group

Interaction level	Group A: data-driven (50 episodes)	Group B: prompt-based (35 episodes)
Level 1 – Transactional	0 (0%)	5 (14.3%)
Level 2 – Situational	43 (86.0%)	30 (85.7%)
Level 3 – Operational	7 (14.0%)	0 (0%)
Level 4 – Co-regulative	0 (0%)	0 (0%)

Discussion

These results matter methodologically. First, once agency is treated as a system property, a multiple-case design stops being just one option among many and becomes more of a requirement. Only comparison across cases brings out the configural variability that a single tool generates across learners, and only comparison keeps us from mistaking the most common pattern for the general rule. Second, triangulation needs rethinking. Its job is not to confirm one individual-level measure with agreeing indicators, but to set agency-as-observed against agency-as-experienced and treat divergence as evidence rather than noise. Third, coding schemes should stay close to the interaction: the episode of co-construction, not the isolated learner move, is the unit worth analysing, and the threshold of interest lives in the interactional dynamic itself.

Practically, then, researchers extending multi-case designs to system-level phenomena should make traceable interaction data (chat logs, screencasts) primary rather than supplementary, anchor self-report to that trace through stimulated recall, and write coding rules a second reader can apply and audit.

Finally, descriptive multi-case evidence, this study included, has by now established that agency in human-AI interaction matters and varies in patterned ways. What is still lacking is a systematic sense of how to develop it. The methodological toolkit sketched here, multi-case, interaction-primary and triangulated at the level of the system, is exactly what intervention studies will need if they are to test whether agency can be taught and whether stronger agency with AI improves learning. Adapting these multi-case study methods to the emergent, system-level character of AI-mediated learning can be a first step in that direction.

Note

Four-level framework adapted from Cukurova (2026):

Level 1 Transactional: students accept AI output with minimal evaluation of argument quality.

Level 2 Situational: student specifies what the essay should contain; AI supplies the reasoning.

Level 3 Operational: student directs the argumentative substance (claims and how to reason) and evaluates AI output against their own standards.

Level 4 Co-regulative: sustained multi-turn exchange demonstrably reshapes students' argumentative thinking.

References

Ahearn, L. M. (2001). Language and agency. *Annual Review of Anthropology, 30*(1), 109–137. https://doi.org/10.1146/annurev.anthro.30.1.109

Cukurova, M. (2026). Agency as a system property in human–AI interaction in education. *British Journal of Educational Technology*, *57*(4), 1065–1070. https://doi.org/10.1111/bjet.70060

Deng, C., Lin, S., Tong, E. K.-m., & Ma, J. (2026). Leveraging hybrid intelligence: How teacher agency influences behavioural engagement in human-AI collaborative writing feedback practices. *International Journal of TESOL Studies, 8*(1), 82–107. https://doi.org/10.58304/ijts.251108

Lee, M., Tan, R. J. Y., Chen, D.-T., Huang, J. S., & Hung, D. W. L. (2025). Exploring interactions between learners and ChatGPT from a learner agency perspective: A multiple case study on historical inquiry. *Education and Information Technologies, 30*(6), 7167–7189. https://doi.org/10.1007/s10639-024-13114-y

Tao, J., & Gao, X. (2026). Learner and teacher agency. *TESOL Quarterly*, *60*(2), 586–611. https://doi.org/10.1002/tesq.70112

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE.