System 143 (2026) 104162

Contents lists available at ScienceDirect

System

journal homepage: www.elsevier.com/locate/system

# GenAI-assisted language teaching (GenAILT): Approach, design, and procedure

Keith Cheng Lin, Jack C. Richards *

*Sydney School of Education and Social Work, The University of Sydney, Education Building A35, Camperdown, NSW, 2006, Australia*

A R T I C L E I N F O

*Keywords:*
Generative AI
Language teaching
Language teaching methods
Pedagogical design
Teacher decision-making

A B S T R A C T

Since the release of ChatGPT in late 2022, Generative Artificial Intelligence (GenAI) has been rapidly taken up in second language teaching and learning. Much of the existing literature reports on perceptions, tools, techniques, and short-term outcomes; however, the broader question of what kind of pedagogy is emerging has received less attention. This conceptual paper addresses that question by analysing GenAI-assisted Language Teaching (GenAILT) through Richards and Rodgers' (2014) framework of Approach, Design, and Procedure. At the level of Approach, we demonstrate that GenAI is deployed under quite different theoretical assumptions, and we distinguish declared theory from the commitments GenAI enacts through its training data, which favour dominant language varieties unless teachers counter them. At the level of Design, we argue that objectives in GenAILT are better understood along a product–process continuum, that GenAI functions as a flexible resource for realising existing syllabus types, and that activity design must build in strategic, critical and ethical engagement with GenAI output. We further propose a four-role matrix for learners along the dimensions of learning orientation and instructional scaffolding, examine how teacher work is redistributed, reframe instructional materials in terms of form and quality assurance, and discuss classroom assessment as a design decision. At the level of Procedure, we illustrate that competent GenAILT depends on teacher decision-making. We close with implications for teacher education and research.

## 1. Introduction

The release of ChatGPT in late 2022 accelerated the widespread adoption of Generative Artificial Intelligence (GenAI) applications across various domains, including second language (L2) teaching and learning. The diversity of GenAI applications, along with their rapidly evolving functions, makes these tools increasingly attractive for teachers, learners and policymakers (Funa & Gabay, 2025). Yet its growing uptake raises questions that go beyond immediate practical applications. What theoretical assumptions about language and language learning underpin the use of GenAI in language teaching? What does pedagogical design look like when GenAI enters the language classroom?

The concern is not new. Earlier work in Computer-Assisted Language Learning (CALL) and Intelligent CALL (ICALL) made strikingly similar arguments about previous generations of technology. Matthews (1992, p. 14) argued that ICALL systems should be "embedded within well-motivated theories of second language acquisition". Oxford (1995) made a related observation, noting that most ICALL developers did not cite or describe any underlying theories of learning, and calling for the field to be linked more explicitly to

* Corresponding author.
*E-mail addresses:* keith.lin@sydney.edu.au (K.C. Lin), tiromoana1001@gmail.com, jack.richards@sydney.edu.au (J.C. Richards).

https://doi.org/10.1016/j.system.2026.104162
Received 24 May 2026; Received in revised form 16 September 2026; Accepted 20 September 2026
Available online 21 September 2026

contemporary theories of language acquisition, learning, and use. Subsequent work in CALL has continued this line of inquiry (e.g., Chapelle, 2008). In this paper, we extend this lineage to GenAI, arguing that the growing use of GenAI in language teaching should not simply be treated as a repertoire of appealing techniques to add to teachers' "bag of tricks", but should be examined in relation to the theoretical foundations, pedagogical design, and practical procedures of language teaching and learning.

This paper draws on Richards and Rodgers' (2014) framework of Approach, Design, and Procedure to analyse *GenAI-assisted Language Teaching (GenAILT)*, proposed here as a provisional umbrella term. GenAILT refers to the teaching and learning of a language in which GenAI systems are used for pedagogical purposes, by teachers, by learners, or by both. GenAI here refers to a subset of artificial intelligence (AI) designed to generate new content, including text, code, images, audio, and video (Law, 2024; Moorhouse & Wong, 2025). Unlike the rule-based systems that preceded it, GenAI is not confined to pre-programmed responses. It is based on large language models (LLMs), which generate content by identifying patterns and relationships in their training data and producing statistically probable outputs in response to prompts (Moorhouse & Wong, 2025). Pedagogical purposes, in turn, encompass teacher-facing uses such as planning lessons, preparing materials, and generating feedback, and learner-facing uses such as practising interaction, seeking explanations, and revising written work (Law, 2024; Wang et al., 2025). We state this scope explicitly because terms in this area are often used interchangeably, and such looseness can obscure what is being claimed (Wang et al., 2025). In naming the term, we take no prior position on whether it amounts to a fixed method, a question we return to in the conclusion.

This paper aims to provide a conceptual account of GenAILT and to offer language teachers and teacher educators a framework for evaluating its pedagogical possibilities and limitations. To achieve this aim, we address three fundamental questions that correspond to the levels of Approach, Design, and Procedure, respectively.

1. What theoretical assumptions about the nature of language and language learning does GenAILT reflect?
2. What implications does GenAILT have for pedagogical design, namely its objectives, syllabus, learning and teaching activities, the roles of teachers, learners, and materials, and assessment?
3. How should teachers make decisions regarding the use of GenAILT in practice?

## 2. GenAILT at the level of approach

According to Richards and Rodgers (2014), the *Approach* level refers to the theoretical foundation of a method, encompassing assumptions and beliefs about the nature of language and language learning that underpin teaching practices. At this level, two theoretical components are specified: a theory of language that accounts for the model of language competence and the basic features of linguistic organisation and use, and a theory of language learning that describes the central processes of learning and the conditions believed to promote successful language acquisition.

To identify the assumptions underlying the use of AI in language teaching, we examined recent empirical studies on the use of GenAI and found that it cannot be described as a teaching method in the strong sense of embodying one stable, declared view of language and learning. Even within a study, the application of AI in teaching may reflect a mixture of several theoretical perspectives. This can be seen in Table 1, which presents four illustrative AI interventions, selected because they are underpinned by different theories at the level of Approach. For each study, we asked what view of language and what view of language learning the intervention implies, and whether the activity does in fact reflect its theoretical assumptions, since the two do not always coincide.

Reading across the rows in Table 1 reveals theoretical heterogeneity. Tai and Chen's (2024) speaking intervention, in which learners hold extended oral dialogues with a chatbot alone or with a peer, treats language as a vehicle for interpersonal exchange and learning as advancing through interaction and, in the paired condition, joint construction. Shi et al.'s (2025) writing intervention, by contrast, has learners draft an essay, prompt ChatGPT to identify errors and improve the text, and then revise on that basis. This feedback-and-revision cycle enacts an implicitly structural view of language, as a set of formal and textual features to be corrected and improved, even though the study itself declares a sociocultural framework. The design treats improvement as accumulating through practice under feedback and as driven by conscious attention to the language system, combining skill-learning and cognitive-code assumptions. Zheng's (2024) reading intervention treats language as a system of decomposable components (e.g., vocabulary, grammar, background knowledge) that learners process individually with on-demand chatbot support, embodying a cognitive theory of language combined with interactionist and skill-learning assumptions about acquisition. Tai et al.'s (2026) genre-based writing intervention sits in a different theoretical neighbourhood altogether by integrating ChatGPT into the Teaching-Learning Cycle. It treats

**Table 1**
GenAI-supported pedagogical interventions in empirical studies and their implied theories of language and language learning.

| Pedagogical intervention | Core learner activity | Theories of language | Theories of language learning |
|---|---|---|---|
| Individual or pair interaction with a GenAI chatbot for speaking practice (Tai & Chen, 2024). | Extended oral dialogue with a chatbot, alone or with a peer | Interactional | Interactionist; Sociocultural |
| Using ChatGPT-generated feedback to revise argumentative essays (Shi et al., 2025). | Cycles of text production, AI feedback, and revision | Structural | Skill learning; Cognitive-code learning |
| Using a GenAI-based chatbot to support reading comprehension (Zheng, 2024). | Reading texts with on-demand chatbot support for comprehension obstacles | Cognitive | Interactionist; Skill learning |
| ChatGPT integrated into genre-based instruction for argumentative writing (Tai et al., 2026). | Deconstruction, joint construction, and independent writing within the Teaching-Learning Cycle | Genre | Sociocultural (genre pedagogy tradition) |

language as a resource for making meaning within socially recognised text types, and learning as a process of apprenticeship into genre conventions through scaffolded co-construction. Across these four cases, the use of GenAI is mobilised within different theoretical commitments, and even within a single intervention, the theoretical assumptions about language and about learning are sometimes layered, and sometimes pull against each other.

This theoretical heterogeneity is a strength because GenAI can be adapted to different curricular settings and institutional constraints (see further discussion in Section 3). However, GenAI is not a neutral tool awaiting a theory from its users. No one explicitly built these systems around a model of language competence or a view of how languages are learned, but their output still implicitly carries enacted theoretical commitments. LLMs learn language from statistical patterns in vast text corpora, so what they produce favours the varieties, registers, and styles that dominate that data (Kasneci et al., 2023). These commitments surface in practice. For example, Yeo et al. (2025) report that the audio-generation function of NotebookLM produces only American-accented speech and does not vary accents on request, and Kobak et al. (2025) document a recognisable stylistic uniformity in GenAI-influenced academic writing. Both findings point toward a narrowing of the language varieties learners encounter through GenAI-generated materials. This narrowing reconnects with long-standing concerns about native-speakerism and inner-circle norms in English language teaching (Holliday, 2006). A more normative worry is that, because GenAI produces fluent output with little effort, teachers may default to transmission-oriented uses such as generating drills or supplying corrective feedback, even where more demanding designs would better serve learning (Moorhouse & Wong, 2025; Pack & Maloney, 2024). Unless teachers notice and counter its defaults, these commitments favour dominant varieties and undemanding pedagogy. This is why teacher mediation matters at the level of Approach. Teachers need to reflect continuously on the theories of language and language learning that their use of GenAI enacts, and equally on the commitments the technology brings with it, before any pedagogical decision has been made.

## 3. GenAILT at the level of design

While the level of Approach concerns the assumptions about language and language learning that GenAILT may reflect, these assumptions do not in themselves determine what a course or lesson looks like. In Richards and Rodgers' (2014) terms, this is the work of *Design*, the level at which objectives are specified, content is selected and organised, activity types are chosen, and roles are assigned to teachers, learners, and instructional materials. Because GenAILT does not reflect a unified theoretical foundation, it does not embody one stable design. Instead, it can be incorporated into very different pedagogical designs depending on whether it is used for controlled practice, communicative interaction, genre modelling, learner support, feedback, or assessment. We will analyse the impact of GenAI on key components of the Design level, including a discussion of classroom assessment, to elaborate on this heterogeneity.

### *3.1. Objectives*

*Objectives* may be *product-oriented*, primarily focused on linguistic attainment, such as grammar, vocabulary, pronunciation, or communicative ability. Alternatively, objectives may be *process-oriented*, focusing on learning behaviours, capacities and processes that learners are expected to develop through instruction (Richards & Rodgers, 2014). The product–process orientation of a method is evident in the emphasis it places on vocabulary and grammatical proficiency, as well as in how errors in grammar and pronunciation are addressed (Richards, 2017). Extending this, the objectives are better understood as distributed along a continuum, not as a strict binary.

We can therefore ask where GenAILT sits on this continuum and what additional objectives it introduces. As shown in Fig. 1, GenAI may be incorporated into a more product-oriented design, into a more process-oriented design, or into designs that attempt to combine both. For a product-oriented design, GenAI may function primarily as a tool for supporting language performance and practice, for example, by generating exercises, supplying corrective feedback, or acting as a conversational partner (e.g., Muñoz Muñoz et al., 2025; Rezai et al., 2024). In a more process-oriented design, GenAI acts more like a learning advisor. It can prompt reflection, advise on strategies, scaffold revision, offer affective support, or adapt feedback to learner needs (e.g., Tram et al., 2024; Wu et al., 2025; Zhan & Yan, 2025).

GenAI's strongest pedagogical potential may lie in designs that combine product- and process-oriented objectives, as Liu et al. (2025) found that it can both support higher-order thinking and undermine it when used uncritically. In particular, its strategic,

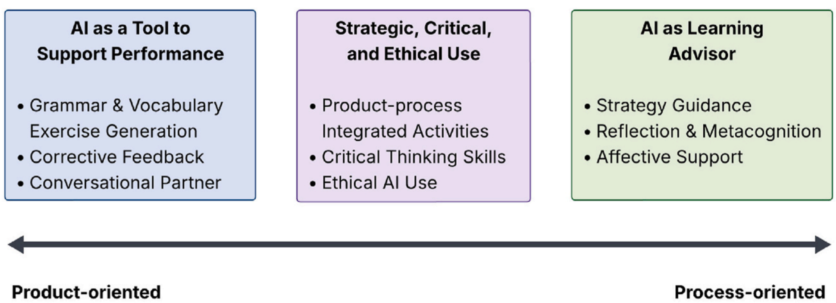


**Fig. 1.** Positioning GenAILT along a product–process continuum.

critical, and ethical use becomes pedagogically relevant, regardless of where the design is initially located on the continuum. This is because of the risks that emerge whenever GenAI is incorporated into language teaching, and these risks work against the very objectives that motivated the use of GenAI in the first place.

The first risk is overreliance. Across writing and other skill domains, learners report depending on GenAI to a degree that they themselves recognise as excessive, and this dependence reduces independent composing, critical thinking, and learner agency (Liu et al., 2025; Lu & Zeng, 2025; Mekheimer, 2025; Shi et al., 2025; Wiboolyasarin et al., 2025). In writing research specifically, this concern is echoed by Li and Wilson's (2025) review, which warns that AI support can become a substitute for learners' own idea generation, content development, and decision-making if it is not carefully scaffolded. The pattern is not confined to self-report. Deng et al.'s (2025) meta-analysis found that effect sizes for higher-order thinking under ChatGPT interventions decreased with longer exposure, suggesting that prolonged unstructured use may erode rather than build the cognitive engagement the tool was meant to support.

Surface engagement is another risk. Studies suggest that learners tend to engage more readily with feedback on grammar, spelling, wording, and sentence-level revision, while feedback on organisation, coherence, argument development, and critical thinking can be harder to interpret and act on (Yeung, 2025; Zhang, Aubrey, et al., 2025). This pattern is unevenly distributed by proficiency and regulatory capacity. Lower-proficiency writers may focus disproportionately on lower-order concerns, while higher-proficiency or more self-regulated writers are better positioned to evaluate AI suggestions critically and engage with higher-order revision (Nguyen et al., 2024; Yeung, 2025; Zhang, Zhou, & Bai, 2025).

Moreover, a prevalent concern is that undisclosed or undocumented GenAI use makes learner work opaque to the teacher, who cannot tell what the learner did, what GenAI did, and where learning occurred or did not occur. This compromises the teacher's capacity to set, monitor, and adjust objectives, regardless of where the design sits on the continuum (Funa & Gabay, 2025; Moorhouse et al., 2023; Pack & Maloney, 2024).

For these reasons, the strategic, critical, and ethical use of GenAI should be treated as a baseline objective of GenAILT, pushing it toward the process-oriented end of the continuum. It is the set of capacities that prevents overreliance, surface engagement, and opacity from undermining whatever other objectives the design is pursuing.

### 3.2. Syllabus

*Syllabus* refers to the way the content of a course is selected and organised, whether in grammatical, lexical, functional, situational, topical, task-based, or text-based terms (Richards & Rodgers, 2014). It is concerned with the principles by which content is prioritised, sequenced, and made available for learning. Some methods rely on a *priori* syllabus specified in advance, whereas others allow content to emerge more contingently through classroom interaction, resulting in a *posteriori* syllabus (e.g., Community Language Learning).

From this perspective, GenAILT does not constitute a syllabus type. Although it can generate language materials rapidly in response to teacher prompts, it is not the same as syllabus design. A syllabus provides an organising principle for content selection and sequencing. GenAI, by contrast, functions primarily as a resource for generating, adapting, and elaborating content within such principles, although it can also be used to develop a syllabus. Therefore, it is better understood as a flexible means through which existing syllabus types may be realised either through text, skills, or grammar. This aligns with recent arguments that GenAI functions as a pedagogical resource rather than a curriculum framework (Moorhouse & Wong, 2025).

The fact that GenAI can be incorporated into multiple, quite different syllabus logics further indicates that it is not a syllabus principle. In a grammatical syllabus, GenAI may be used to generate grammar-focused practice materials, including drills and other controlled exercises, sentence-level reordering or paraphrasing tasks, vocabulary practice, and contextualised examples of target structures (Peachey & Crichton, 2024). In a functional syllabus, it may support learners' engagement with speech acts and transactional language through simulated, task-based exchanges such as interviews, discussions, and role-plays (Du & Daniel, 2024; Yang et al., 2022). In a task-based syllabus, it may assist learners in planning, rehearsing, completing, and reflecting on pedagogic and real-world tasks through scaffolded, goal-oriented interaction and feedback (Shafiee Rad & Roohani, 2025; Yang et al., 2022). In a text-based syllabus, it may be used to model and analyse discourse types such as narratives, reports, explanations, essays, and emails by providing genre-sensitive feedback on structure, coherence, and language use (Polakova & Ivenz, 2024; Zhang, Aubrey, et al., 2025). These suggest that GenAI is a versatile instructional resource that can be aligned with multiple ways of organising course content.

What GenAI changes more substantially is the mode through which a syllabus can be realised in practice. In addition to supporting predetermined content (e.g., textbooks), GenAI allows teachers to generate explanations, examples, tasks, texts, audio, and other resources responsively as learner needs arise during instruction (Moorhouse & Wong, 2025). In this sense, GenAI may make it easier to mediate between a priori and a posteriori syllabus. Teachers may still work within a broader pre-planned syllabus while using GenAI to elaborate, adapt, or supplement content contingently in real time (e.g., a teacher realises that students need additional practice on inferential reading, which is not sufficiently provided for in their textbook). Thus, GenAILT increases the responsiveness of syllabus enactment without removing the need for an underlying curricular logic.

This responsiveness is pedagogically significant, but it also raises important questions regarding what AI-generated content is selected, validated, and prioritised pedagogically (Moorhouse & Wong, 2025; Pack & Maloney, 2024). Without clear syllabus principles and objectives, such content may become excessive, insufficiently filtered, or only loosely aligned with course goals. Because GenAI responds effectively to immediate prompts, it may support local instructional needs without ensuring longer-term coherence across a course unless teachers actively maintain a principled sequence and keep a human in the loop (Moorhouse & Wong, 2025). Additionally, AI-generated output cannot be assumed to be accurate or reliable, making validation and teacher judgment essential (Pack & Maloney, 2024). Thus, it heightens the need for principled content selection, sequencing, and teacher judgement.

### 3.3. Types of learning and teaching activities

*Types of learning and teaching activities* show how broader assumptions about language and learning at the Approach level are realised in classroom practice (Richards & Rodgers, 2014). Different methods are often distinguished most clearly at this level, since activity types make visible what learners are actually asked to do and how learning is expected to occur. Activities may be oriented toward grammatical accuracy, communication, problem-solving, text analysis, or other forms of participation, thereby linking objectives and syllabus design to classroom work (Richards & Rodgers, 2014).

GenAI has been integrated across a wide range of activity domains, including oral interaction and conversational practice (Wiboolyasarin et al., 2025), writing instruction across multiple composing stages (Shi et al., 2025), and task-based language teaching (Li et al., 2026). This breadth of integration means the tool itself should not be the unit of analysis. Significance must lie in how the activity around the tool is structured.

To analyse the structure of an activity, we can ask again what changes GenAI has made. Feedback is no longer bounded by a teacher turn, a peer review session, or a marking cycle, but is continuous and available on demand. As a consequence, the boundary between production and evaluation blurs, since learners can query, revise, and re-query within a single activity episode. Cognitive work is also split differently between learner and tool, as planning, lexical retrieval, and argument construction can all be partially delegated to GenAI (see Chen et al., 2025; Godwin-Jones, 2024). What the learner actually does during an activity can differ substantially from what its surface description suggests.

These shifts open a productive tension, where learners can engage with GenAI output as a stimulus for reflection and revision, or they can treat it as an endpoint, incorporating suggestions without evaluation. Shi et al. (2025) found that students using ChatGPT for writing revision achieved significantly higher scores than those using conventional automated feedback, but qualitative analysis reveals widespread overreliance and a reported loss of creative agency. The performance gain and the learning cost showed up together in the same activity.

This reinforces our earlier argument that the strategic, critical, and ethical use of GenAI should be a baseline objective. At the level of activity design, this means that evaluative engagement with GenAI output must be built into activities. Learners should be asked to

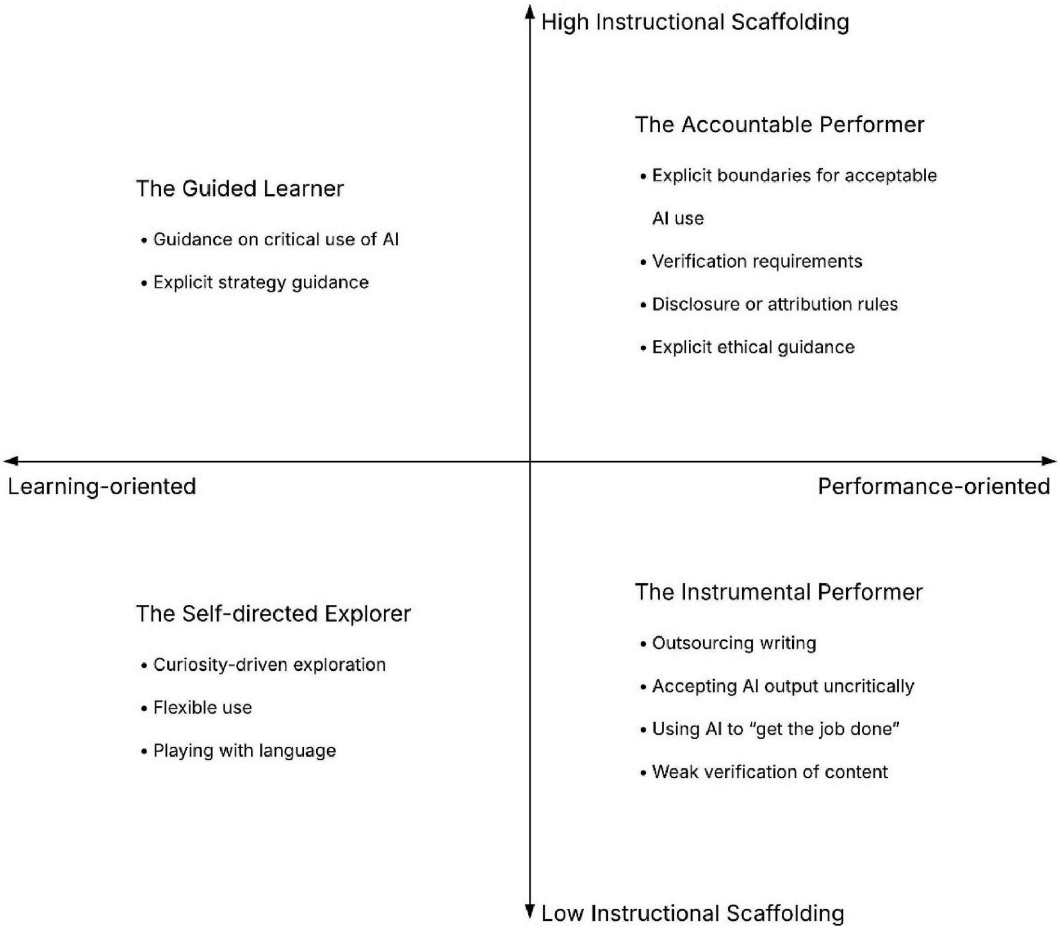


**Fig. 2.** Learner roles in GenAILT: A matrix of learning orientation and instructional scaffolding.

analyse the GenAI output as a learning process. For example, learners might draft a paragraph before consulting GenAI and then justify which suggestions they accept or reject; compare two GenAI outputs and identify which one better fits the target genre; or, in oral interaction, annotate which pragmatic moves the GenAI performs and which it misses. The common feature of such designs is that the AI-generated output becomes an object of analysis within the activity, leading to a process-oriented learning. This keeps the cognitively demanding work with the learner, and gives the strategic, critical, and ethical use of GenAI a concrete location in the lesson.

### *3.4. Learner roles*

*Learner roles* refer to how a method positions learners within the learning process, including the activities they perform, the control they exercise over content, their patterns of interaction, and whether they are viewed as performers, initiators, problem-solvers, or active processors of language (Richards & Rodgers, 2014). Over time, learner roles have shifted from passive recipients of instruction to more learner-centred participants in learning, with increased attention given to learner autonomy and learning strategies as part of the learner's contribution to language development (Benson, 2001; Nunan, 1988; Oxford, 1990).

The previous activity section leaves a further question open. If evaluative engagement with GenAI output is what distinguishes learning-oriented use from performance-oriented use, and if such engagement depends on how the activity is structured, then learners will be positioned very differently depending on the purpose of GenAI use and how the activity or teacher guides them to engage with it. Thus, the purpose of use and the degree of pedagogical guidance can be treated as two dimensions along which learner roles in GenAILT vary.

Fig. 2 presents a matrix that proposes four learner roles as heuristic categories along these dimensions. On the horizontal axis, drawing on goal orientation theory (Dweck, 1986; Elliott & Dweck, 1988), we distinguish GenAI use aimed at language development (i.e., learning-oriented) from use aimed at task completion (i.e., performance-oriented). On the vertical axis, instructional scaffolding refers to the degree of guidance and constraint built into pedagogical design (Wood et al., 1976).

The four learner roles presented in Fig. 2 suggest different ways in which GenAILT may position learners. The *Self-directed Explorer* is given relatively high freedom to use GenAI as an independent resource for low-stakes, learning-oriented activity. This may be an autonomy-promoting role, though without pedagogical guidance, learning outcomes are more likely to be incidental and harder to assess (Sockett, 2014). The *Guided Learner* is guided toward critical and strategic use of AI to develop their language. This may be the strongest role for intentional language development, though the degree of guidance requires careful calibration, as over-scaffolding can work against the development of learner autonomy (Daniel et al., 2016). The *Instrumental Performer* is positioned primarily as a task-completer who uses AI instrumentally with limited oversight. This is arguably the least desired role which carries the greatest risks of over-reliance (Li & Wilson, 2025), surface engagement and weak verification (Zhang, Zhou, & Bai, 2025), and academic integrity concerns (Funa & Gabay, 2025). The *Accountable Performer*, by contrast, uses GenAI for performance-oriented purposes under explicit expectations of verification, disclosure, and responsibility for the final product. This may be the most sustainable role in high-stakes contexts, though its learning value depends on whether accountability requirements push learners toward genuine self-monitoring or merely toward meeting external criteria.

In reality, a learner may be positioned anywhere in the matrix. Their use of GenAI may be both learning-oriented and performance-oriented, or learning might occur during a performance-oriented use of GenAI. This means that the learner roles are not as clear-cut as the matrix might imply. Movement along the horizontal axis follows from how a course is framed and what it rewards. Objectives that name language development, classroom activities that guide learners toward strategic, critical and ethical use of GenAI, and assessment criteria that credit the process of development rather than the quality of the finished text, pull the use of GenAI toward a learning orientation. Where a task rewards only the product, learners have little reason to treat GenAI as anything other than a means of completing it (see also discussion about assessment of learning in Section 3.7). Movement along the vertical axis follows from the degree of guidance and constraint the teacher builds into the design, including how tightly the tool's behaviour is configured, how far the task specifies what learners may use it for, and whether learners are taught to interrogate what it returns. The sequencing decision discussed in Section 3.7, where learners judge their own draft before soliciting GenAI's appraisal, is one such move.

What this matrix contributes is that it asks teachers to consider what purpose learners are using GenAI for, what challenges or risks learners may encounter under different purposes, what kind of support they can provide, and what moves are available if a learner is positioned somewhere other than where the course intends.

### *3.5. Teacher roles*

*Teacher roles* refer to how a method positions teachers in relation to the instructional system, encompassing four sub-issues: the *functions* they fulfil, the degree of *control* they exercise over how learning takes place, their responsibility for *content*, and the *interactional patterns* they establish with learners (Richards & Rodgers, 2014). Teacher roles are tightly coupled to learner roles, and have shifted historically from the audiolingual conception of the teacher as primary source and model, to more facilitative conceptions in Communicative Language Teaching, Task-Based Language Teaching and constructivist methods, and to the reflective and ecological conception of the teacher as a practitioner who observes, listens, and adapts to learners' emerging needs (Richards, 2017; Richards & Rodgers, 2014).

GenAILT changes how teacher work is divided across the four sub-issues (functions, control, content, and interactional patterns). Several traditional functions (e.g., providing content, modelling target language, generating practice, supplying corrective feedback) can now be partially performed by GenAI (Jeon & Lee, 2023; Moorhouse & Wong, 2025), but a set of meta-functions is added, such as validating GenAI output, calibrating its use to particular learners, maintaining a "human in the loop" (Moorhouse & Wong, 2025; Pack

& Maloney, 2024). Teachers are also required to mediate across macro-, meso-, and micro-level demands (The Douglas Fir Group, 2016). At the macro level, they are often expected to implement institutional policies on acceptable AI use. At the meso level, they need to align AI-supported tasks with the curriculum. At the micro level, they need to respond to learners' immediate needs, proficiency levels, and learning difficulties. Across these three levels, teachers' functions converge on a single task, namely negotiating the use of GenAI between institutional constraints, curricular coherence, and local pedagogical responsiveness.

Furthermore, control becomes more diffuse, as learners interact with GenAI outside the teacher's sightline (Godwin-Jones, 2024), shifting teacher control from moment-to-moment management toward earlier task design and later review. Teachers' responsibility for content also increases. GenAI produces fluent text very quickly, but the teacher still has to judge whether it is accurate, culturally appropriate, and aligned with the course. The interactional patterns of the teacher–learner dyad now become a teacher–GenAI–learner triad. The teacher's job is increasingly to orchestrate productive interaction between learners and GenAI, where the teacher actively designs the conditions under which learner–GenAI interaction takes place. GenAI thus moves much of teacher work into design and review, and raises the level of professional judgement and decision-making the work demands (see Section 4 for a discussion of teacher decision-making).

### 3.6. The role of instructional materials

*The role of instructional materials* refers to how a method positions materials within the instructional system, including their *primary goal* (whether they present content, practise content, facilitate communication, or enable independent learning), their *form* (textbook, audio, software, realia), their *relation to other sources of input*, and the *teacher abilities* they assume (Richards & Rodgers, 2014). Different methods assign materials different roles. For example, in Task-Based Language Teaching, they create the need for negotiation of meaning and interaction; in Text-Based Instruction, they model the features of texts learners are expected to produce. Earlier forms of technology-mediated materials, including CALL resources, web-based materials, and adaptive platforms, had made instructional materials more flexible and less tied to printed textbooks (Chapelle, 2008).

What GenAI changes is the speed, scale, multimodality, and distributed control of material generation. Materials can be generated in response to a teacher's prompt or a learner's query, at whatever point in the instructional cycle is useful, and can take whatever form the moment requires (Law, 2024). Teachers may still prepare materials in advance, as they always have, but they may equally generate or adapt them mid-lesson in response to emerging learner needs, or have learners generate their own materials outside class. This expansion reconfigures each of Richards and Rodgers' (2014) four dimensions for analysing materials, as summarised in Table 2.

Of these four dimensions, form is the most visible change. A teacher who once chose between a textbook, an audio file, and a worksheet now works with a single tool that produces all three. The other three dimensions follow from this. A single artefact can serve more than one purpose at once. An AI-generated dialogue can present new language, give learners something to practise, and provide feedback on their attempts within one episode, collapsing distinctions that earlier materials kept separate. The relation to other inputs becomes permeable because the boundary of the planned instructional environment no longer holds. A learner who queries a chatbot at home is bringing additional input into the course without the teacher having selected or sequenced it. And the abilities materials assume of teachers shift from selecting, sequencing, and exploiting vetted artefacts to prompting, evaluating, and quality-assuring fluent multimodal output in real time (Pack & Maloney, 2024).

These changes raise the stakes and redistribute who does the work. Quality assurance migrates from publishers to teachers, since AI-generated materials are neither edited nor piloted before use. As discussed before, AI-generated materials also carry representational risks related to inner-circle accents and stylistic uniformity (see Section 2), which may narrow learners' exposure to language varieties. Teachers, hence, need to be aware of these risks and take quality assurance responsibilities when applying GenAILT.

### 3.7. Classroom assessment

Assessment is not clearly located in Richards and Rodgers' (2014) framework, but it surfaces in connection with feedback at the level of Procedure. Teaching and assessment are primary instructional activities in virtually every method, and the decisions that shape assessment are made when a course is planned. Biggs' (1996) principle of constructive alignment makes the point directly. A course works when its objectives, its teaching and learning activities, and its assessment are designed together, so that the activities elicit the

**Table 2**
Reconfiguration of materials across Richards and Rodgers' (2014) four dimensions under GenAILT.

| Dimension | Earlier instructional materials | GenAILT materials |
|---|---|---|
| Primary goal | Often designed for relatively specialised functions, such as presenting content, supporting practice, facilitating communication, or enabling independent learning | Potentially multifunctional, since a single AI-generated artefact may present input, support practice, provide feedback, and prompt reflection within the same learning episode |
| Form | Increasingly varied across print, audio, video, software, web-based resources, and realia, but usually selected or prepared before use | Highly mutable and multimodal, with materials generated, revised, or reformatted on demand across modes |
| Relation to other input | Usually positioned within a designed instructional environment, even when supplemented by authentic or digital resources | More permeable, since learners and teachers can generate additional input inside or outside the planned instructional environment |
| Teacher abilities assumed | Ability to select, adapt, sequence, and exploit materials in relation to syllabus goals and learner needs | Ability to prompt, evaluate, curate, adapt, and quality-assure generated materials in real time and over time |

performances the objectives describe and the assessment establishes whether they were attained. Our focus here is classroom assessment, the assessment that teachers plan and carry out within a course, commonly discussed in terms of *assessment of learning*, *assessment for learning* and *assessment as learning* (Phakiti & Leung, 2024).

Assessment makes inferences about what learners know (i.e., *competence*) through what they show (i.e., *performance*) (Phakiti & Leung, 2024). This work proceeds in three stages. A learner produces a performance, a teacher or examiner makes inferences about what that performance shows about competence, and in some cases, feedback follows. Fig. 3 outlines how the assessment of, for, and as learning distribute this work and where GenAI can enter it.

Fig. 3 shows that assessment of learning usually ends in judgment. It seeks to determine whether students have met a course's objectives, and does not primarily aim to provide feedback for learning (Phakiti & Leung, 2024). This makes the threat to valid inference serious, since an uninvigilated performance may have been produced jointly by the learner and GenAI, or in some cases solely by GenAI. Therefore, many institutions have returned to invigilated examinations, but this retreat narrows what a course can assess, particularly sustained writing (Moorhouse et al., 2023). The strategies collected in Moorhouse et al.'s (2023) review are attempts to cope with this. Teachers are advised to run existing tasks through GenAI to see what the tool alone can produce, to stage assessment so the process of writing is visible, and to build AI use into the task with acknowledgment required. However, in assessment of learning, GenAI can easily be seen as a problem to be contained so that valid inferences remain possible, rather than an opportunity to enhance assessment and learning. Therefore, it is necessary to incorporate assessment for and as learning in any GenAILT design.

GenAI plays a different role in assessment for learning, in that it can support the teacher's work of giving feedback. What changes is not the source of feedback so much as who decides what feedback a learner receives. Muñoz Muñoz et al. (2025) compared direct written corrective feedback from ChatGPT with feedback from a trained teacher across four writing sessions and found that both improved students' writing, with ChatGPT performing better across the assessed criteria. They nonetheless position the tool as complementary to the teacher rather than a substitute, since generating useable feedback depends on carefully designed prompts, and judging whether that feedback suits a particular learner and context calls for the teacher's own evaluation. Thus, it is the teacher who decides which feedback functions to delegate to the tool, what to review before it reaches learners, and what to reserve for their own judgment, so that the feedback learners receive still rests on a teacher's gatekeeping.

Assessment as learning treats the appraisal of one's own work as itself a learning activity, exercised through self-assessment, peer-assessment, and reflection (Phakiti & Leung, 2024). The learner is the one expected to draw the evaluative inference, and, as Fig. 3 shows, the one who carries all three stages. GenAI can take this work over entirely, so self-assessment outsourced to the tool exercises none of the judgment it is meant to develop and may contribute to the deskilling associated with uncritical AI adoption (Hyland, 2026). What is different from assessment for learning is that the key design decision here is sequencing. Learners can first judge a draft against the criteria, then solicit GenAI's appraisal, and finally examine discrepancies between the two. Used in this sequence, AI becomes an object of reflection within the self-regulatory cycle (Hou & Zhou, 2025). The teacher's role is to prepare learners for it by setting the criteria they judge against, teaching learners how to interrogate what the tool returns, and building the habit of self-reflection before consulting the tool. Designed this way, assessment as learning positions the learner as a Guided Learner in the sense set out in Section 3.4.

Across the three functions of assessment, GenAI can assist at every stage. Classroom assessment under GenAILT should be

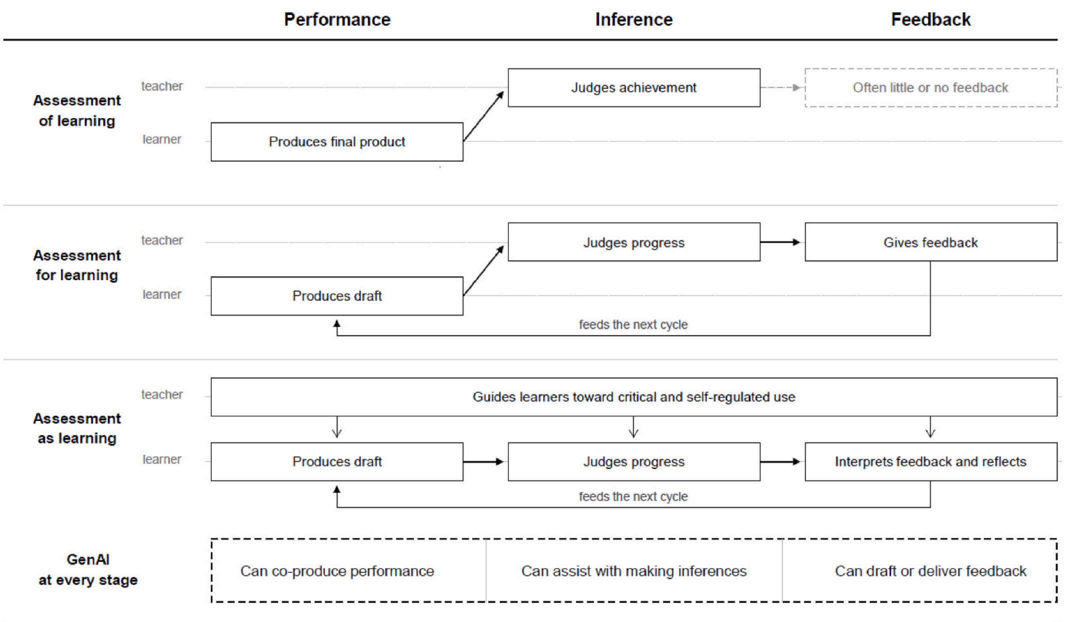


**Fig. 3.** Classroom assessment in GenAILT.

considered not only in terms of how valid inferences can be made but also how GenAI can support feedback, self-reflection, self-regulation and learning overall, rather than being treated as a risk to be contained.

## 4. GenAILT at the level of procedure

*Procedure* is the level at which Approach and Design are realised in classroom behaviour. Richards and Rodgers (2014, p. 36) describe it as the “moment-to-moment techniques, practices, and behaviors that operate in teaching a language,” and identify three dimensions: the activities used to *present* new language, the ways activities are used for *practising* language, and the techniques used for giving *feedback* on learner output. In what follows, we use these three dimensions as our analytic anchor for what teachers do in a GenAILT lesson.

We argue that GenAI changes the *decisions* teachers must make within each of the three dimensions, and that these decisions are distributed across time. To capture this temporal distribution, we draw on Jackson's (1968) account of teacher decision making as taking place in three stages, namely preactive (planning before the lesson), interactive (in the moment of teaching), and postactive (reflection after the lesson). Following Westerman (1991), we further argue that competent GenAILT depends on the teacher's capacity to integrate the three into a feedback loop, so that what is learned about students in one lesson informs the planning of the next. Crossing Richards and Rodgers' three procedural dimensions with Jackson's three decision-making stages yields a three-by-three grid of decisions that maps the decisions teachers make in GenAILT procedure. Table 3 summarises this grid, and the hypothetical lesson that follows illustrates how the cells operate together in practice.

We illustrate these decisions through a hypothetical intermediate speaking lesson built around a job interview role-play, where the syllabus specifies hedging and self-presentation as the focus, with target expressions including “I would say”, “I tend to”, and “it depends on”. The illustration follows the lesson through its three phases and shows how postactive decisions at each phase feed into subsequent planning, which is what makes the grid operative.

In the presentation phase, the teacher's preactive decision concerns whether and how to combine textbook materials with AI-generated content, which is among the most commonly reported uses of GenAI in language teaching (Law, 2024). The teacher locates a short transcript of a sample exchange in the textbook and decides to extend it with additional AI-generated content. During the class, learners work in pairs to underline the hedges in the transcript and discuss what each one is doing. One pair asks whether “I guess” would work in the same position. Rather than answering directly, the teacher guides the students to ask GenAI in front of the class: “Can I replace ‘I would say’ with ‘I guess’ in this sentence: ‘I would say meeting customers’ needs is a priority’? Is there any difference in the meaning I express?” The teacher evaluates whether the generated explanations are accurate and decides whether further examples differentiating the two expressions are needed. Postactively, she reflects on the AI interaction record and finds that learners engaged more readily with the generated examples than with the transcript, as they drew on pre-existing knowledge of a familiar expression “I guess”. She decides to include a comparison activity in future lessons, contrasting target hedges with learner-familiar alternatives.

In the practice phase, learners pair off and conduct the role-play with a GenAI chatbot configured as the interviewer. Chen et al. (2025) found that role-play with a GenAI agent improved EFL learners’ speaking as much as role-play with peers, with greater gains in motivation, though not all learners responded equally well to the agent. The preactive decision is therefore how much of the activity to delegate to GenAI, and the interactive one is which learners need the teacher rather than the tool. The teacher sets up the chatbot beforehand with a prompt specifying the job (a part-time café position), the register (formal but friendly), and an instruction to ask follow-up questions when answers are short. Interactively, she circulates and reads the room. Two learners are giving one-word answers, and the chatbot is accepting them without probing further. She pauses the pair, asks them to read their last three exchanges aloud, and points out where a fuller answer would be expected. A second pair is doing well. One learner asks the chatbot to repeat its question more slowly, and her partner picks up a phrase from the GenAI and uses it in her own answer. The teacher leaves them to continue and notes the exchange for later use. Postactively, reviewing her notes and GenAI record, she decides to make turn length an opening discussion point and present communicative strategies with GenAI in the next class.

In the feedback phase, preactively, the teacher decides to delegate corrective feedback to GenAI. The delegation is partial by design, since GenAI feedback tends to improve grammar and sentence-level features while feedback combining GenAI with a teacher produces greater gains in organisation and critical thinking (Zhang, Aubrey, et al., 2025). Interactively, she monitors the chatbot's feedback as she moves between pairs. In one case, the chatbot does not point out issues with “I have worked there last year”, but the teacher notices

**Table 3**
Teacher decision-making in GenAILT procedure.

| | Preactive (planning) | Interactive (during the lesson) | Postactive (reflection and evaluation) |
|---|---|---|---|
| **Presentation** | Anticipating where generated input is needed and where teacher-prepared input is preferable. | Deciding when to generate further input on demand and when to curate or override it in response to learner uptake. | Reviewing which generated input worked and revising preactive planning accordingly. |
| **Practice** | Designing activities that exploit adaptive affordances of GenAI while preserving objectives. | Evaluating learner practice while retuning activities and redirecting GenAI. | Identifying which adaptive moves supported practice and which displaced it to inform future interactive decisions. |
| **Feedback** | Allocating feedback functions between the teacher and GenAI across the three stages. | Monitoring generated feedback and intervening where it misaligns. | Analysing generated feedback records to identify learners' recurrent difficulties for subsequent planning. |

the oddness and asks the chatbot whether the form is correct. The chatbot now flags the time reference, and the learner reformulates. In another pair, the chatbot offers a correction that the teacher considers too prescriptive for an interview context, and she intervenes briefly to explain that the original student phrasing was appropriate. Postactively, she downloads the transcripts and finds three patterns. Several learners avoided the target hedging expressions entirely. Two produced the present perfect accurately in slower, writing-like turns but not in spontaneous speech. And the chatbot missed three grammatical errors of the same type. She plans a short focused-practice activity on hedging for the next lesson and revises the chatbot configuration to attend to time reference.

These decisions are the ordinary substance of competent teaching, made newly visible by the presence of GenAI that can perform aspects of the work if the teacher does not (e.g., arguably providing a more well-rounded explanation of "I guess" vs. "I would say" and examples). However, these teacher decision-making processes cannot be replaced by GenAI, though they can be supported by it as a mediation for teachers' planning and reflection. We therefore return to our argument that at the procedure level, what makes GenAILT work (or not work) is teachers' moment-to-moment decisions and their ability to interconnect the three stages of decision making. Procedure in GenAILT is hence a sequence of teacher decisions about when to delegate, intervene, validate, and redesign. The implication for teacher education is that competent GenAILT cannot be developed by training teachers to operate the tool. It depends on the kinds of decision-making that have long defined good language teaching, now exercised in conditions that make their absence more consequential. This is why professional GenAI competence has been framed as combining knowledge of how the tools work with the pedagogical judgement to decide when their use serves learning (Moorhouse & Wong, 2025).

## 5. Conclusion

Across the three levels, we have shown that GenAILT does not exhibit the internal coherence that would conventionally qualify it as a method, and we have not argued that it should. Richards (1990, p. 35) describes methodology not as "something fixed, a set of rigid principles and procedures that the teacher must conform to", but as "a dynamic, creative, and exploratory process that begins anew each time the teacher encounters a group of learners". The contribution of this paper is to show where, within the levels of Approach, Design, and Procedure, that dynamic process takes place, and what teachers take on at each level. We have also proposed GenAILT based on the extended Approach, Design, and Procedure framework at each of its three levels. At Approach, the theoretical commitments in GenAILT are embedded in training data rather than declared, so teachers must surface what nobody has stated and decide what to counter. At Design, the components widen, since classroom assessment becomes unavoidable and the teacher–learner dyad becomes a teacher–GenAI–learner triad. At Procedure, crossing the procedural dimensions with preactive, interactive, and postactive decision-making brings the teacher's decisions into view. Across all three levels, theoretical and design work no longer arrives settled from a method's originator. It rests with the teacher, which is what the post-method literature anticipated (Kumaravadivelu, 1994).

For teacher education, preparation for GenAILT should be organised around the theoretical orientation at Approach, design choices across the components at Design, and decision-making at Procedure. The need is pressing, since teacher educators themselves report lacking the confidence and competence to address what GenAI implies for curriculum, instruction, and assessment (Moorhouse & Kohnke, 2024). For researchers, we suggest that studies of GenAI in language teaching identify their position at each of the three levels. Without this, two studies that both report learning gains can appear to be making the same claim when, in fact, they are making very different ones. A study that reports gains under a task-based design and a study that reports gains under a genre-based design are reporting the success of different theoretical commitments at Approach, different organising principles at Design, and different sequences of teacher decisions at Procedure. Aggregating them as evidence that "GenAI improves language learning" obscures the specific approach, design and procedure that produced the gain and makes the finding harder to interpret, replicate, or apply. Locating each study within the framework allows the field to ask sharper questions, namely what the intervention assumed about language and learning, how it was designed, and how teachers and learners operated within it, rather than treating "AI in language teaching" as a single object of inquiry.

## CRediT authorship contribution statement

**Keith Cheng Lin:** Writing – review & editing, Writing – original draft, Conceptualization. **Jack C. Richards:** Conceptualization, Writing – review & editing.

## Declaration of generative AI

During the preparation of this work, the authors used the enterprise account of Copilot (Microsoft) to assist with checking citations against references, APA 7th style, readability of sentences, and unnecessary repetition. After using this tool, the author reviewed and edited as needed and takes full responsibility for the content of the published article.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References


Benson, P. (2001). *Teaching and researching autonomy in language learning*. Longman.

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*(3), 347–364. https://doi.org/10.1007/BF00138871

Chapelle, C. A. (2008). Technology and second language acquisition. *Annual Review of Applied Linguistics, 27*, 98–114. https://doi.org/10.1017/S0267190508070050

Chen, A., Jia, J., Li, Y., & Fu, L. (2025). Investigating the effect of role-play activity with GenAI agent on EFL students' speaking performance. *Journal of Educational Computing Research, 63*(1), 99–125. https://doi.org/10.1177/07356331241299058

Daniel, S. M., Martin-Beltrán, M., Peercy, M. M., & Silverman, R. (2016). Moving beyond yes or no: Shifting from over-scaffolding to contingent scaffolding in literacy instruction with emergent bilingual students. *TESOL Journal, 7*(2), 393–420. https://doi.org/10.1002/tesj.213

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education, 227*, Article 105224. https://doi.org/10.1016/j.compedu.2024.105224

Du, J., & Daniel, B. K. (2024). Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. *Computers and Education: Artificial Intelligence, 6*, Article 100230. https://doi.org/10.1016/j.caeai.2024.100230

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41*(10), 1040–1048. https://doi.org/10.1037/0003-066X.41.10.1040

Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology, 54*(1), 5–12. https://doi.org/10.1037/0022-3514.54.1.5

Funa, A. A., & Gabay, R. A. E. (2025). Policy guidelines and recommendations on AI use in teaching and learning: A meta-synthesis study. *Social Sciences & Humanities Open, 11*, Article 101221. https://doi.org/10.1016/j.ssaho.2024.101221

Godwin-Jones, R. (2024). Distributed agency in language learning and teaching through generative AI. *Language, Learning and Technology, 28*(2), 5–31. https://hdl.handle.net/10125/73570.

Holliday, A. (2006). Native-speakerism. *ELT Journal, 60*(4), 385–387. https://doi.org/10.1093/elt/ccl030

Hou, F., & Zhou, C. (2025). How AI shapes self-regulated learning in foreign language education. *Discover Computing, 28*, Article 344. https://doi.org/10.1007/s10791-025-09801-0

Hyland, K. (2026). Writing in the AI era: Rethinking writing, research and teaching. *Journal of Second Language Writing, 72*, Article 101302. https://doi.org/10.1016/j.jslw.2026.101302

Jackson, P. W. (1968). *Life in classrooms*. Holt, Rinehart and Winston.

Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies, 28*(12), 15873–15892. https://doi.org/10.1007/s10639-023-11834-1

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kobak, D., González-Márquez, R., Horvát, E.-Á., & Lause, J. (2025). Delving into LLM-assisted writing in biomedical publications through excess vocabulary. *Science Advances, 11*(27), Article eadt3813. https://doi.org/10.1126/sciadv.adt3813

Kumaravadivelu, B. (1994). The postmethod condition: (E)merging strategies for second/foreign language teaching. *Tesol Quarterly, 28*(1), 27–48. https://doi.org/10.2307/3587197

Law, L. L.-h. (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. *Computers and Education Open, 6*, Article 100174. https://doi.org/10.1016/j.caeo.2024.100174

Li, M., & Wilson, J. (2025). AI-integrated scaffolding to enhance agency and creativity in K-12 English language learners: A systematic review. *Information, 16*(7), Article 519. https://doi.org/10.3390/info16070519

Li, Y., Shadiev, R., & Chiu, T. K. F. (2026). Applying generative artificial intelligence to task-based language teaching and learning: A systematic review and meta-analysis. *TechTrends, 70*, 150–163. https://doi.org/10.1007/s11528-025-01140-7

Liu, J., Sihes, A. J. B., & Lu, Y. (2025). How do generative artificial intelligence (AI) tools and large language models (LLMs) influence language learners' critical thinking in EFL education? A systematic review. *Smart Learning Environments, 12*, Article 48. https://doi.org/10.1186/s40561-025-00406-0

Lu, D. J., & Zeng, Y. (2025). Exploring the use of ChatGPT-generated model texts as a feedback instrument: EFL students' text quality and perceptions. *Innovation in Language learning and teaching*. Advance online publication. https://doi.org/10.1080/17501229.2025.2525341

Matthews, C. (1992). Going AI: Foundations of ICALL. *Computer Assisted Language Learning, 5*(1–2), 13–31. https://doi.org/10.1080/0958822920050103

Mekheimer, M. (2025). Generative AI-assisted feedback and EFL writing: A study on proficiency, revision frequency and writing quality. *Discover Education, 4*, Article 170. https://doi.org/10.1007/s44217-025-00602-7

Moorhouse, B. L., & Kohnke, L. (2024). The effects of generative AI on initial language teacher education: The perceptions of teacher educators. *System, 122*, Article 103290. https://doi.org/10.1016/j.system.2024.103290

Moorhouse, B. L., & Wong, K. M. (2025). *Generative artificial intelligence and language teaching*. Cambridge University Press. https://doi.org/10.1017/9781009618823

Moorhouse, B. L., Yeo, M. A., & Wan, Y. (2023). Generative AI tools and assessment: Guidelines of the world's top-ranking universities. *Computers and Education Open, 5*, Article 100151. https://doi.org/10.1016/j.caeo.2023.100151

Muñoz Muñoz, B. C., Nassaji, H., & Bello Carrillo, F. I. (2025). ChatGPT-generated versus human direct corrective feedback on L2 writing. *System, 134*, Article 103805. https://doi.org/10.1016/j.system.2025.103805

Nguyen, A., Hong, Y., Dang, B., & Huang, X. (2024). Human-AI collaboration patterns in AI-assisted academic writing. *Studies in Higher Education, 49*(5), 847–864. https://doi.org/10.1080/03075079.2024.2323593

Nunan, D. (1988). *Syllabus design*. Oxford University Press.

Oxford, R. (1990). *Language learning strategies: What every teacher should know*. Newbury House.

Oxford, R. (1995). Linking theories of learning with intelligent computer-assisted language learning (ICALL). In V. M. Holland, J. D. Kaplan, & M. R. Sams (Eds.), *Intelligent language tutors: Theory shaping technology* (pp. 359–369). Lawrence Erlbaum.

Pack, A., & Maloney, J. (2024). Using artificial intelligence in TESOL: Some ethical and pedagogical considerations. *Tesol Quarterly, 58*(2), 1007–1018. https://doi.org/10.1002/tesq.3320

Peachey, N., & Crichton, R. (2024). *AI activities and resources for English language teachers*. British Council. https://doi.org/10.57884/DKKR-QF54

Phakiti, A., & Leung, C. (2024). *Assessment for language teaching*. Cambridge University Press.

Polakova, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. *Cogent Education, 11*(1), Article 2410101. https://doi.org/10.1080/2331186X.2024.2410101

Rezai, A., Namaziandost, E., & Hwang, G.-J. (2024). How can ChatGPT open promising avenues for L2 development? A phenomenological study involving EFL university students in Iran. *Computers in Human Behavior Reports, 16*, Article 100510. https://doi.org/10.1016/j.chbr.2024.100510

Richards, J. C. (1990). *The language teaching matrix*. Cambridge University Press. https://doi.org/10.1017/CBO9780511667152

Richards, J. C. (2017). *Curriculum development in language teaching* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/9781009024556

Richards, J. C., & Rodgers, T. S. (2014). *Approaches and methods in language teaching* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/9781009024532

Shafiee Rad, H., & Roohani, A. (2025). AI language alchemists: Unleashing task-based chatbots to enhance speaking proficiency, shape attitudes, and foster a translanguaging space. *Journal of Educational Computing Research, 63*(7–8), 1659–1688. https://doi.org/10.1177/07356331251363281

Shi, H., Chai, C. S., Zhou, S., & Aubrey, S. (2025). Comparing the effects of ChatGPT and automated writing evaluation on students' writing and ideal L2 writing self. *Computer assisted Language learning*. Advance online publication. https://doi.org/10.1080/09588221.2025.2454541

Sockett, G. (2014). *The online informal learning of English*. Palgrave Macmillan. https://doi.org/10.1057/9781137414885
Tai, H.-Y., Lin, M.-F., & Chen, Y.-S. (2026). Incorporating ChatGPT into genre-based instruction for argumentative writing among EFL college students. *International Journal of Applied Linguistics, 36*(1), 767–781. https://doi.org/10.1111/ijal.12777
Tai, T.-Y., & Chen, H. H.-J. (2024). Improving elementary EFL speaking skills with generative AI chatbots: Exploring individual and paired interactions. *Computers & Education, 220*, Article 105112. https://doi.org/10.1016/j.compedu.2024.105112
The Douglas Fir Group. (2016). A transdisciplinary framework for SLA in a multilingual world. *The Modern Language Journal, 100*, 19–47. https://doi.org/10.1111/modl.12301
Tram, N. H. M., Nguyen, T. T., & Tran, C. D. (2024). ChatGPT as a tool for self-learning English among EFL learners: A multi-methods study. *System, 127*, Article 103528. https://doi.org/10.1016/j.system.2024.103528
Wang, Y., Zhang, T., Yao, L., & Seedhouse, P. (2025). A scoping review of empirical studies on generative artificial intelligence in language education. *Innovation in Language learning and teaching*. Advance online publication. https://doi.org/10.1080/17501229.2025.2509759
Westerman, D. A. (1991). Expert and novice teacher decision making. *Journal of Teacher Education, 42*(4), 292–305. https://doi.org/10.1177/002248719104200407
Wiboolyasarin, W., Wiboolyasarin, K., Tiranant, P., Jinowat, N., & Boonyakitanont, P. (2025). AI-driven chatbots in second language education: A systematic review of their efficacy and pedagogical implications. *Ampersand, 14*, Article 100224. https://doi.org/10.1016/j.amper.2025.100224
Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x
Wu, X.-Y., Radloff, J. D., Yeter, I. H., Wang, L., & Chiu, T. K. F. (2025). Designing artificial intelligence chatbots for self-regulated learning from a systematic review based on Habermas's three interests. *Interactive learning environments*. Advance online publication. https://doi.org/10.1080/10494820.2025.2563086
Yang, H., Kim, H., Lee, J. H., & Shin, D. (2022). Implementation of an AI chatbot as an English conversation partner in EFL speaking classes. *ReCALL, 34*(3), 327–343. https://doi.org/10.1017/S0958344022000039
Yeo, M. A., Moorhouse, B. L., & Wan, Y. (2025). From academic text to talk-show: Deepening engagement and understanding with google NotebookLM. *TESL-EJ, 28*(4), 1–14. https://doi.org/10.55593/ej.28112int
Yeung, S. (2025). University students' engagement with generative AI-supported automated writing evaluation (AWE) feedback. *Journal of Second Language Writing, 68*, Article 101203. https://doi.org/10.1016/j.jslw.2025.101203
Zhan, Y., & Yan, Z. (2025). Students' engagement with ChatGPT feedback: Implications for student feedback literacy in the context of generative artificial intelligence. *Assessment & Evaluation in Higher Education*, 1–14. https://doi.org/10.1080/02602938.2025.2471821
Zhang, Z., Aubrey, S., Huang, X., & Chiu, T. K. F. (2025). The role of generative AI and hybrid feedback in improving L2 writing skills: A comparative study. *Innovation in Language learning and teaching*. Advance online publication. https://doi.org/10.1080/17501229.2025.2503890
Zhang, Z., Zhou, S., & Bai, B. (2025). A systematic scoping review of self-regulated learning with AI in language education. *International Journal of Applied Linguistics*. https://doi.org/10.1111/ijal.70030. Advance online publication.
Zheng, S. (2024). The effects of chatbot use on foreign language reading anxiety and reading performance among Chinese secondary school students. *Computers and Education: Artificial Intelligence, 7*, Article 100271. https://doi.org/10.1016/j.caeai.2024.100271