2nd ACM SIGCHI CHIGreece-Cyprus Chapters Symposium 2026
November 6–7, 2026 | Athens, Thessaloniki, Nicosia & Online

Pedagogical architectures of student–LLM interaction in L2 learning: A PRISMA systematic review and action research design for Greek

Spartakos Anagnostaras, Iro Voulgari, & Stathis Selimis
Department of Speech & Language Therapy, University of the Peloponnese, Greece

1 THE PROBLEM

- **AI in classrooms:** Deployed without sufficient empirical knowledge, including L2 contexts
- **Higher Ed bias:** Initial research targets Higher Education; findings do not apply directly to schools
- **Surface vs. Deep:** LLMs boost fluency (grammar/vocabulary) but lack deep cognitive impact
- **Retention gap:** L2 retention post-LLM withdrawal is rarely evaluated
  - **Liu et al. (2026):** Learning gains emerged at 1 month: LLM group retained skills vs. control regression

2 IT IS THE ARCHITECTURE, NOT THE LLM *PER SE*

- **Interaction design:** Research often neglects prompts, system rules, and interaction constraints
- **Sung & Kang (2025):** Withholding ready-made answers lowered satisfaction but yielded higher text scores
- **Pedagogical friction:** Desirable difficulty and productive struggle are key to deep learning

3 THEORETICAL FRAMEWORK

- **Constructivism & discovery learning** (Alanazi, 2016): Active knowledge co-construction through dialogue
- **Feedback model** (Hattie & Timperley, 2007): *Task*, *process*, *self-regulation* and *self* levels
- **TPACK framework** (Mishra & Koehler, 2006): Technology, pedagogy, and content integration

4 OBJECTIVE & RESEARCH QUESTIONS

Objective: Examine how student–LLM interaction architecture shapes feedback quality and learning retention

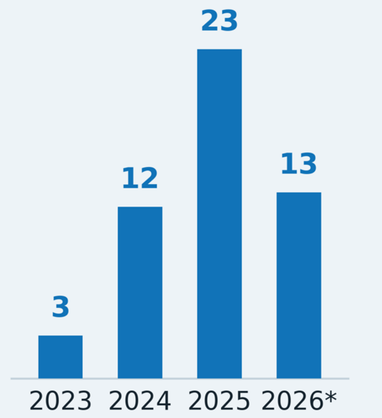
Main question: How do teacher-mediated vs. direct ChatGPT interaction architectures support L2 Greek?

Q1: How does ChatGPT feedback to the teacher align with Hattie & Timperley (2007) feedback levels?
Q2: What is the quality of direct ChatGPT feedback to lower secondary learners?
Q3: How do L2 Greek skills develop and persist post-AI support (retention)?
Q4: What interaction strategies do secondary school learners adopt with ChatGPT?
Q5: How does metalinguistic awareness evolve across L1-L2?

5 AN EMERGING FIELD: PRISMA 2020 REVIEW

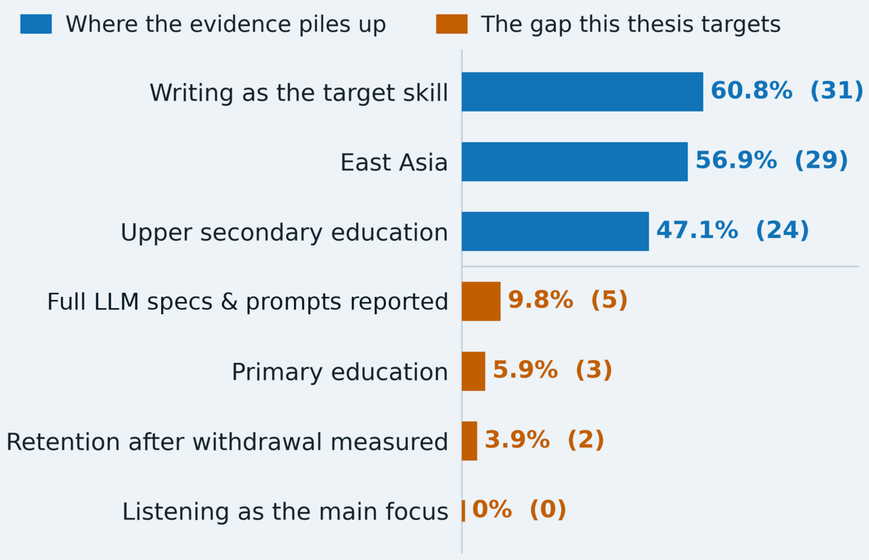
Systematic Review (2020–2026)

***n* = 51 studies**
Scopus *n* = 43
ERIC *n* = 8



No eligible studies before 2023
*2026 data: Q1 only

6 EVIDENCE BREAKDOWN & GAP ANALYSIS



7 BRIDGING THE GAP: THIS STUDY

Gap in the evidence	This study
Primary education underrepresented	Focus on primary (n = 16) & lower secondary (n = 6)
English-language dominance	Greek as L2 (low-resource & limited-exposure context)
Short interventions (3–6 weeks)	3-year prospective design (2026–2028)
Focus on written output	Dual focus: written texts & audio recordings
Emphasis on motivation over linguistic gains	Balanced tracking of language performance & retention
Generic LLM feedback vs. teacher	Scaffolding quality via Hattie & Timperley feedback levels
Omission of LLM specs & prompts	Full reporting of LLM specs, versions & prompts

8 ACTION RESEARCH DESIGN

- **Methodology:** Prospective collective case study and practical action research
- **Setting & Duration:** 'Ellada' school, Moldova (3-year prospective study: 2026–2028)
- **Participants:** Total *N* = 22 Romanian-speaking pupils of Greek origin

STUDY A: Primary cohort (*n* = 16)
Teacher ⇄ ChatGPT → Pupils

STUDY B: Lower secondary cohort (*n* = 6)
Pupils ⇄ ChatGPT

9 RESEARCH DATA & TOOLS

- **Standardised tests:** 'Let's Speak Greek' (Anastasiadi-Symeonidi et al., 2011); DVIQ (Stavrakaki & Tsimpli, 2000)
- **Custom tools:** Mid-term & delayed retention tests; observation forms; feedback rubrics
- **Data sources:** Student texts & recordings; ChatGPT interaction logs; teacher's instruction log; reflective diary

10 CONTRIBUTION

- **Comparative architecture:** Mediated vs. direct LLM interaction models in school settings
- **Feedback mapping:** Categorising LLM feedback levels (Hattie & Timperley) for L2 learners
- **Empirical baseline:** First longitudinal dataset on LLM-assisted Greek L2 learning

11 OPEN QUESTIONS

1. **Scaffolding:** How to fade scaffolding in an AI tool designed never to stop helping?
2. **Evaluation:** Learner satisfaction vs. actual learning outcome in LLM interaction?
3. **Validity:** How to account for LLM technology drift across a 3-year longitudinal study?

REFERENCES

Alanazi, A. (2016). A critical review of constructivist theory and the emergence of constructionism. *American Research Journal of Humanities and Social Sciences*, *2*, 1–8.

Anastasiadi-Symeonidi, A., Papadopoulou, D., Tzevelekou, M., Varlokosta, S., Iakovou, M., Kantzou, V., Lytra, V., Sioupi, A., Stamouli, S., Terzi, A., & Chondrogianni, V. (2011). *Ας μιλήσουμε Ελληνικά I, II, III [Let's Speak Greek I, II, III].* Monochromia.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112.

Liu, Z.-M., Hwang, G.-J., Chen, C.-Q., Chen, X.-D., & Ye, X.-D. (2026). Integrating large language models into EFL writing instruction: Effects on performance, self-regulated learning strategies, and motivation. *Computer Assisted Language Learning*, *39*(3), 466–490. https://doi.org/10.1080/09588221.2024.2389923

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge. *Teachers College Record*, *108*(6), 1017–1054.

Stavrakaki, S., & Tsimpli, I.-M. (2000). Διαγνωστικό τεστ γλωσσικής νοημοσύνης για παιδιά σχολικής και προσχολικής ηλικίας: Στάθμιση, στατιστική ανάλυση, ψυχομετρικές ιδιότητες [Diagnostic Verbal Intelligence Quotient test for school- and preschool-age children: Standardization, statistical analysis, and psychometric properties]. *Proceedings of the 8th Panhellenic Conference of Speech and Language Therapists* (pp. 95–106). Ellinika Grammata.

Sung, M.-C., & Kang, S. (2025). GPT API-based chatbots as adaptive and facilitative tutors for L2 English process writing. *Korean Journal of English Language and Linguistics, 25*, 448–471. https://doi.org/10.15738/kjell.25..202504.448