Language Learning & Technology
ISSN 1094-3501 CC BY-NC-ND

2026, Volume 30, Issue 1
pp. 1–25

ARTICLE

Agency to autonomy in mediated data-driven learning: A longitudinal study

Soyeon Moon, Seoul National University

Sun-Young Oh, Seoul National University*

Abstract

Drawing on Sociocultural Theory (SCT), this longitudinal case study examines how secondary EFL learners exercise agency, engage, and develop autonomy within digitally mediated data-driven learning (DDL). It began with a 16-month compilation of a local learner corpus, followed by 7 months of four pedagogical phases—paper-based, hands-on, customized, and self-directed—progressing from teacher-guided instruction to autonomous engagement. Multimodal data reveal that learner autonomy in DDL goes beyond technical mastery of corpus tools; it unfolds as a socially, emotionally, and contextually mediated process shaped by identity, affect, and strategic growth. Suho, the Planner, progressed through structured, scaffolded support, while Jimin, the Seeker, exercised exploratory agency through multimodal engagement aligned with personal interests. Theoretically, the study extends SCT by demonstrating reciprocal mediation in digital learning environments and highlighting the interdependence between agency, identity, and affect. Pedagogically, findings underscore the central role of teacher expertise—particularly adaptive technological, pedagogical, and content knowledge—in making DDL accessible and sustaining learner engagement. Methodologically, the study highlights the value of longitudinal, multimodal, and ecologically situated approaches for capturing the process of agency, engagement, and autonomy.

Keywords: data-driven learning; learner agency; autonomy; sociocultural theory

Language(s) Learned in This Study: *English*

APA Citation: Moon, S., & Oh, S.-Y. (2026). Agency to autonomy in mediated data-driven learning: A longitudinal study. *Language Learning & Technology*, *30*(1), 1–25. https://doi.org/10.64152/10125/73665

Introduction

Originally proposed by Tim Johns (1991), Data-Driven Learning (DDL) is recognized as a groundbreaking approach in language education. By engaging learners directly with corpus data (Boulton, 2009), DDL fosters both inductive discovery and deductive reasoning, particularly when effectively scaffolded (Flowerdew, 2008). Its potential to enhance language awareness and learner autonomy aligns with constructivist and sociocultural perspectives on learning (O'Keeffe, 2021).

Despite its benefits, DDL remains underused in secondary contexts, partly due to constraints such as time, curriculum coverage, and usability of existing corpus tools (Crosthwaite & Baisa, 2024). Native-speaker corpora dominate, while learner-generated or graded corpora are underexplored, limiting pedagogical accessibility (Boulton & Vyatkina, 2021). Many studies focus on procedures or short-term outcomes, often overlooking how learners internalize corpus practices—a gap especially consequential where sustained scaffolding is needed (see also Boulton & Vyatkina, 2021).

A persistent challenge lies in the black box of learner development in DDL. Few studies examine how learners interpret corpus data, how mediation shapes learning, and how autonomy emerges (Crosthwaite & Boulton, 2024). This gap is pronounced in secondary classrooms, where learners depend on teacher scaffolding, peer collaboration, and affective support. Most studies offer snapshots rather than

*** Corresponding Author:** Sun-Young Oh, sunoh@snu.ac.kr

developmental perspectives (Boulton & Vyatkina, 2021).

Sociocultural Theory (SCT) provides more than an interpretive lens—it offers a theoretical key to this black box. Learning is socially mediated through teachers, peers, and tools (Vygotsky, 1978), and applied to DDL, SCT traces learners' progression from supported participation to self-regulated practice. Yet, longitudinal classroom studies of adolescent DDL engagement remain scarce, and teachers often hesitate to introduce corpus tools to novice learners (Chambers, 2019), underscoring the need to examine how sustained mediation fosters independent use and autonomy.

To address these gaps, this study follows two Korean high school learners across four pedagogical phases, from teacher-guided tasks to self-directed exploration. Grounded in SCT, it investigates how interactions with teachers, peers, and tools supported the internalization of corpus practices and the emergence of learner autonomy. This approach highlights the reciprocal, context-sensitive nature of mediation in digitally mediated environments (Lantolf & Thorne, 2006; Wertsch, 1998). Situating adolescent DDL engagement within SCT, the study shows how agency is exercised, engagement unfolds, and autonomy develops—insights that short-term designs cannot capture.

Literature Review

This review applies SCT to DDL research, highlighting gaps in process-oriented studies and framing the current study as a response.

A Sociocultural Framework for Learner Agency, Autonomy, and Engagement

This study adopts a sociocultural perspective grounded in Vygotsky's (1978) view of learning as socially mediated and contextually situated. Within this framework, learners engage with more capable others and cultural tools while operating in the Zone of Proximal Development (ZPD), gradually internalizing collaborative activity as self-regulation. Building on subsequent SCT elaborations (e.g., Lantolf & Thorne, 2006) and related SLA research, we use this perspective as a coherent lens to differentiate agency, autonomy, and engagement, thereby reducing potential conceptual overlap.

Agency refers to the socioculturally mediated capacity to act (Ahearn, 2001, p. 112), encompassing the ability to make choices, interact with others, and use cultural tools within contextual and interpersonal settings. Lantolf and Thorne (2006) link agency to intentionality, goal orientation, and the capacity to interpret and assign relevance to events. In classroom contexts, Van Lier (2008) highlights learners' capacity to notice, interpret, and transform opportunities for action, while the Douglas Fir Group (2016) frames agency as both a means and a goal of language learning, enabling learners to shape their learning conditions and, ultimately, themselves.

Autonomy represents the enacted form of agency, emerging progressively through sustained social and material mediation. While all learners initially possess agency, autonomy develops over time, reflecting the capacity to direct one's own learning without constant external support. From an SCT perspective, autonomy is socially situated, relational, and developmental (Toohey & Norton, 2003), contrasting with Self-Determination Theory, which treats autonomy as an innate psychological need (Deci & Ryan, 2000).

Engagement is a socially mediated, observable process central to learning (Philp & Duchesne, 2016). Within SCT, engagement is dynamic and realized through interaction with tools, peers, and self-regulatory resources, rather than simply time spent on tasks. Behavioral, cognitive, and emotional dimensions are enacted through mediated actions such as corpus searches, collaborative dialogue, or private reflection. Engagement functions as the mechanism for exercising agency and supports the progressive emergence of autonomy.

In summary, agency represents potential, engagement enacts this potential through mediated activity, and autonomy emerges as the realized, self-regulated outcome. With this theoretical framework established, we now turn to its application in DDL, where understanding learner engagement is crucial for tracing the

exercise of agency and the development of autonomy.

DDL: From Potential to Mediated, Process-Oriented Understanding

DDL allows learners to engage with authentic language data via corpora, fostering language awareness and supporting the development of learner autonomy (Boulton, 2009; O'Keeffe, 2021). Despite its promise, implementation at the secondary level remains limited, and research has largely focused on short-term interventions in tertiary settings (Boulton & Vyatkina, 2021). Although some adolescent-focused studies exist (e.g., Moon & Oh, 2017), they are few in number and offer little insight into learners' underlying processes. Reflecting this, process-focused implementation is scarce: a systematic review of CALL studies (2011–2015) found that most work concentrated on tertiary-level learners and writing outcomes, leaving the "how" and "why" of learning largely unexplored (Pérez-Paredes, 2019).

To understand how learners transform corpus potential into enacted understanding, DDL can be conceptualized as a process-oriented, mediated activity, in which learners actively interpret, negotiate, and refine linguistic patterns over time. Rather than treating corpora as static repositories, this approach foregrounds learners' unfolding interactions with tools, peers, and teachers, highlighting the role of mediation in fostering engagement and gradually developing autonomy (Flowerdew, 2008; Huang, 2011; Park, 2012). Through repeated cycles of hypothesis-forming, noticing, and self-correction, learners transform corpus-derived possibilities into enacted learning, a process captured by microgenetic analysis (Lantolf & Thorne, 2006; Swain, 2006). This perspective emphasizes that learning unfolds dynamically over time, underscoring the need to examine moment-to-moment engagement rather than focusing solely on outcomes (Park, 2012).

Existing SCT-aligned DDL research further demonstrates how pedagogical mediation fosters autonomy. Flowerdew (2008) notes that raw corpus data can overwhelm learners, and meaningful learning occurs when data is scaffolded through well-designed tasks and peer discussion. Huang (2011) positions learners as "language researchers," framing learning as mediated through intra- and interpersonal dialogues. Park (2012) traces moment-by-moment learner–corpus interactions, revealing the progressive nature of learning and emphasizing process over product.

However, gaps remain. Adolescent learners are underrepresented, despite their reliance on teacher guidance, peer negotiation, and emotional scaffolding in structured, assessment-oriented environments. Languaging and microgenetic analyses (Swain, 2006) are rarely applied to adolescent DDL, limiting insight into real-time internalization. Teachers' mediating roles also require further attention: effective DDL implementation hinges on Technological Pedagogical Content Knowledge (TPACK), which many teachers lack in time-constrained classrooms (Chambers, 2019).

Building on these gaps, this study emphasizes capturing how adolescent learners exercise agency through sustained, context-sensitive mediation. Integrating SCT with Schmidt's (2023) *Planner*/*Seeker* typology provides a framework for understanding both the mediational processes that support autonomy development and the learner differences that shape corpus engagement. Guided by these considerations, the study addresses the following research questions:

1. How do adolescent EFL learners exercise agency, manifest engagement, and develop autonomy through mediated DDL?

2. How do different forms of mediation shape learners' agency, engagement, and autonomy in DDL?

The Study

Grounded in SCT, this longitudinal case study examines how different forms of mediation shaped the development of learner autonomy in DDL over time.

Context and Participants

This longitudinal study, informed by a 2016 pilot, was conducted at a public high school in Seoul, Korea, with IRB approval (Seoul National University IRB No. 21017052/002-008) and participant/parental consent. A local learner corpus was compiled from writing samples collected between April 2017 and July 2018 from two cohorts of 100 first-year students each.

In March 2018, all students in the 2018 cohort completed a Career Aptitude Test and participated in three preliminary corpus-based grammar sessions delivered during regular English classes to introduce basic DDL concepts.[1] Seven students from this cohort subsequently volunteered for an after-school DDL class and formed the primary participant group for the focal cases.[2] The study was co-designed by both authors; the first author acted as instructor, while the second provided methodological consultation without direct instructional contact.

Among the four students who completed all instructional phases, two focal participants—Suho and Jimin (pseudonyms)—were selected for in-depth analysis. Their contrasting learning styles aligned with Schmidt's (2023) *Planner*/*Seeker* typology, illustrating diverse developmental paths within the same context.

Suho, a reflective and cooperative student, began learning English at age eight without overseas experience. Despite his diligence in class, he rarely used English outside school and described it as "a great mountain to conquer" (interview, July 23, 2018). Identified as a *Planner*, he was less proficient and often anxious about grammar. He described feeling "lost in complex rules" and aimed to move beyond rote memorization to "find a clearer, more logical way to understand grammar." This preference for structured, deliberate learning framed his early struggles and motivated his engagement with the intervention.

Jimin, quiet and introspective with a strong interest in art and design, had higher proficiency from living abroad and described English as "fun and breezy" (interview, July 23, 2018). She nonetheless showed gaps in formal grammar terminology and aimed to "make English feel creative again," noting that school-based instruction had been "boring and full of rules," reflecting her *Seeker*-oriented engagement throughout the study.

Despite differences in background and style, both students struggled with grammar and sought to move beyond fragmented, rule-heavy instruction, providing a meaningful basis for comparing how agency was exercised, engagement unfolded, and autonomy emerged within the DDL environment.

Corpora

Participants worked with four corpora matched to their proficiency and learning context. Two standard online corpora—the Corpus of Contemporary American English (COCA) and the British National Corpus (BNC)—served as general reference resources. A graded reader corpus, compiled from three publishers' e-books (103 files; 17,449 types; 1,361,058 tokens), was categorized into lower, upper, and combined levels for easier navigation.[3]

The local learner corpus, TLC, includes 712 learner-produced texts (9,952 types; 112,485 tokens) on topics such as self-introduction and future goals, reflecting interlanguage patterns for guided analysis and error correction. Finally, a pedagogic corpus (251 files; 38,032 types; 829,383 tokens) was compiled from textbooks and national exam passages to reinforce taught language points.

Four Phases of DDL Implementation

Sustained teacher scaffolding was essential, as DDL requires learners to integrate new content, digital tools, and a novel pedagogical approach simultaneously (Boulton, 2009). Accordingly, the main DDL intervention—the primary research context for this study—was organized into four phases, progressing from teacher-led instruction to self-directed corpus learning. Initially offered as an after-school class, the program gradually became independent and self-paced over the full intervention period (Aug 2018–Feb

2019), concluding with post-study interviews in March 2019.

The instructional design drew on a 2016 pilot study, which tested DDL feasibility for secondary learners and informed the development of age-appropriate, corpus-based materials. For the main intervention, these materials were refined using the local learner corpus and supplemented with graded-reader and pedagogic corpora.

Before the after-school class began, the 2018 cohort completed three preliminary corpus-based lexicogrammar sessions during regular English classes to build initial familiarity with corpus-informed noticing. The after-school DDL program then unfolded across four sequential phases, from highly scaffolded instruction to increasingly self-directed work. Table 1 summarizes the pedagogical goals, representative tasks, and mediational scaffolds for each phase, illustrating the gradual transfer of responsibility from teacher to learner in line with SCT principles (see Appendix A for full details).

Table 1

Illustrative Tasks and Mediational Scaffolds Across Four DDL Phases

Phase & duration	**Pedagogical goal**	**Representative tasks**	**Primary mediators**	**Transition to next phase**
Phase 1: Paper-based DDL (Aug–Sep 2018)	Introduce pattern recognition and basic corpus concepts without the cognitive load of technology.	Learners analyze printed, large-font concordance lines from native-speaker and pedagogic corpora to identify patterns of perceptual verbs (e.g., *see* / *watch* / *hear* + *object* + *bare infinitive*). They mark patterns, discuss findings in small groups, and collaboratively revise problematic learner examples through comparison.	Human: Teacher (explicit modelling, pointing out salient patterns); peers (group discussion, simplified examples) Materials: Printed concordance lines; colored pens	Learners develop the ability to identify pre-selected patterns under teacher guidance, preparing them to handle digital corpus tools.

Phase 2: Hands-on DDL (Oct 2018)	Develop foundational technical skills with corpus software and online interfaces in a structured setting.	With teacher guidance, learners use AntConc to generate concordance lines for light verbs (e.g., *do*, *have*, *make*) from both learner and graded reader corpora, comparing frequencies and usage patterns in pairs and groups.	Human: Teacher (technical support, clarifying queries); peers (collaborative problem-solving) Tools: AntConc, MonoConc, COCA, BNC interfaces	Learners gain basic corpus-search literacy and confidence using tools, enabling them to explore self-selected linguistic targets.
Phase 3: Customized DDL (Nov 2018–Jan 2019)	Foster learner agency by allowing learners to select their own linguistic focus and materials based on personal interest.	Jimin and her group chose to investigate vocabulary from a favorite movie. They used COCA to explore collocates and usage contexts of words such as *enlightening* and *top-notch*, engaging in collaborative meaning negotiation.	Human: Teacher (strategic guidance, e.g., on dictionary use); peers (co-constructing interpretations, hypothesis testing) Symbolic/Materials: Personal interest (movie), online dictionaries, corpora	Learners demonstrate the ability to generate their own research questions and apply corpus tools strategically, preparing for full independence.
Phase 4: Self-directed DDL (Jan–Feb 2019)	Promote autonomy and the internalization of DDL strategies through independent, self-regulated projects.	Suho independently investigates auxiliary verbs using COCA, segmenting concordance lines for note-based analysis and later refining his approach with teacher feedback.	Human: Teacher (asynchronous email feedback) Internal: Private speech (think-aloud protocols), prior linguistic knowledge (e.g., familiar novels) Tools: Screen recording software, email	Learners manage their own corpus-based inquiry—from topic selection to analysis and reflection—demonstrating internalized DDL literacy.

Phase 1 introduced paper-based DDL over five sessions, emphasizing noticing grammatical patterns in large-print concordance lines (Appendix B). With teacher guidance, learners progressed from identifying patterns in native-speaker concordance lines to comparing learner and pedagogic corpora, refining lexical choices and discussing common usage patterns, such as the frequent use of *make* where *let* might also be appropriate. Following Phase 1, learners engaged in ongoing extensive reading of graded novels, which continued throughout Phases 2–4, alongside DDL activities. Phase 2 involved exploring COCA, BNC, Lex Tutor, AntConc, and MonoConc across three guided sessions, with a focus on (lexico)grammatical features. Phase 3 allowed learners to select a focus—grammar or vocabulary—and analyze self-selected materials (novels, films, web texts) using online resources. The phase concluded with class presentations during regular lessons, peer feedback, and reflection. Phase 4 comprised independent, corpus-based projects aligned with learners' interests, such as TED Talks, novels, and grammar references. Four students completed this phase, submitting reflective video reports and receiving asynchronous feedback.

Data Collection and Analysis

Data for the two focal cases were collected across multiple stages to capture their evolving engagement with corpus tools. Observations, recordings, interviews, and learner artifacts served as the primary data sources. To ensure instrument validity and clarity, a pre-survey was administered to all first-year students, with several items revised based on feedback. The final survey, together with a diagnostic grammar test aligned with the school curriculum, was then given to the seven volunteer learners. Interview protocols were also refined through pilot testing to enhance clarity and flow. Table 2 summarizes the data collection methods used in the study, detailing the instruments and procedures that supported this multi-sourced, triangulated design.

Table 2

Data Collection Methods

Phase	Data collection methods
Pre-study	pre-survey, diagnostic grammar test (30′), pre-study individual interview (20′)
1	observation (50–100′, 5 sessions per learner), impromptu individual interview (15′), teacher and learner logs, task cards, learning materials
2	observation, screen and audio recording (100–120′, 3 sessions per learner), impromptu individual interview (15′), intermediate group interview (30′), teacher and learner logs
3	observation, screen and audio recording (60–100′, 3 sessions per learner), intermediate individual interview (20′), whole group reflection (60′), teacher and learner logs
4	screen and audio recording, think-aloud protocol (31–62′, 4–5 sessions per learner), emails, text messages, teacher and learner logs
Post-study	post-study individual interview (30′)

The first author served as both instructor and researcher, enabling rich, naturalistic data collection. This dual role required reflexivity to minimize potential bias, such as an unconscious desire for pedagogical success. To mitigate this, pedagogical reflections were recorded separately in a research journal. The second author, without instructional contact, independently coded 10% of transcripts to ensure reliability; discrepancies were discussed and major analytical decisions were jointly reviewed.

Lessons and breaks were recorded using three camcorders to capture both structured and informal interactions. From Phase 2 onward, screen and audio recordings via oCAM documented real-time activity, including mouse movements and window changes. Transcripts were annotated consistently: ">>" marked a new DDL episode, ">" a processing step, and tools were labeled in brackets (e.g., [C] for

COCA). For instance, "[B] *popping*: 273" indicates that *popping* appeared 273 times in the BNC. Phase 4 included think-aloud protocols, in which learners gradually verbalized their thoughts more freely and in greater detail. Semi-structured interviews were conducted before, during, and after the intervention, with the final stage probing deeper insights. Supplementary data included teacher and learner logs, DDL materials, task cards, emails, and text messages.

Data transcription proceeded alongside collection, beginning with rough drafts and later expanded to detailed versions incorporating gestures, tone, and facial expressions to reflect learner mediation. An iterative analysis (Copland & Creese, 2015) combined open coding in Atlas.ti (version.8.4.26.0) with manual axial and selective coding until saturation (Strauss & Corbin, 1998). Guided by SCT, Phase 2 analysis focused on learner mediation because the DDL episodes showed sufficiently consistent structures for systematic comparison. Mediation was coded by domain and initiator (learner, teacher, peer[s]), including behaviors such as private speech, questioning, peer support, and direct corpus interaction. Full frequency tables are provided in Appendix C to support transparency and credibility.

Sensitive data were securely stored per institutional ethics guidelines. Trustworthiness was reinforced through participant clarification (Merriam, 1998), double-checking by MA-level non-experts, and expert review, ensuring balanced and credible interpretations of learner engagement and mediation.

Findings: Developmental Trajectories of Suho and Jimin

This section provides a comparative overview of Suho and Jimin. Suho shows a *Planner* profile using structured, analytical strategies, while Jimin displays a *Seeker* profile shaped by exploratory, interest-driven engagement. Table 3 summarizes their trajectories. This overview prepares the SCT-informed analysis of how their agency was exercised, engagement took shape, and autonomy developed over time.

Table 3

Comparative Developmental Trajectories of Suho (Planner) and Jimin (Seeker) Through Mediated DDL

Phase & goal	Learner	Key behaviors & challenges	Primary mediators & manifestations of agency, engagement, and autonomy
Phase 1: Paper-based DDL Goal: Introduce pattern recognition	Suho	Initially overwhelmed by data density; hesitant in group discussions; highly anxious about being correct.	Other-regulation: Teacher (direct scaffolding, emotional support); peers (structured examples). Agency: Expressed as questioning for clarification and affective self-regulation (private speech, e.g., "Don't shrink").
	Jimin	Initially anxious about grammar terms; hesitant to speak in group discussions; often asked to go last (e.g., "Can I do it later?", "I'm embarrassed").	Other-regulation: Teacher (verification and confirmation); minimal peer mediation. Agency: Independently explored paper-based corpus materials, solved tasks autonomously; low private speech.

Phase 2: Hands-on DDL Goal: Develop technical skills	Suho	Struggled with corpus software; avoided trial-and-error; relied heavily on a peer for technical help.	Other-regulation: Peer (technical support from Youngmo); teacher (process guidance). Internalization: High frequency of private speech (languaging) to process information and manage tasks.
	Jimin	Explored graded reader corpora; engaged with content independently; showed increasing confidence in tool usage.	Other-regulation: Teacher (confirmation); peers (prompts for exploration). Agency: Increased private speech and active engagement with graded-reader corpus via LexTutor, collaborating more with peers.
Phase 3: Customized DDL Goal: Foster learner agency	Suho	Chose to focus on challenging grammar points; engaged in collaborative inquiry with a peer, taking initiative in searches.	Collaborative Agency: Co-constructed knowledge with peer (Youngmo), shifting from receiving help to proposing joint action (e.g., "Shall we find it ourselves?").
	Jimin	Shifted from grammar to collaborative vocabulary work; led peer discussions; maintained curiosity-driven exploration.	Collaborative Agency: Negotiated meaning with peers within ZPD; minimal teacher intervention; integrated corpus and familiar content.
Phase 4: Self-directed DDL Goal: Promote autonomy	Suho	Initially hesitated due to fear of making mistakes; guided through teacher encouragement and peer modeling; developed decontextualized sentence-segmentation strategy.	Self-regulation: Teacher-provided affective mediation (e.g., email) and peer modeling helped overcome an initial block, followed by the development and refinement of cognitive strategies, showing emerging autonomy.
	Jimin	Explored corpora and TED talks independently; applied strategies across contexts; maintained sustained engagement and curiosity.	Self-regulation: Applied strategic exploration independently; integrated corpus and multimedia content; demonstrated autonomy and strategic agency.

Suho: "Oh, yeah, this is the very sentence! Like this!"

Suho exemplifies a *Planner* learner, favoring structured, step-by-step approaches to language learning (Schmidt, 2023). In Phase 1, he faced cognitive overload and hesitated during group discussions ("Where's the sentence you mentioned?") (1-1).[4] Teacher mediation—circling key phrases and prompting noticing—helped him adopt analytic strategies beyond his independent capability, reflecting ZPD-guided learning (Vygotsky, 1978), as evident in moments of realization such as "Aha!" (1-1). A diagnostic grammar test (July 20, 2018) showed that he was unfamiliar with perceptual verbs plus bare infinitives, causing initial hesitation (e.g., *She **saw** a boy **run** in and **start** talking to the swimming officials*) that decreased with guided practice.

Despite these gains, Suho remained cautious in group discussions, expressing self-doubt: "How dare I do it? (…) I don't know what to do" (1-2). That day, he wrote "Don't shrink" in his learner log—a quiet self-encouragement reflecting the emotional dimension of learning. This underscores the role of affective mediation (Swain, 2006), where internal dialogue and self-regulation foster agency. During an impromptu individual interview (Aug 30, 2018), Suho shared, "They seem to see things I can't," also noting his discomfort as one of the few boys in the group. The teacher emphasized collaborative learning and reassured him that his unease would ease over time. This instance illustrates responsive mediation, addressing emotional and social needs while creating a safe space for risk-taking and growth.

Over time, peer interaction further mediated Suho's development, fostering more confident participation. Excerpt 1 shows he initially hesitated while observing peers' grammar-focused discussion but soon sought clarification. As a *Planner*, he benefited from structured peer scaffolding, transforming earlier hesitation into active pattern recognition. Seo-hee and Yumi acted as more capable peers, pointing to patterns and offering simplified examples ("Like here—*waiting*, *walking*"). Suho's private muttering ("Ah, -*ing*? (...) You mean -*ing*?") manifests languaging (Swain, 2006), allowing him to externalize abstract input, test hypotheses (-*ing*), and eventually extend understanding independently to -*ed* endings—demonstrating microgenetic internalization from intermental to intramental comprehension.

Excerpt 1

(1-3)

All excerpts were translated from Korean into English.

Seo-hee:	Have you done that? (...) *A little while ago*.
Yumi:	These are all *be verbs*. (pointing at Seo-hee's worksheet)
Seo-hee:	Wow, how did you know that?
Suho:	(Tapping Seo-hee on the arm) What is it? What is it? (...) Which part are you talking about?
Seo-hee:	Here, in the front part. (pointing at Suho's worksheet) The subject is like this. (...) If you look closely, you'll see (underlying Suho's worksheet). *Be verbs* (...) like *were* or *is*. Just find something in common. (...) *Verbs*, *verbs*. *Be verbs* and then *participial constructions*.
Yumi:	Like here—*waiting*, *walking*.
Seo-hee:	Yeah.
Yumi:	*Questioned*, *condemned*.
Suho:	Ah, -*ing*? (...) You mean -*ing*? (muttering -*ing* and circling -*ed* endings by himself)

During Phase 2, Suho struggled with corpus tools. He continued to rely heavily on teacher mediation, which was provided as needed. In addition, he increasingly interacted with his peer, Youngmo, for technical support, reflecting ZPD-mediated peer scaffolding.[5] This collaborative support helped Suho remain cognitively engaged, often using private speech ("Huh? What's this?", "Wow, interesting," "Is this sentence in group A?") (2-1) to externalize thoughts to process input. Although his direct manipulation of tools was limited (2.7%), the high proportion of teacher mediation (48.6%), combined with peer support from Youngmo (17.6%) and frequent private speech (31.1%), underscores his preference for guided, socially mediated learning over independent exploration (see Appendix C).

In Phase 3, Suho focused on more challenging grammar and increasingly initiated collaborative inquiry, demonstrating clear exercise of agency. Excerpt 2 illustrates how peer interaction mediated his shift from guided problem-solving to more autonomous, analytic engagement. Private speech evolved into collaborative dialogue with Youngmo, negotiating distinctions such as *where* versus *when*, while Suho

handled COCA searches and Youmgmo managed BNC queries—complementary roles supporting Suho's *Planner*-type agency. Teacher scaffolding occurred only as necessary ("Oh, just like *um* or *uh*—colloquial filler"), reflecting movement within the ZPD from other- to self-regulation (Vygotsky, 1978). Suho's initiative ("Shall we find it ourselves?") exemplified agency exercised, with collaborative languaging becoming a self-sustaining tool. Expressions of enjoyment ("This is fun") indicate growing cognitive and affective engagement.

Excerpt 2

(3-1)

Suho:	Wait, *century where*? *In South Korea in the 21st century, where individuals' freedom and rights are maximized and firmly protected* (...) So, is *century where* correct?
Youngmo:	Let's see, (…) for the country, maybe.
Suho:	It sounds weird. I think it should be *when*.
Youngmo:	For *South Korea*, *where* works, but for *21st century*, *when* is better.
Suho:	Well, (…) Shall we find it ourselves?
Youngmo:	Let's do it! You search BNC, I'll do COCA.
	[a few minutes later]
Suho:	I found five.
Teacher:	You got five in BNC? Read the line?
Suho:	Hmm (…) *Political peace and the model er... for his policy* (…) Ah, what does *er* mean?
Teacher:	Oh, just like *um* or *uh*—colloquial filler.
	[a few minutes later]
Youngmo:	I found 55.
Suho:	That's a lot more. (…) How many times does *century when* appear? Let's search!
Youngmo:	Sure, why not?
Suho:	This is fun.
Youngmo:	Yeah, I agree.
	[a few minutes later]
Suho:	I found 149.
Teacher:	Compared to five, what do you think?
Suho:	Huge difference.
Teacher:	It seems in England most people use *when*.
Youngmo:	I found it, too—in the U.S., *where* came up 55 times, *when* 460; in England, 5 and 149.
Teacher:	Wow, you took all those notes? Amazing. (…) In England *when* dominates, but in the U.S. *where* also appears quite a bit.
Suho:	Exactly—*when* really stands out in England.

In Phase 4, Suho initially felt anxious about making all decisions independently, a pattern familiar from earlier phases ("What if I go in the wrong direction?") (1-3), and he hesitated, ultimately skipping the first

session (text message, Jan 13, 2019). The teacher responded with both emotional and cognitive support, sending an encouraging email and sharing Youngmo's sample independent work (Jan 14–16, 2019), which reignited his motivation (text message, Jan 16, 2019). Reflecting later, Suho admitted, "At first, I had no idea what to do, but after watching Youngmo, it didn't seem so hard" (interview, Mar 12, 2019). These events show how emotional and cognitive support—through teacher encouragement and peer modeling—helped Suho progress within his ZPD, consistent with SCT (Vygotsky, 1978).

Once he resumed, Suho completed four independent learning sessions (Sessions 2–5). Drawing on his *Planner* orientation, he first checked grammar rules in his workbook and then used corpora to confirm or refine them. He postponed using the graded reader corpus until finishing all borrowed books and relied primarily on COCA and BNC due to his slow reading pace. This careful sequencing and growing confidence shaped his approach to demanding corpus tasks.

To process challenging concordance lines, Suho often skipped or shortened difficult sentences, copying segments into notes and isolating them from context, which sometimes limited comprehension. For instance, he reduced *Downstairs in casualty Diriham Ali, 21, patient number 35 in the log book, is being treated for four gunshot wounds* to *Patient number 35 in the log book is being treated for four gunshot wounds* (4-2), overlooking surrounding words like *victims of war* or *arms*. In another case, he removed connectors, changing *Although if there are two people at odds you may lose a bit because they are not going to get the idea together* to simply *There are two people at odds you may lose a bit* (4-3), distorting meaning even with a translator. Suho commented, "It's confusing to figure out what the sentence really means without full context," reflecting his habit of isolating sentences (4-3).

Targeted teacher scaffolding via email (Feb 3, 2019) on meaningfully segmenting concordance lines helped Suho adjust his approach, encouraging him to infer meaning from context rather than relying solely on dictionaries. This boosted his confidence in reading full sentences, illustrating how mediation supported his exercise of agency and emerging autonomy.

By the fifth session, Suho incorporated the graded reader corpus, drawing on familiar texts like *Dracula* and *Sky Jack* to interpret concordance lines. Visualizing scenes helped him contextualize grammar patterns and move from basic segmentation to flexible inferencing, reflecting active cognitive engagement. His exclamation—"Oh, yeah, this is the very sentence! Like this!" (4-5)—demonstrated emotional investment and motivation. By selecting and applying graded reader texts, Suho exercised agency, supporting the development of autonomy. Frequent private speech highlighted his metacognitive awareness, guiding strategic decisions and attention. The final session extended by about 15 minutes, indicating sustained focus and deeper engagement.

Excerpt 3 shows that Suho planned to add the graded reader corpus only after finishing his borrowed books. He noted that prior familiarity helped him identify grammar points more efficiently, recalling, "Yeah, when I looked at the sentence, I could kind of remember which part it was from." This vivid recall, together with his metacognitive awareness, illustrates how familiar, learner-controlled contexts can enhance his confidence and focus during corpus work.

Excerpt 3

(Interview, March 12, 2019)

Teacher: You had been thinking about adding the graded reader corpus, but you only included it in the last period.

Suho: Ah, yeah.

Teacher: I wish you had added the graded reader corpus a bit earlier.

Suho: I planned to add the novel corpus after finishing my reading, but I only managed to finish one of the four books I borrowed.

Teacher: Oh, I see. It took you a while to finish, didn't it? (...) How did it feel when you added the concordance lines from the literature?

Suho: Well, since I had already read it once, (...)

Teacher: So, the context was fresh in your mind, right?

Suho: Yeah, when I looked at the sentence, I could kind of remember which part it was from.

Teacher: Exactly, and that made it much easier for you to identify the grammar point, right?

Suho: I agree, that is exactly it.

In summary, Suho initially struggled with dense corpus data, group discussions, and the software. Through teacher scaffolding, peer modeling, and his own efforts, he gradually gained confidence and engagement. By Phase 3, he participated more actively, identifying discussion points and collaborating with peers, reflecting the development of both strategic and social skills. Challenges in independent decision-making still caused hesitation, but peer examples and teacher guidance helped him refine sentence-segmentation strategies and apply inferencing effectively. Familiar contexts from the graded reader corpus further supported self-regulation, enjoyment, and autonomy through exercised agency. Overall, these developments show how learner-controlled contexts, combined with social and instructional scaffolding aligned with ZPD principles, facilitated the gradual emergence of independent corpus literacy.

Jimin: "Time went so fast, like solving interesting riddles together."

Jimin exemplifies a *Seeker*-oriented learner, marked by exploratory engagement and self-directed inquiry (Schmidt, 2023). In Phase 1, although she hesitated to speak ("Can I do it later? I'm embarrassed") (1-1), she engaged deeply with the materials, showing early cognitive involvement and emerging agency within her ZPD: 'I didn't know the grammatical term at first, but the example sentences made the meaning of perception verbs clear' (learner log, Aug 23, 2018).

In Phase 2, Jimin's learning shifted toward tool-mediated exploration. Her technology mediation (15.1%)—much higher than Suho's (2.7%)—reflected active hypothesis testing and exploration of alternative lexical patterns. Teacher mediation (43.4%) primarily provided confirmatory checks rather than foundational scaffolding, while peer mediation (28.3%) involved increasingly active discussion of corpus findings with peers, especially after engaging with the literature-based corpus (see Appendix C). Together, these patterns indicate a growing preference for tool-supported, self-directed inquiry, consistent with her *Seeker* orientation.

A key turning point occurred when Jimin explored graded reader corpora, such as *Wuthering Heights* and *Jane Eyre*, via LexTutor. Familiar stories sparked curiosity and confidence, increasing private speech, tool use, and peer collaboration. This episode illustrates the interplay of tool-mediated action and social mediation, as literary context and peer prompts co-constructed understanding. Moving from hesitant browsing to enthusiastic engagement, her proactive *Seeker* profile became clearer, reflecting learner agency that fostered autonomous exploration. Excerpt 4 presents a representative episode.

Excerpt 4

(2-2)

Yumi: (tapping Jimin) Hey, look, here's something you read, *Witches of Pendle*.

Jimin: Really? Oh, I also found *Dorian Gray* and *Gladiator*. (increasing the size of each book)

Yumi: Click on the word, separately.

Jimin: (already clicking) Wow, this is cool. (scrolling down and moving on to examine other novels) *Love Story*. (…)

Teacher: Let's take a look at the site. Is it okay?

Jimin: I love it. (clicking on other sources) *Jane Eyre*.

In Phase 3, Jimin integrated peer collaboration with personal interests, shifting from grammar-focused study to vocabulary work inspired by a favorite movie (interview, Oct 30, 2018). When Jungwon's laptop malfunctioned, she and Jimin worked together on her device, while Yumi continued nearby. This flexible collaboration promoted spontaneous interaction among the three, enhancing engagement and enjoyment, as reflected in Jimin's comments: "It's really fun doing it like this with the three of us" (3-1).

Meanwhile, teacher intervention in dictionary use served as pivotal pedagogical mediation, reorienting Jimin's approach to DDL (Excerpt 5). By framing the dictionary as a tool for confirmation rather than a primary source, she scaffolded a hypothesis-driven inquiry that guided Jimin from simple information retrieval to active meaning-making. Jimin's smiling response ("I'll try that") reflected her exercise of agency and expanded cognitive engagement. Peer interaction further supported meaning negotiation and collaborative knowledge construction, reinforcing socially mediated internalization.

Excerpt 5

(3-1)

Jimin: Let's find one on *Naver*[6], then check the corpus.

Jungwon: Oh yeah, that's *enlightening*.

Jimin: I guess it is restrictive?

>> [D] *enlightening*: 계몽적인 [7] (tapping Jungwon)

Jungwon: 계몽적인, like awakening, right? (…)

Teacher: (approaching) What are you two working with?

Jungwon: We looked up *enlightening*.

Teacher: You looked up in the dictionary? (…) Haven't you checked the corpus yet?

Jungwon: We're about to check COCA now to see how many times the word is used.

Teacher: Frequency information is important. So far, we've used the dictionary first.

>> [C] *enlightening*

(…) But, as you can see, words or phrases before or after the target word, or the context, tone, or writing style, can provide clues for understanding the word.

Jimin: Ah, yes, that makes sense.

Teacher: It's not too late to use the dictionary after fully examining the corpus data, as much as possible. It's like testing a hypothesis, inferring from corpus data. Then, we can check if the guess was correct.

Jimin: (smiling) I'll try that.

After the teacher's mediation, Jimin and her peers analyzed *top notch* by engaging directly with the corpus to infer meaning ("*Cloud*, *garden*, *team* (…) there's always some noun following *top notch*"). Jimin extended the analysis with contextualized examples ("*Top notch legal teams*, *top notch engineers*, *top notch facilities and staff*") (3-1). This episode illustrates rapid internalization of the scaffolded practice and a shift in the learners' collective agency—from passive dictionary users to active, collaborative constructors of meaning.

As her confidence grew, Jimin's *Seeker* orientation became increasingly apparent, prompting her to lead peer-led corpus inquiry. In Excerpt 6, she guided the group in exploring *popping pills*, with Jungwon examining prepositional patterns and Yumi adding interpretive insights. The teacher offered subtle scaffolding to encourage a more targeted search, but Jimin retained control over meaning-making, inferring nuances and negotiating interpretations with peers. Drawing on familiar movie scenes, the learners connected the phrase to narrative context, deepening understanding. Their playful, socially engaging exchange—including laughter and role play—sustained motivation. Jimin's initiative demonstrates her exercise of agency and growing autonomy with corpus tools, while her peers' engagement indicates a shift in collective agency as the group moved from guided participation to collaborative meaning-making.

Excerpt 6

(3-1)

Jimin:	Why don't we search for *popping*—was it *popping pills*?
Jungwon:	Let's check BNC. (…) Lots of prepositions here. *Popping at*, *popping over*.
Teacher:	What did you type in the search bar?
Jimin:	Just *popping*.
Teacher:	Why don't you try typing *popping pills* all?
Jungwon:	>> [C] *popping pills*: 26
	It shows 26 cases.
Jimin:	Looks like it means taking a pill.
Yumi:	Like *popping candy*? Putting the pill in your mouth, like that? (pretending)
Jimin:	Yeah, sounds like casual everyday expression.
Jungwon:	(…) *She was drinking very heavily and popping pills at the same time*.
Yumi:	Sounds like drugs.
Teacher:	(looking at the screen) See? With *drinking*, *drink* (…) it seems *drug*-related. (…) Remember the movie context?
Jimin:	So, (…)
Jungwon:	Because they were all patients, I think, it came from taking medicine.
Yumi:	Right, there were several scenes like that.
Jimin:	Not illegal drugs—more like habitual medicine use?
Jungwon:	Like a grandfather taking all his meds at once—*Pak*! (pretending)
Yumi:	You look like a real grandfather. (laughing)
Jimin:	(smiling) Shall we look for something else?

As the session progressed, the group negotiated meaning collaboratively. In Excerpt 7, Jungwon first interpreted *Godspeed* as slang but revised her judgement after noticing most corpus examples came from news sources. Jimin observed its formal tone and interpreted it in context, prompting deeper peer discussion. She linked it to "good luck," while Yumi added cultural insights. A brief dictionary check confirmed their interpretation, yet Jimin's integration of corpus data, peer input, and contextual reasoning solidified understanding. Their playful interaction—including winks and laughter—sustained motivation and reflected collective agency as the group moved toward co-constructed, autonomous meaning-making.

Excerpt 7

(3-1)

Jungwon:	(patting Jimin's shoulder) It sounds like slang. (…) Wait, it can't be right. Slang doesn't appear in the news. (scrolling down, noticing many sources are news articles)
Jimin:	It might not be slang.
Jungwon:	Ah, yeah, you're right. (examining more concordance lines)
Jimin:	*All the best*! *Godspeed*! *I made sure not to say*, '*Good luck*.' *Godspeed on your journey*. *I wish him Godspeed*.
Yumi:	*Good luck and Godspeed*. It sounds like "good luck." Yeah, it's like "good luck."
Jungwon:	Same here. So, *good luck and Godspeed*. I guess it's also just "good luck."
Jimin:	Isn't it an interjection?
Yumi:	You know, Americans wink when they wish someone good luck.
Jimin:	Oh, that's right.
Yumi:	That's exactly how it feels.
Jungwon:	[D] *godspeed*: 성공, 행운의 축복 [8] You can either space out the words or leave them together. It says here it's a blessing for success.
Yumi:	Should we do the same next time? (giving a wink)
Jungwon:	*Godspeed to you*! (laughing)
Yumi:	Like this. (gesturing) Jungwon, for your interview tomorrow, I'll give you *Godspeed*!
Jimin:	Me too! (laughing)

In the following session, Jimin reflected, "Time went so fast, like solving interesting riddles together" (3-2), capturing the enjoyment and ownership fostered through shared discovery. Building on this engagement, in Phase 4 she persistently explored word meanings, selecting multiple corpora based on her own rationale, testing hypotheses, and applying linguistic knowledge independently, as shown in the following examples:

"BNC has fewer examples than COCA, so I'll check *derive* in COCA again. (…) The dictionary says it's British? Then I'll look it up in BNC for better examples, since BNC is from the UK."

"*Throughout* seems like a conjunction, so I'll check the graded reader corpus with AntConc. (…) It seems to indicate a period or general state at a certain time, but I'm still not sure, so I'll verify it through COCA."

"*Cultural* means culture, but *horticultural* is new to me. I think it may be British, so I'll look it up in BNC." (4-2)

This approach illustrates Jimin's growing autonomy, agentive decision-making, and reflective engagement, showing how self-directed corpus selection, scaffolded by ZPD principles, supports hypothesis-driven learning. She sometimes explored additional examples to refine her understanding, attending closely to context, collocates, and tone. For instance, when examining *abundant*, she reflected:

"*Abundant*? I guess (…) it seems like, it's used differently in each sentence. In some cases, it seems to describe a shape or appearance—perhaps from a novel? And then in other places, it feels kind of (…)

scientific? So it's the same word, but it doesn't seem to mean exactly the same thing. It's like it has a lot of different meanings." (4-2)

Although generally accurate, she sometimes expressed uncertainty—"It could be right, or maybe not"—and doubted her inferences when they conflicted with dictionary definitions ("Oh, then my inference was totally off") (4-2). This hesitation reflects her *Seeker* orientation, indicating a willingness to keep exploring despite uncertainty. Timely scaffolding—an encouraging email—helped her regain confidence, and her cheerful reply ('I'd like to try something new next time 😊,' text message, Jan 23, 2019) signals the confident exercise of agency and readiness to take risks in self-directed learning.

Subsequent sessions involved TED Talks, where she proactively engaged with content she had previously avoided (text message, Jan 29, 2019). She paused videos to look up unfamiliar words and even watched another lecture when one ended early. Demonstrating strategic self-regulation, she maintained focus despite the loud Lunar New Year holiday at her grandmother's house: "It was so noisy outside that I just couldn't concentrate. So I figured it'd be much better to just listen to a TED Talk on marine animals—one of my interests—to maintain engagement" (text message, Feb 8, 2019).

Jimin developed a reciprocal relationship between corpus data and TED content, using lecture insights to guide corpus exploration and deepen vocabulary. She inferred from concordance lines that *confession* meant "announcement" or "revealing," linking it to the TED context: "The thing that person said earlier sounds like that too. (…) Like, *I have a confession to make*. That's my rough guess." Her curiosity sustained engagement: "What kind of *confession* would the speaker make in front of a crowd? The topic doesn't seem cheerful. I'll keep watching" (4-3).

Similarly, she identified consistent patterns across sources. For example, in the case of *go viral*, she described it as spreading "by word of mouth or social media, like a virus" (4-3). For *cascade*, a vivid TED Talk image of a chain reaction harming shrimp and coral reefs helped her grasp the idea of "continuity." She then confirmed this through corpus examples, such as *energy generates a cascade of new particles* and *a slow cascade of books*, concluding that *cascade* means "a continuous flow", effectively integrating lecture imagery, corpus data, and dictionary evidence (4-4).

Through this strategic, flexible engagement, Jimin's inferencing confidence grew. She reflected on prior guesses and refined meanings: "*Detachment* is not exactly 'separation,' but close—like 'keeping a distance,' as I inferred, confirming my earlier guess," and later, "It's like what I thought before, but the dictionary is more detailed. Another word came to mind, though I'm unsure if it fits" (4-5). These episodes illustrate her agentive, self-directed approach, combining multi-modal input, corpus analysis, and reflective practice to construct nuanced, context-sensitive vocabulary.

In summary, Jimin's development illustrates how personal interest and meaningful engagement create conditions for exercising learner agency. Initially hesitant due to prior grammar-focused instruction, she became increasingly active through collaborative vocabulary tasks aligned with her interests. In Phase 4, she flexibly integrated corpus data and TED Talks to deepen understanding, demonstrating strategic autonomy, reflective practice, and sustained engagement. Having examined both learners, their trajectories show how individuals uniquely navigate the four DDL phases, with patterns and contrasts explored further in the Discussion.

Discussion

This section presents the study's contributions, showing how SCT frames the exercise of learner agency toward autonomy in mediated DDL.

Theoretical Implications: Extending SCT in Digital DDL Contexts

This study extends SCT by demonstrating the reciprocal nature of mediation in digitally mediated DDL environments, where human (teacher, peers), tool-based (corpora, dictionaries, TED Talks), and internal

(private speech, prior knowledge) mediators interact dynamically. Whereas classic SCT often emphasizes unidirectional flows—from expert to novice or from tool to internalized use—this study highlights a dialogic process, clearly illustrated in Jimin's case. The TED Talk initially mediated her vocabulary learning, yet she exercised agency to reinterpret and repurpose it, generating targeted corpus queries. This feedback loop reshaped the tool itself, exemplifying a reciprocal dialogue between learner and technology and aligning with Wertsch's (1998) concept of agency-as-mediated action in a multimodal context. In DDL, such co-construction fosters autonomous decision-making and sustained engagement, showing that autonomy emerges as both internally regulated and socially supported through private speech, tool use, and strategic choices.

The findings also underscore the interplay of agency, identity, and affect. Consistent with Pavlenko and Lantolf (2000), who argue that language learning engages the whole person—where identity, desire, and emotion function as mediators—these results show that agency cannot be fully understood apart from affective and identity dimensions. Suho's trajectory illustrates this: his *Planner* identity triggered high anxiety about mistakes ("What if I go in the wrong direction?") (1-3), constraining his agency and requiring emotional support from the teacher. His Phase 4 cognitive breakthrough was sparked not only by reasoning but by affective resonance with a familiar literary text (*Dracula*). For adolescents, such emotional engagement powerfully mediates cognitive development, confirming that DDL is not purely cognitive but also emotionally engaging. Overall, the study reconceptualizes SCT in digital DDL contexts as a reciprocal, affectively and identity-laden process, capturing the lived dynamics of adolescent learners navigating multimodal learning environments.

Pedagogical Implications: The Centrality of Teacher TPACK in Mediated DDL

Beyond theoretical contributions, the study offers practical insights for secondary EFL classrooms, where corpus tools are often underused. While DDL is learner-centered, technology alone does not guarantee independent learning; meaningful mediation is essential. SCT-aligned research (Flowerdew, 2008; Huang, 2011; Park, 2012) emphasizes teacher and peer scaffolding, particularly for younger learners who might be overwhelmed by raw corpus data. Central to this process is the teacher's TPACK, flexibly responding to learners' evolving needs (Mishra & Koehler, 2008).

Teacher interventions—emotionally supporting Suho, guiding Jimin's dictionary use, or modeling peer behavior—illustrated responsive, embodied TPACK. Strategic choices, such as beginning with paper-based concordances, scaffolding dense input, and connecting tasks to learner interests, made potentially overwhelming tools accessible and meaningful. Well-designed tasks and peer discussion prevent cognitive overload while sustaining engagement (Flowerdew, 2008). Teachers' linguistic expertise also helped learners notice patterns, revise hypotheses, and build vocabulary, especially in high-stakes contexts. Without sufficient TPACK, DDL risks becoming an added cognitive burden; with it, it promotes learner agency.

Ultimately, effective DDL relies on orchestration, not tools alone. Teachers who integrate corpus tools with thoughtful task design, emotional attunement, and linguistic insight can better scaffold learner independence. TPACK should be understood as a dynamic, practice-based competence that evolves alongside learners' needs and classroom ecology. Professional development must therefore go beyond tool training to address when, how, and why to use corpus tools and how to sustain learners' motivation and engagement.

Methodological Implications: Longitudinal, Multimodal, and Ecological Approaches

This study provides two key methodological insights for DDL research. First, it underscores the value of longitudinal, multimodal approaches. Short-term interventions or pre/post-test designs cannot capture the complex, non-linear exercise of learner agency and the development of autonomy. In-depth case studies tracing learning over time and across multiple modes—combining microgenetic observation with intra- and interpersonal interactions—reveal how learners progressively internalize strategies, self-regulate, and engage in socially mediated learning. Triangulated data sources, including screen recordings, think-aloud

protocols, interviews, and observations, are crucial for identifying subtle yet critical moments of agentive action.

Second, an ecological approach to data collection is essential. Mediational events often occur beyond the corpus interface, as seen in Suho's peer scaffolding in Phase 1 or Jimin's collaborative dialogues in Phase 3. Studies focusing only on log files or screen activity risk overlooking these dynamics. Capturing the full learning ecology links SCT theory with classroom practice and provides a robust model for future secondary-level DDL research.

Conclusion

The exercise of learner agency and the development of autonomy in DDL are complex, socially- and affectively-mediated phenomena that vary according to learner identity, rather than reflecting uniform technical skill acquisition. In this study, learner agency refers to the initiative and control exercised through mediated actions, whereas autonomy denotes the capacity that emerges as learners internalize these mediated actions into self-regulated activity. Adolescent learners' interactions with corpus-based tools were influenced by cognitive strategies, affective resonance, identity orientation, and strategic decision-making.

This study contributes theoretically, pedagogically, and methodologically. Theoretically, it extends SCT by conceptualizing mediation as a reciprocal process in digitally mediated contexts, highlighting the intertwined roles of agency, identity, and affect. By building on Pavlenko and Lantolf's (2000) affective and identity-mediated view of SCT, this study bridges theoretical propositions and classroom-based evidence, showing how these dimensions dynamically shape learner agency trajectories. Pedagogically, teacher TPACK is crucial for coordinating tools, tasks, peer discussion, and affective support, transforming complex resources into opportunities for learner autonomy, in line with SCT evidence on guided interaction and microgenetic support. Methodologically, the study demonstrates the value of longitudinal, ecological, and multimodal approaches for capturing the nuanced, moment-by-moment dynamics of learner development.

Although the small, voluntary sample limits generalizability, the rich, triangulated data offer detailed insights into learner development. Building on the teacher's pivotal role in mediating DDL, future research could explore how emerging tools—particularly generative AI—may act as mediators, presenting new challenges and opportunities for SCT-informed pedagogy, TPACK development, and the promotion of learner autonomy in digital environments.

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

Boulton, A. (2009). Data-driven learning: Reasonable fears and rational reassurance. *Indian Journal of Applied Linguistics*, *35*(1), 81–106.

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology*, *25*(3), 66–89. https://doi.org/10.10125/73450

Chambers, A. (2019). Towards the corpus revolution? Bridging the research–practice gap. *Language Teaching*, *52*(4), 460–475. https://doi.org/10.1017/S0261444819000176

Copland, F., & Creese, A. (2015). *Linguistic ethnography: Collecting, analysing and presenting data*. SAGE Publications.

Crosthwaite, P., & Baisa, V. (2024). A user-friendly corpus tool for disciplinary data-driven learning: Introducing CorpusMate. *International Journal of Corpus Linguistics*, *29*(4), 595–610. https://doi.org/10.1075/ijcl.23056.cro

Crosthwaite, P., & Boulton, A. (2024). DDL is dead? Long live DDL! Expanding the boundaries of data-driven learning. In H. Tyne, M. Bilger, L. Buscail, M. Leray, N. Curry, & C. Pérez-Sabater (Eds.), *Discovering language: Learning and affordance*. Peter Lang.

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, *11*(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

Douglas Fir Group. (2016). A transdisciplinary framework for SLA in a multilingual world. *The Modern Language Journal*, *100*(S1), 19–47. https://doi.org/10.1111/modl.12301

Flowerdew, L. (2008). Corpus linguistics for academic literacies mediated through discussion activities. In D. Belcher, & A. Hirvela (Eds.), *The oral-literate connection: Perspectives on L2 speaking, writing, and other media interactions* (pp. 268–287). University of Michigan Press.

Huang, L. (2011). Language learners as language researchers: The acquisition of English grammar through a corpus-aided discovery learning approach mediated by intra- and interpersonal dialogues. *Studies in Second Language Acquisition*, *33*(1), 89–122. https://doi.org/10.1017/S0272263110000740

Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. *ELR Journal*, *4*, 1–16.

Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford University Press.

Merriam, S. B. (1998). *Qualitative research and case study applications in education*. Jossey-Bass Publishers.

Mishra, P., & Koehler, M. J. (2008). Introducing technological pedagogical content knowledge. In A. C. Bishop, C. E. Dwyer, & J. L. Forman (Eds.), *Handbook of technological pedagogical content knowledge (TPACK) for educators* (pp. 3–29). Routledge.

Moon, S., & Oh, S.-Y. (2017). Unlearning overgenerated *be* through data-driven learning in the secondary EFL classroom. *ReCALL*, *30*(1), 48–67. https://doi.org/10.1017/S0958344017000246

O'Keeffe, A. (2021). Data-driven learning, theories of learning and second language acquisition: In search of intersections. In P. Pérez-Paredes & G. Mark (Eds.), *Beyond concordance lines: Corpora in language education* (pp. 35–55). John Benjamins.

Park, K. (2012). Learner–corpus interaction: A locus of microgenesis in corpus-assisted L2 writing. *Applied Linguistics*, *33*(4), 361–385. https://doi.org/10.1093/applin/ams012

Pavlenko, A., & Lantolf, J. P. (2000). Second language learning as participation and the (re)construction of selves. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 155–177). Oxford University Press.

Pérez-Paredes, P. (2019). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011–2015. *Computer Assisted Language Learning*, *35*(1–2), 36–61. https://doi.org/10.1080/09588221.2019.1667832

Philp, J., & Duchesne, S. (2016). Exploring engagement in tasks in the language classroom. *Annual Review of Applied Linguistics*, *36*, 50–72. https://doi.org/10.1017/S0267190515000094

Schmidt, N. (2023). Unpacking second language writing teacher knowledge through corpus-based pedagogy training. *ReCALL*, *35*(1), 40–57. https://doi.org/10.1017/S0958344022000106

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). SAGE Publications.

Swain, M. (2006). Languaging, agency and collaboration in advanced second language learning. In H. Byrnes (Ed.), *Advanced language learning: The contributions of Halliday and Vygotsky* (pp. 95–108). Continuum.

Toohey, K., & Norton, B. (2003). Learner autonomy as agency in sociocultural settings. In D. Palfreyman, & R. C. Smith (Eds.), *Learner autonomy across cultures: Language education perspectives* (pp. 58–72). Palgrave Macmillan.

van Lier, L. (2008). Agency in the classroom. In J. P. Lantolf, & M. E. Poehner (Eds.), *Socio-cultural theory and the teaching of second languages* (pp. 163–186). Equinox.

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

Wertsch, J. V. (1998). *Mind as action*. Oxford University Press.

Endnotes

1. The Career Aptitude Test, administered on March 19, 2018, is a nationally used tool to help high school students explore potential career paths.

2. This case study is part of a larger project involving seven learners (all names are pseudonyms). The current analysis focuses on Suho and Jimin as representative cases.

3. The graded reader corpus was sourced from the open-access Lex Tutor website (lextutor.ca), which provided excerpts from leveled e-books including Cambridge English Readers (Levels 1-5), Oxford Bookworms (Levels 1-6), and Penguin Readers (both fiction and non-fiction, Levels 1-5).

4. Observation and think-aloud data are labeled by Phase–Session (e.g., 1–1); all other sources are cited by date.

5. In Phase 1, Youngmo showed little engagement, but during Phase 2, he became more active—partly due to his interest in the computer programming—and began offering voluntary assistance to peers, including Suho.

6. *Naver* is a widely used South Korean online platform that provides search, dictionary, and other language-related tools.

7. Translated as “enlightening, awakening.”

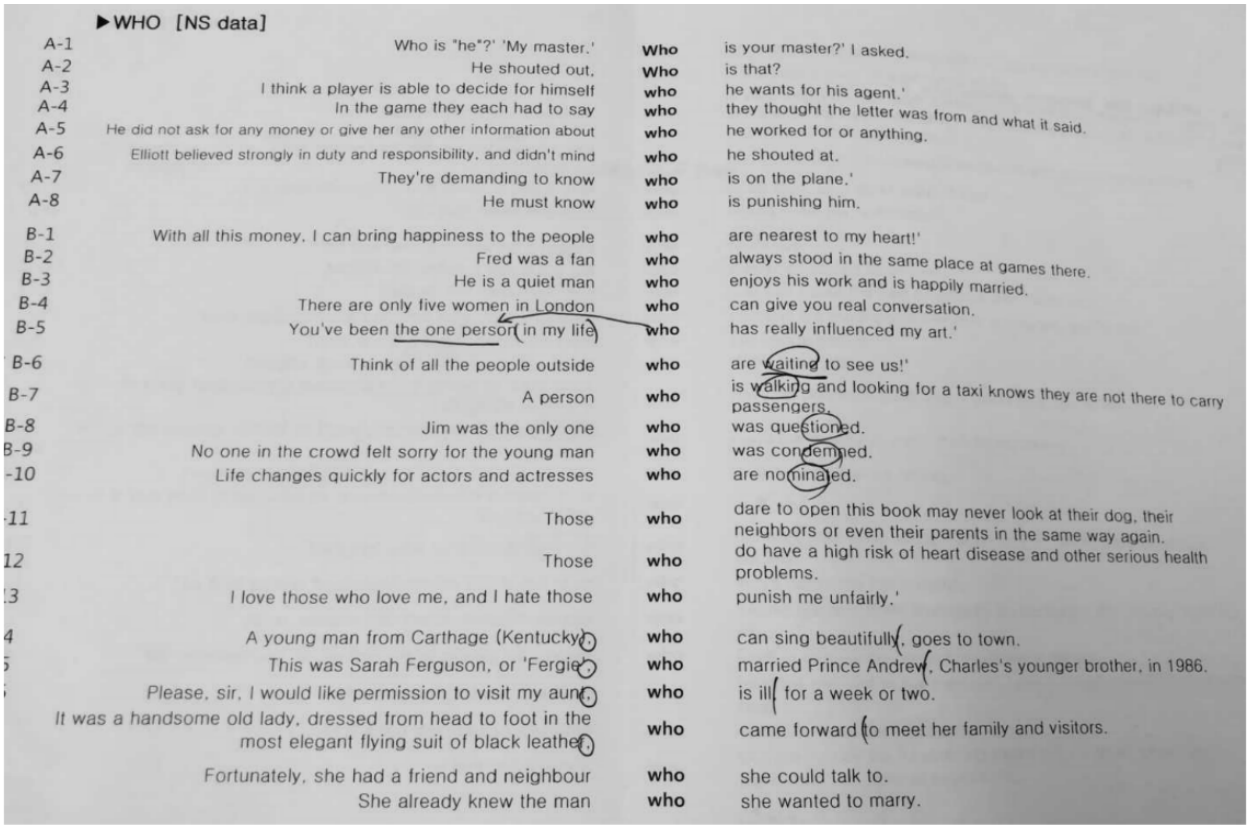
8. Translated as “a blessing for success, good fortune.”

Appendix A. Detailed Chronological Schedule of DDL Activities (Phases 1–4)

This appendix presents the complete schedule of tasks across Phases 1–4, including session dates, duration, learning topics, and participant involvement. Each phase included multiple sessions, with activities ranging from paper-based exercises and hands-on corpus analysis to customized and self-directed tasks. Sessions incorporated both formal instruction and interviews, and Phases 3 and 4 included individualized routes for the two focal participants to accommodate their differing learning focuses.

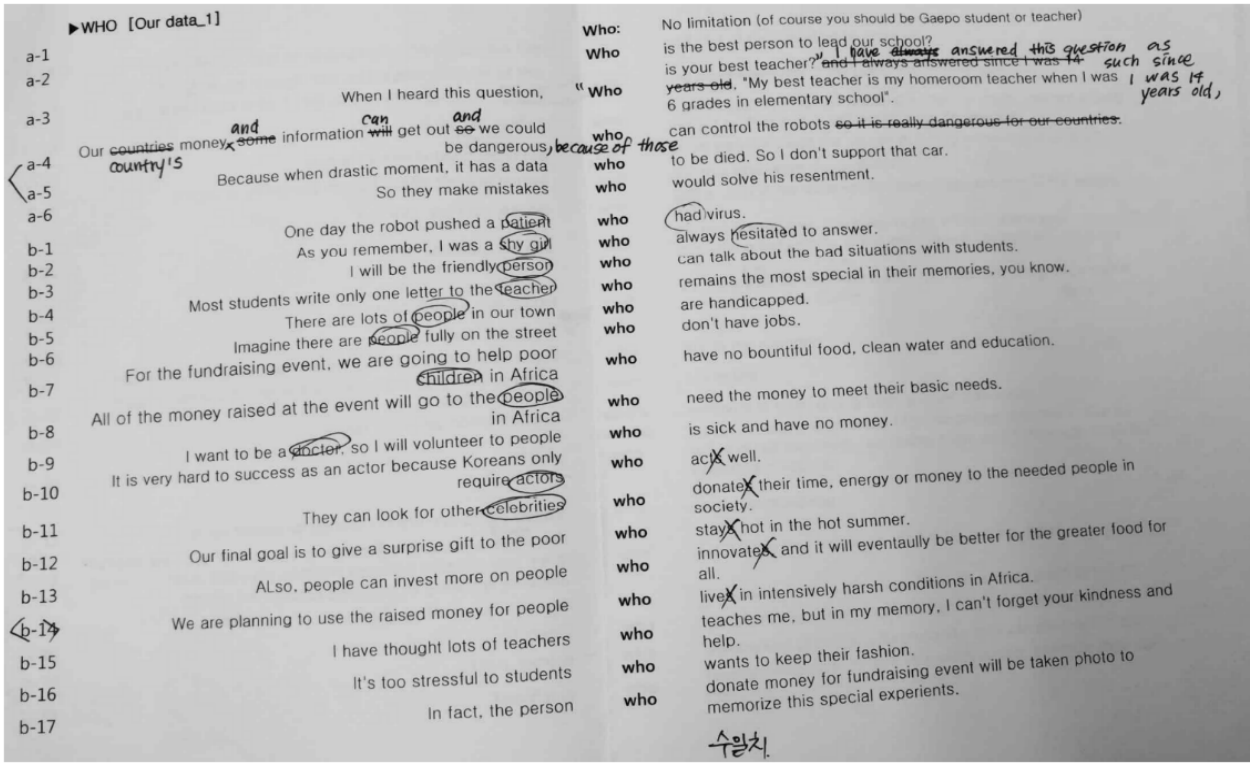
Phase 1 **Paper-based DDL** **(Aug–Sep 2018)** **5 sessions** **(50–100 min each)**	**Phase 2** **Hands-on DDL** **(Oct 2018)** **3 sessions** **(100–120 min each)**	**Phase 3** **Customized DDL** **(Nov 2018–Jan 2019)** **3 sessions** **(60–100 min each)**		**Phase 4** **Self-directed DDL** **(Jan–Feb 2019)** **4–5 sessions** **(31–62 min each)**	
• Orientation (50 min) • Perceptive verbs (60 min) • Causative verbs (90 min) • Impromptu individual interview (15 min) • Relative pronouns: who, whom (100 min) • Relative pronouns: whose, which (90 min)	• What (120 min) • That (100 min) • Impromptu individual interview (15 min) • Light verbs (102 min) • Intermediate group interview (30 min)	Suho • Relative pronouns and adverbs (80 min) • Infinitives and gerunds (70 min) • Overview of grammatical items with pedagogic corpus (60 min) • Intermediate individual interview (20 min) • Presentation in regular classroom (20 min) • Whole group reflection (60 min)	Jimin • Movie vocabulary 1 (100 min) • Movie vocabulary 2 (90 min)	Suho • Giving up • Passive voice (37 min) • Auxiliary verbs (36 min) • Infinitives and gerunds (31 min) • Relative pronouns and adverbs (49 min)	Jimin • Reading texts (62 min) • Reading texts (61 min) • TED Talks (61 min) • Reading texts, TED Talks (61 min) • Reading texts (62 min)

Appendix B. Sample Paper-based DDL Materials



▶WHO [NS data]

A-1	Who is "he"?' 'My master.'	Who	is your master?' I asked.
A-2	He shouted out,	Who	is that?
A-3	I think a player is able to decide for himself	who	he wants for his agent.'
A-4	In the game they each had to say	who	they thought the letter was from and what it said.
A-5	He did not ask for any money or give her any other information about	who	he worked for or anything.
A-6	Elliott believed strongly in duty and responsibility, and didn't mind	who	he shouted at.
A-7	They're demanding to know	who	is on the plane.'
A-8	He must know	who	is punishing him.
B-1	With all this money, I can bring happiness to the people	who	are nearest to my heart!'
B-2	Fred was a fan	who	always stood in the same place at games there.
B-3	He is a quiet man	who	enjoys his work and is happily married.
B-4	There are only five women in London	who	can give you real conversation.
B-5	You've been the one person in my life	who	has really influenced my art.'
B-6	Think of all the people outside	who	are waiting to see us!'
B-7	A person	who	is walking and looking for a taxi knows they are not there to carry passengers.
B-8	Jim was the only one	who	was questioned.
B-9	No one in the crowd felt sorry for the young man	who	was condemned.
-10	Life changes quickly for actors and actresses	who	are nominated.
-11	Those	who	dare to open this book may never look at their dog, their neighbors or even their parents in the same way again.
12	Those	who	do have a high risk of heart disease and other serious health problems.
13	I love those who love me, and I hate those	who	punish me unfairly.'
4	A young man from Carthage (Kentucky),	who	can sing beautifully, goes to town.
5	This was Sarah Ferguson, or 'Fergie',	who	married Prince Andrew, Charles's younger brother, in 1986.
	Please, sir, I would like permission to visit my aunt,	who	is ill for a week or two.
	It was a handsome old lady, dressed from head to foot in the most elegant flying suit of black leather,	who	came forward to meet her family and visitors.
	Fortunately, she had a friend and neighbour	who	she could talk to.
	She already knew the man	who	she wanted to marry.



▶WHO [Our data_1]

		Who:	No limitation (of course you should be Gaepo student or teacher)
a-1		Who	is the best person to lead our school?
a-2		Who	is your best teacher? ~~and I always answered since I was 14 years old~~. "My best teacher is my homeroom teacher when I was 6 grades in elementary school".
a-3	When I heard this question,	" Who	I have ~~always~~ answered this question as such since I was 14 years old,
a-4	Our ~~countries~~ country's money, ~~and some~~ information ~~will~~ can get out ~~so~~ and we could be dangerous,	who	can control the robots ~~so it is really dangerous for our countries~~. because of those
a-5	Because when drastic moment, it has a data	who	to be died. So I don't support that car.
a-6	So they make mistakes	who	would solve his resentment.
b-1	One day the robot pushed a patient	who	had virus.
b-2	As you remember, I was a shy girl	who	always hesitated to answer.
b-3	I will be the friendly person	who	can talk about the bad situations with students.
b-4	Most students write only one letter to the teacher	who	remains the most special in their memories, you know.
b-5	There are lots of people in our town	who	are handicapped.
b-6	Imagine there are people fully on the street	who	don't have jobs.
b-7	For the fundraising event, we are going to help poor ~~children~~ in Africa	who	have no bountiful food, clean water and education.
b-8	All of the money raised at the event will go to the people in Africa	who	need the money to meet their basic needs.
b-9	I want to be a ~~doctor~~, so I will volunteer to people	who	is sick and have no money.
b-10	It is very hard to success as an actor because Koreans only require actors	who	act well.
b-11	They can look for other celebrities	who	donate their time, energy or money to the needed people in society.
b-12	Our final goal is to give a surprise gift to the poor	who	stay hot in the hot summer.
b-13	ALso, people can invest more on people	who	innovate and it will eventaully be better for the greater food for all.
b-14	We are planning to use the raised money for people	who	live in intensively harsh conditions in Africa.
b-15	I have thought lots of teachers	who	teaches me, but in my memory, I can't forget your kindness and help.
b-16	It's too stressful to students	who	wants to keep their fashion.
b-17	In fact, the person	who	donate money for fundraising event will be taken photo to memorize this special experients.

수일치

Appendix C. Mediation Types by Domain

Suho's Case

Domain (%)		Type	Initiator	Frequency	Episode stage
Self-mediation (31.1)		Private speech on data	Learner	15	103, 104, 108, 109, 110, 111, 201, 204, 205, 206, 207, 301, 303, 304, 306
		Private speech on process	Learner	8	103, 107, 110, 201, 204, 301, 303, 304
Other-mediation	Teacher (48.6)	Answering teacher's questions	Teacher	11	201, 202, 204, 205, 206, 207, 301, 302, 303, 304, 306
		Requesting confirmation from teacher	Learner	11	107, 110, 111, 204, 205, 206, 301. 302, 303, 305, 306
		Asking questions to teacher	Learner	8	103, 110, 111, 205, 206, 304, 305, 306
		Receiving help from teacher	Teacher	5	110, 111, 202, 204, 205
		Requesting help from teacher	Learner	1	102
	Peer (17.6)	Receiving help from peers	Peers	4	106, 204, 302, 303
		Requesting confirmation from peers	Learner	3	102, 108, 111
		Asking questions to peers	Learner	2	111, 303
		Sharing corpus data with peers	Learner	1	110
		Discussing corpus data shared by peers	Peers	1	110
		Offering opinions	Learner	1	110
		Helping other peers (facilitated by teacher)	Teacher	1	204
Technology mediation (2.7)		Direct corpus interaction	Learner	2	109, 306

Note. Numbers in episode strage column indicate session and DDL episode (e.g., 108 = Session 1, the 8th DDL episode).

Jimin's Case

Domain (%)		Type	Initiator	Frequency	Episode stage
Self-mediation (13.2)		Private speech on data	Learner	4	108, 204, 205, 206
		Private speech on process	Learner	3	111, 204, 206
Other-mediation	Teacher (43.4)	Requesting confirmation from teacher	Learner	10	102, 111, 201, 202, 204, 205, 206, 207, 304, 306
		Answering teacher's questions	Teacher	7	204, 205, 206, 207, 301, 302, 303
		Asking questions to teacher	Learner	4	110, 205, 206, 306
		Requesting help from teacher	Learner	2	205, 207
	Peer (28.3)	Receiving help from peers	Peers	6	104, 106, 110, 202, 205, 206
		Asking questions to peers	Learner	3	204, 205, 301
		Sharing corpus data with peers	Learner	3	204, 206, 304
		Requesting confirmation from peers	Learner	1	204
		Requesting help from peers	Learner	1	103
		Helping peers	Learner	1	301
Technology mediation (15.1)		Direct corpus interaction	Learner	8	104, 105, 107, 108, 109, 204, 301, 306

About the Authors

Soyeon Moon obtained her Ph.D. in English Education from Seoul National University and teaches English as a foreign language at a public high school in Korea. Her research interests focus on learner corpus-based methodologies for language learning and teaching.

E-mail: soyeonmoon@snu.ac.kr

ORCiD: https://orcid.org/0009-0007-1494-6795

Sun-Young Oh received her Ph.D. in Applied Linguistics from the University of California, Los Angeles. She is a professor in the Department of English Education at Seoul National University and an affiliated professor at the Learning Sciences Research Institute. Her research focuses on learner corpus research and technology-enhanced language learning, including corpus-based pedagogy and AI-mediated learning and feedback. She is the corresponding author.

E-mail: sunoh@snu.ac.kr

ORCiD: https://orcid.org/0000-0001-6686-5427