ISSN: 2669-2341 10.33422/ijsfle.v5i1.1445

Beyond Memorisation: An AI-Mediated Course Design for Vocabulary Learning in Higher Education

Angel Yim Hung Fan

Lingnan University, Hong Kong

fanyh@connect.hku.hk, https://orcid.org/0000-0001-7349-6332

Cite this article as: Fan, A. Y. H. (2026). Beyond Memorisation: An Ai-Mediated Course Design for Vocabulary Learning in Higher Education. *International Journal of Second and Foreign Language Education, 5*(1), 48-85. https://doi.org/10.33422/ijsfle.v5i1.1445

Abstract

English language teaching (ELT) in higher education has undergone successive pedagogical transformations, evolving from rote memorisation to task-based and communicative approaches, integrating computer-assisted language learning (CALL), and now entering an era of artificial intelligence (AI)–mediated learning. This paper argues that AI adoption should not be viewed as a pedagogical rupture, but as a continuation of a long-standing shift toward learner-centred, meaning-focused, and interaction-driven language education. Situating AI within this trajectory, the paper contends that successful integration relies not on technological novelty, but on research-informed pedagogy, ethical awareness, and evidence-based practice. Echoing a longstanding caution against allowing education to be driven by the measurement of outcomes at the expense of purpose, it asks not what is technologically possible, but what is pedagogically meaningful. To illustrate this argument, the paper presents an action research project documenting the ongoing construction of a new elective course on vocabulary development at a research-focused liberal arts university in Hong Kong. Moving beyond rote memorisation, the course enhances vocabulary acquisition through active learning, textual analysis, and extensive reading, employing techniques such as morphological analysis and contextual inference. Drawing on literary and non-literary materials, it adopts a scaffolded design transitioning from guided instruction to independent learning, fostering critical thinking alongside lexical growth. Through discussions, presentations, and written analyzes understood as a recursive, problem-solving process, students apply new vocabulary in academic and real-world contexts—offering a practical model for pedagogically grounded, AI-mediated vocabulary learning.

Keywords: Action Research, English Language Teaching, Lexical Development, AI-mediated Language Classrooms, TESOL

1. Introduction

1.1. The Current Moment in English Language Teaching

The arrival of generative artificial intelligence in the language classroom has provoked reactions ranging from uncritical enthusiasm to anxious foreboding. Some commentators greet AI as a transformative force that will revolutionize how languages are taught and learned; others warn that it threatens academic integrity, deskills both teachers and students, and reduces the rich, social activity of learning to the transactional consumption of machine-generated

output. What unites these competing positions is a shared assumption that AI represents a fundamental break with everything that came before—a rupture that demands either celebration or resistance.

This paper proceeds from a different premise. It argues that the entry of AI into English language teaching (ELT) is best understood not as a question of *whether* the technology disrupts pedagogy, but of *how* it can be meaningfully integrated into established traditions of language education. When viewed against the long arc of ELT's development over the past century, AI does not appear as an alien intrusion but as the latest expression of a movement that has been underway for decades: a steady shift away from rote memorization and teacher-centred transmission toward reasoning, learner agency, and autonomy. The central question, therefore, is not how to defend against AI or how to deploy it for its own sake, but how to frame it pedagogically so that it serves rather than supplants the deeper purposes of education.

This framing matters because the field's recent history offers a sobering precedent. Zawacki-Richter, Marín, Bond, and Gouverneur (2019), in their systematic review of 146 studies on AI applications in higher education, found a "dramatic lack of critical reflection" on the pedagogical and ethical implications of these technologies, alongside a striking absence of educators themselves: only a small minority of first authors came from education departments, while computer science and STEM disciplines dominated. The result, they argued, is a body of research rich in technical sophistication but impoverished in theoretical and pedagogical grounding. If AI is to fulfil its promise in language education, this gap must be closed—and it can only be closed by educators who approach the technology from the standpoint of learning and teaching rather than computation.

1.2. Aim and Scope

This paper makes a dual contribution. First, it advances a conceptual argument that situates AI within the historical trajectory of ELT and within established theoretical traditions in learning science, contending that AI-mediated learning is a continuation rather than a rupture. Second, it grounds this argument in a concrete illustration: the ongoing construction of a new elective course on vocabulary development at a research-focused liberal arts university in Hong Kong. The course is presented as an action research project, documenting both the design rationale and the integration of AI within a deliberately pedagogically grounded framework.

The choice of vocabulary as the site of this illustration is purposeful. Vocabulary learning has long been associated with the most mechanical forms of instruction—word lists, flashcards, and rote memorization—and thus offers a particularly demanding test case. If AI can be integrated into vocabulary teaching in a way that deepens reasoning rather than automating recall, then the broader argument gains considerable force. Conversely, vocabulary is also a domain where the temptation to use AI as a shortcut is acute: a student can ask a chatbot to

define, translate, or even deploy a word with no cognitive effort whatsoever. The course described here confronts this tension directly, positioning AI as a scaffold that surfaces options while leaving the evaluative and generative work with the learner.

1.3. Structure of the Paper

The paper is organized as follows. Section 2 reviews the historical trajectory of ELT across its major pedagogical phases—from the grammar-translation era, through the communicative turn and task-based learning, to the CALL era and the rise of learner autonomy—and identifies the through-line that connects them. Section 3 develops the theoretical framework, drawing on constructivism, self-determination theory, the knowledge-telling/knowledge-transforming distinction, and theories of scaffolding and cognitive load to argue that AI represents continuity rather than rupture. Section 4 establishes vocabulary as a productive test case, reframing lexical learning as a matter of depth and expansion rather than mere acquisition. Section 5 sets out the methodology, framing the project as the design-oriented first phase of an action research cycle and specifying its setting, participants, and anticipated data sources. Section 6 presents the course design in detail, mapping its learning outcomes, content, and blended, reading-into-writing architecture onto the theoretical framework. Section 7 turns to AI as scaffold in practice, offering worked examples of student–AI interaction and showing how the course cultivates critical AI literacy. Section 8 details the assessment regime and supporting materials. Section 9 confronts the anticipated challenges and ethical safeguards the design must address, from academic integrity and over-reliance to privacy, bias, and the irreducible role of the teacher. Section 10 situates the project within the wider field, discussing its implications for technology integration, educational purpose, and the state of AI research in higher education. Section 11 concludes.

2. Literature Review - The Historical Trajectory of English Language Teaching

To understand why AI should be seen as continuity rather than rupture, it is necessary to recall how ELT has evolved. The field's history is not a random succession of fashions but a discernible movement, propelled by shifting pedagogical priorities, from passive reception toward active engagement and agency. Tracing this movement reveals the trajectory into which AI naturally fits.

2.1. The Grammar-Translation Era

For much of the nineteenth and early twentieth centuries, the dominant approach to language teaching was the grammar-translation method. As Richards and Rodgers (2014) describe in their authoritative survey of language teaching approaches, this tradition centred on the explicit study of grammatical rules, the translation of texts between the target language and the learner's

first language, and the memorization of vocabulary in isolated, decontextualized lists. Linguistic accuracy was paramount; the goal was the correct reproduction of forms rather than communicative fluency. The teacher occupied a position of unquestioned authority, the repository of knowledge to be transmitted to comparatively passive students.

Within this paradigm, learning a language was fundamentally an exercise in recall. Vocabulary acquisition meant committing word-equivalents to memory; grammatical competence meant mastering rules and applying them precisely in translation exercises. The model treated language as an object to be analyzed and dissected rather than a medium to be used. While the grammar-translation method produced learners capable of reading classical texts and parsing complex sentences, it was notoriously poor at developing the ability to communicate in real time and in authentic situations.

2.2. The Communicative Turn

By the latter half of the twentieth century, dissatisfaction with the grammar-translation method and its successors gave rise to what is now known as the communicative turn. Littlewood (1981), in his foundational introduction to communicative language teaching, articulated the central insight of this movement: that the purpose of language learning is communication, and that pedagogy should therefore prioritize meaning-focused interaction and authentic language use over the mechanical mastery of forms. Language was no longer to be treated merely as an object for analysis but as a tool for genuine communication, deployed in contexts that approximated real-world use.

This represented a profound reorientation. Attention shifted from the accurate reproduction of isolated structures to the negotiation of meaning in interaction. Errors came to be seen not as failures to be eradicated but as natural features of an interlanguage in development. The learner moved from the periphery toward the centre of the pedagogical process, and the classroom was reconceived as a space for communicative activity rather than rote drilling. The communicative turn thus marked the first decisive step in the long movement from passive reception toward active engagement.

2.3. Task-Based Learning

The communicative turn found a more systematic and theoretically elaborated expression in task-based language teaching. Ellis (2003), in his comprehensive treatment of task-based learning and teaching, explains how purposeful communicative tasks—activities in which learners use language to achieve a genuine outcome—became the central organizing principle of instruction. Rather than practicing language forms for their own sake, learners engage in tasks that require them to use the target language to solve problems, exchange information, or accomplish goals, with the negotiation of meaning at the heart of the process.

Task-based learning deepened the commitment to authenticity and learner engagement that the communicative turn had inaugurated. It positioned the learner as an active agent who learns language by using it, drawing on the cognitive and interactional processes that purposeful communication demands. The teacher's role shifted further still, from authority and transmitter toward facilitator and designer of meaningful tasks. In this sense, task-based learning continued and consolidated the trajectory away from passive reception.

2.4. The CALL Era and the Rise of Learner Autonomy

The closing decades of the twentieth century and the opening years of the twenty-first introduced a new dimension: the integration of digital technologies into language learning. Levy (1997), in his early conceptualization of computer-assisted language learning (CALL), examined how computers could expand opportunities for practice, provide individualized feedback, and grant learners access to authentic materials beyond the confines of the classroom. CALL was not merely a matter of digitizing existing practices; it carried with it the potential to reshape the relationship between learner, teacher, and content.

Crucially, the CALL era coincided with, and helped to advance, a growing emphasis on learner autonomy and self-direction. Benson (2011), in his definitive account of autonomy in language learning, traces how the proliferation of digital tools enabled and encouraged learners to take greater responsibility for their own learning processes—setting goals, selecting resources, monitoring progress, and directing their own development. Technology, in this framing, was valuable not as an end in itself but as a means of fostering the agency and self-regulation that autonomous learning requires.

2.5. Identifying the Trajectory

Viewed in sequence, these phases reveal a consistent and cumulative movement. The grammar-translation era treated the learner as a passive recipient of transmitted knowledge, the communicative turn repositioned the learner as an active participant in meaning-making, task-based learning consolidated this by organizing instruction around purposeful use, and the CALL era extended the movement by fostering autonomy and self-direction. Across these shifts runs a clear trajectory: from passive reception toward engagement, from teacher authority toward learner agency, from the reproduction of forms toward reasoning and meaningful use.

It is into this trajectory that AI enters. The question is not whether AI is powerful—it manifestly is—but how it is pedagogically framed. As with every previous technology, from the language laboratory to the personal computer, what matters is integration rather than adoption for its own sake. AI fits the trajectory of ELT precisely to the extent that it is deployed to deepen engagement, agency, and reasoning rather than to revert to the mechanical transmission and recall that the field has spent a century moving beyond. This is the sense in which AI represents

continuity rather than rupture—a claim that the next section grounds in established theoretical traditions.

3. Theoretical Framework: AI as Continuation, Not Rupture

If the historical argument establishes that AI fits the trajectory of ELT, the theoretical argument explains *why* a pedagogically sound integration of AI is consistent with established understandings of how learning works. This section draws on four interlocking theoretical traditions—constructivism, self-determination theory, the knowledge-telling/knowledge-transforming distinction, and theories of scaffolding and cognitive load—to demonstrate that AI, properly framed, supports rather than subverts the deep learning these traditions describe.

3.1. Constructivist Foundations

The first foundation is constructivism. In contrast to transmission models that conceive of learning as the transfer of knowledge from teacher to student, constructivism holds that knowledge is actively constructed by the learner. Jonassen (1999), in his influential account of constructivist learning environments, argues that meaningful learning occurs when learners engage in authentic, complex tasks that require them to build understanding through active intellectual work rather than passive absorption. Knowledge, in this view, is not delivered but made.

This principle has direct implications for AI. If learning is a process of construction, then the pedagogical value of any AI tool depends entirely on whether it supports or short-circuits that process. An AI system that simply supplies answers—that constructs knowledge *on behalf of* the learner—undermines the very mechanism by which learning occurs. But an AI tool thoughtfully designed to prompt, to surface options, to provoke reflection, and to occasion the learner's own intellectual work can support the construction process rather than replace it. The constructivist lens thus reframes the question of AI from "what can the machine do?" to "what cognitive work does the tool leave with the learner?"—a reframing that runs throughout the course design presented below.

3.2. Self-Determination Theory

The second foundation is self-determination theory. Deci and Ryan (2000), in their elaboration of the psychological needs that underpin human motivation, identify autonomy, competence, and relatedness as fundamental drivers of deep, self-sustaining engagement. When learners experience a sense of control over their learning (autonomy), a growing sense of capability (competence), and meaningful connection (relatedness), they are far more likely to engage deeply and persist.

Self-determination theory adds a motivational dimension to the case for AI as continuity. The historical trajectory toward learner autonomy described in Section 2 is, in this light, not merely a pedagogical preference but a response to a fundamental psychological need. If AI is deployed in ways that enhance learners' sense of control and capability—by giving them tools to pursue their own questions, to test their own hypotheses about language, and to monitor their own progress—then it aligns with the principles of self-determination and supports deep learning. If, conversely, AI is deployed in ways that diminish autonomy and competence—by making decisions for learners and removing the opportunity to develop capability—then it works against these principles, however impressive its technical performance. The design challenge, therefore, is to position AI so that it augments rather than erodes learner agency.

3.3. Knowledge-Telling versus Knowledge-Transforming

The third foundation comes from research on writing and composition. It may seem incongruous to ground an account of so recent a technology in scholarship from the 1970s and 1980s, but the apparent anachronism is deliberate and is itself part of the argument. The claim advanced throughout this framework is one of continuity rather than rupture, and the choice of foundational sources enacts that claim: the cognitive questions raised by AI in writing are not new, and the demonstration that decades-old theories of composition map cleanly onto them is the strongest available evidence that they are not. These are not, in any case, theories of technology; they are accounts of how human cognition operates during writing and learning, pitched at a level of abstraction that outlasts any particular tool. The architecture of working memory did not change in 2022. What changed is that a tool arrived which can either tax or relieve that architecture depending on how it is used—and it is precisely an enduring theory of learning, rather than one improvised to explain the tool itself, that gives us independent purchase to judge whether the tool helps or harms.

On that foundation, Bereiter and Scardamalia (1987), in their seminal psychology of written composition, distinguish between two fundamentally different modes of cognitive engagement: knowledge-telling and knowledge-transforming. Knowledge-telling involves the relatively straightforward retrieval and reproduction of existing information—writing down what one already knows, with the text generated by little more than topic and genre cues. Knowledge-transforming, by contrast, requires the writer to reorganize, evaluate, and improve ideas, using the demands of the task to reshape understanding itself. The former is reproductive; the latter is generative and developmental. This distinction extends the earlier insight of Flower and Hayes (1981), whose cognitive process model recast writing as a recursive act of problem-solving in which planning, translating, and reviewing continually feed back into one another; what Bereiter and Scardamalia (1987) add is an account of why some writers exploit this recursiveness to transform their thinking while others do not. The transforming process is cognitively demanding precisely because it asks the writer to hold a content problem and a

rhetorical problem in tension at once—a demand that aligns closely with the account of cognitive load developed by Sweller (1988), who shows that genuine learning depends on directing finite working-memory resources toward productive, schema-building effort rather than toward extraneous or merely reproductive activity.

This distinction is directly relevant to the integration of AI—and that it should be so relevant, decades after it was first drawn, is exactly the point. Used uncritically, AI invites knowledge-telling at its most degraded: the learner becomes a mere conduit for machine-generated content, reproducing without transforming. The danger is not simply that the writing is unoriginal but that the cognitive work which makes writing a vehicle for learning is offloaded entirely onto the machine—the very effort that Sweller (1988) identifies as the engine of schema formation is bypassed, and the demanding dual problem-space described by Bereiter and Scardamalia (1987) collapses into a single act of transcription. But AI can equally be enlisted in the service of knowledge-transforming—provided that students are required to critique, revise, and justify rather than copy.

When learners treat AI output as raw material to be interrogated, evaluated, and reworked, the technology becomes an occasion for precisely the reorganization and improvement of ideas that knowledge-transforming entails; the model's draft becomes a problem to be solved rather than a solution to be accepted, reintroducing the recursive review and revision that Flower and Hayes (1981) place at the centre of skilled composition. Understood this way, AI can function as a form of scaffolding in the Vygotskian sense elaborated by Wood, Bruner, and Ross (1976), supporting the learner in attempting cognitive operations that lie just beyond independent reach—provided the scaffold is designed to be progressively withdrawn rather than permanently leaned upon. It can likewise serve the autonomy that Deci and Ryan (1985) identify as central to durable motivation, since the act of evaluating and overruling a machine's suggestions returns evaluative authority to the learner rather than ceding it.

The decisive variable, once again, is pedagogical framing. The same tool can occasion the most degraded knowledge-telling or the most productive knowledge-transforming, and which of these it becomes is determined not by the technology but by the task in which it is embedded. That a forty-year-old distinction should so precisely capture this choice confirms the framework's guiding premise: AI does not pose a new pedagogical problem so much as sharpen an old one. AI-mediated learning, properly designed, should deepen reasoning and informed decision-making, not automate thinking.

3.4. Scaffolding and Cognitive Load

The fourth foundation is the closely related pair of scaffolding theory and cognitive load theory. The concept of scaffolding derives from Vygotsky's (1978) notion of the Zone of Proximal Development—the gap between what a learner can accomplish independently and what they

can accomplish with appropriate support. Learning, in this account, occurs precisely within this zone, when support enables performance beyond the learner's current independent ability. Wood, Bruner, and Ross (1976), in their foundational study of tutoring, gave this idea its enduring name and specified its defining feature: scaffolding is *temporary* support, calibrated to the learner's needs and gradually withdrawn as competence increases. The goal of the scaffold is its own eventual removal.

Cognitive load theory provides a complementary mechanism. Sweller (1988) demonstrated that learning is impaired when the demands placed on working memory exceed its limited capacity, and that effective instruction manages cognitive load so that learners can devote their mental resources to the schema-building that constitutes genuine learning. Scaffolding, in these terms, works in part by managing cognitive load—by handling peripheral demands so that the learner can concentrate on the core intellectual task.

Together, these theories furnish the most important principle for the pedagogical positioning of AI. If AI functions as an *authority*—a source of answers to be accepted rather than interrogated—then learning risks collapsing into dependency, the scaffold becoming a permanent crutch rather than a temporary support. But if AI functions as a *scaffold*—temporary support that manages cognitive load while preserving the learner's reasoning, and that is progressively withdrawn as competence grows—then it serves the very developmental process these theories describe. The critical requirement is that students must interrogate, verify, and justify AI suggestions; the evaluative authority must remain with the learner. This principle is the keystone of the course design.

3.5. Synthesis

These four traditions converge on a single conclusion. Constructivism insists that knowledge must be actively built by the learner; self-determination theory insists that autonomy and competence drive deep engagement; the knowledge-telling/knowledge-transforming distinction insists that genuine learning requires the reorganization rather than the reproduction of ideas; and scaffolding and cognitive load theory insist that support must be temporary, calibrated, and ultimately withdrawn. In every case, the value of AI depends not on its technical capabilities but on whether it is framed to support these processes or to circumvent them. AI-mediated learning, properly designed, deepens reasoning and informed decision-making rather than automating thinking. Pedagogical framing—integration rather than adoption for its own sake—is decisive. This is the theoretical sense in which AI is continuation, not rupture.

4. Vocabulary as a Test Case for AI Integration

Having established both the historical and theoretical grounds for treating AI as continuity, the paper now turns to the specific domain in which these arguments are tested: vocabulary

learning. Vocabulary is chosen deliberately, both because it has historically been associated with the most mechanical forms of instruction and because it offers a particularly clear illustration of the difference between rote retention and informed lexical decision-making.

4.1. Why Vocabulary?

Vocabulary occupies a peculiar position in language education. On the one hand, it is indispensable: without words, there is no communication, and lexical knowledge is among the strongest predictors of overall language proficiency. On the other hand, vocabulary teaching has long been the domain most resistant to the pedagogical advances chronicled in Section 2. While grammar instruction was transformed by the communicative turn and skills like speaking and writing were reconceived around authentic use, vocabulary teaching in many contexts continued to rely on word lists, flashcards, and rote memorization—the very practices the field had otherwise moved beyond.

This makes vocabulary an exacting test case. If AI can be integrated into vocabulary teaching in a way that advances reasoning and agency, the broader argument that AI is continuity rather than rupture gains substantial credibility. And because vocabulary is precisely the domain where AI most readily tempts learners toward effortless shortcuts—a definition, a synonym, a sentence supplied on demand—it is also the domain where the principle of AI-as-scaffold-not-authority is most severely tested.

4.2. Depth Over Size

A crucial insight from vocabulary research is that lexical knowledge is not merely a matter of size—the sheer number of words a learner knows—but of depth. Nation (2001), in his comprehensive treatment of vocabulary learning, distinguishes the many dimensions of knowing a word: not only its form and basic meaning, but its collocations (the words it characteristically combines with), its register (the contexts and levels of formality in which it is appropriate), its connotations, and its grammatical behaviour. To know a word deeply is to be able to use it flexibly and appropriately across contexts, not merely to recognise or translate it.

Schmitt (2008), in his synthesis of research on instructed second language vocabulary learning, similarly emphasises that meaningful vocabulary use depends on this depth of knowledge and that instruction should therefore aim well beyond the memorisation of word-meaning pairs. The implication is significant: a pedagogy that treats vocabulary learning as the accumulation of definitions captures only the most superficial layer of what it means to know a word. Genuine lexical competence requires engagement with collocation, register, and contextual flexibility—precisely the dimensions that demand reasoning rather than recall.

4.3. Repetition and Meaningful Exposure

Depth of knowledge does not emerge from a single encounter. Webb (2007), in his study of the effects of repetition on vocabulary knowledge, demonstrates that multiple encounters with a word across varied contexts substantially strengthen retention and enrich the learner's understanding of its range of meanings and uses. It is not mere repetition that matters, but repetition that is meaningful—encounters in which the word is met in genuine contexts that reveal new facets of its behaviour.

This finding has direct implications for course design. If durable, deep vocabulary knowledge requires multiple meaningful encounters across contexts, then a course must be structured to provide such encounters systematically, through extensive reading, discussion, and writing in which target vocabulary recurs in varied and authentic settings. It also clarifies the limitation of traditional instruction: memorisation from lists provides repetition without the contextual variety that builds depth, and thus often fails to produce durable transfer to real use.

4.4. From Acquisition to Expansion

These insights suggest a fundamental reframing of vocabulary pedagogy: a shift from *acquisition* to *expansion*. Rather than asking students simply to acquire new words by committing definitions to memory, instruction can ask them to expand their lexical competence by justifying lexical choices, analysing collocations, evaluating register, and discriminating among near-synonyms. This reframing moves vocabulary learning from rote retention toward informed lexical decision-making—from knowledge-telling toward knowledge-transforming, in Bereiter and Scardamalia's (1987) terms.

It is within this reframing that AI finds its proper role. In a pedagogy of expansion, AI can surface options—generating candidate synonyms, illustrating collocations, suggesting alternatives—but the cognitive work of evaluating, selecting, and justifying those options remains with the learner. AI becomes a tool for widening the field of possibilities the learner must reason about, not a substitute for the reasoning itself. This is the principle that animates the course design presented in the following section.

4.5. Tools, Prompts, and Privacy-Compliant Configuration

The principle articulated above is only as sound as the tools through which it is enacted. Two are used in the course, chosen precisely because they externalise lexical reasoning rather than shortcut it: an institutionally hosted generative-AI service and a large reference corpus. Their selection reflects the same pedagogical logic that governs the rest of the design—each tool is valuable insofar as it presents evidence the learner must interpret, not answers the learner may simply adopt.

The generative-AI tool is **LingnanChatGPT**, the university's in-house conversational service (ITSC, n.d.). The decision to use an institutionally hosted rather than a public deployment is deliberate and has two rationales, one ethical and one pedagogical. Ethically, an in-house service keeps interaction data within the institution's governance framework rather than routing student input to a third-party commercial endpoint, which matters both for compliance with the university's data-protection obligations and for the transparency owed to students. Data are accessed only with students' informed consent, obtained at enrolment, and students retain the right to decline data collection without penalty. Pedagogically, the same configuration yields a rich record of the *process* of lexical decision-making: with consent, the service preserves the chat history, including the prompts students issue, the ways they subsequently modify those prompts, and the intervals between exchanges. This process record is itself an object of instruction and assessment—it makes visible whether a student is interrogating and revising the model's output (knowledge-transforming) or accepting it wholesale (knowledge-telling), a distinction that a finished text alone conceals.

Prompts are designed to elicit options for evaluation rather than verdicts to be trusted. Students are taught to frame requests so that the model surfaces alternatives and the reasoning behind them, while the act of judgement remains with the learner. Representative prompts include: *"List six adjectives that could describe 'success' in academic writing, and for each, note the kind of context it typically appears in—I will decide which fits my sentence"*; *"Here is a sentence from my draft. Suggest three alternative verbs and explain the connotation of each; do not tell me which is best"*; and *"I think 'put a finger on' may be too informal for a research abstract. Give me the arguments for and against, and two more formal candidates, so I can judge."* In each case the prompt is constructed to keep the model in the role of an option-generator whose suggestions the student must then verify—crucially, against corpus evidence rather than against the model's own authority.

That verification is the task of the corpus tool: the **Corpus of Contemporary American English (COCA)** (Davies, 2008), a balanced reference corpus of contemporary American English—originally some 450 million words and now exceeding one billion—drawn from spoken, fiction, magazine, newspaper, and academic sources. COCA is free to use following a simple registration, which keeps it accessible to students beyond the life of the course. Two features make it especially suited to a pedagogy of expansion. First, its section-based architecture allows register and genre to be examined empirically, so that a phrase's suitability for academic prose becomes a matter of observable distribution rather than intuition. Second, every search generates its own persistent URL, which means a teacher can send a student the exact query and its results to explore, and a student can submit a query link as evidence for a lexical decision. This second feature underpins a distinctive assessment move borrowed from the "kibbitzer" tradition: rather than correcting a student's draft, the teacher can leave a corpus link beside a questionable phrase, turning a correction into an invitation to investigate.

Worked examples of both tools in use are provided as appendices. **Appendix A** presents a fully worked demonstration built from a single authentic case—a Master's student's draft abstract on which the teacher left corpus links instead of corrections—in which five real phrases (*high criticism*, *put a finger on*, *reason of*, *gives the impression to be*, and *heavy success*) are diagnosed through COCA queries, each illustrating a different lexical relationship: collocation, register/distribution, colligation, complementation, and, in the final self-directed item, collocation once more. **Appendix B** contains the two output activities that follow from this demonstration: Activity 2, in which students replay the four queries, state the rule they discover in the form *"The data show that…"*, and then formulate their own adjective-plus-noun query; and Activity 3, the module's flagship task, in which students close-read a published systematic review, corpus-check three of its content words across the collocation, colligation, and register query types, and complete a mini-thesis sentence linking their finding to the nature of natural academic English. Together these appendices instantiate the principle set out in 4.4: the tools widen the field of lexical possibilities, and the corpus supplies the evidence against which the learner—not the machine—reaches and defends a decision.

4.6. Assessment of the Output Tasks

Each of the module's three assessed outputs—the AI-critique log (Activity 1), the lexical justification annotations (Activity 3), and the summative word presentation—is supported by a dedicated criterion-referenced rubric. The full instruments appear in **Appendix C**; this section explains their design and justifies the decisions behind them.

The three rubrics share a single five-band architecture. Each criterion is scored across five levels descending from full mastery to non-performance, and, where departmental convention applies, these bands are anchored to approximate IELTS levels (8 down to 4). This anchoring is deliberate rather than cosmetic. The word presentation is an existing summative assessment in LUE3008 *English Vocabulary Across Curriculum*, whose established rubric already uses IELTS-anchored bands with fixed weightings of 50% for content, 30% for delivery, and 10% each for organisation/visual aids and peer review. Rather than impose a competing scoring logic, the module retains that instrument without modification and constructs the two new rubrics in the same idiom, so that scores generated by all three tasks remain mutually interpretable and legible to colleagues teaching elsewhere on the programme. Preserving the presentation rubric intact also maintains alignment with departmental practice and minimises the retraining burden on tutors who already use it.

The two new instruments—the AI-critique log rubric and the lexical justification annotation rubric—are designed to render Pillar One's central constructs assessable. Where a generic writing rubric would reward correctness in the abstract, these rubrics reward the specific behaviours the module seeks to cultivate: naming the lexical relationship at stake (collocation,

colligation, register/distribution, or complementation), verifying claims against corpus evidence rather than intuition, and articulating what the data reveal. In the AI-critique log rubric, these behaviours are operationalized through four criteria—detection of lexical issues (30%), corpus verification (30%), quality of revision (25%), and reflection (15%)—whose weightings place the evidential heart of the task, detecting a problem and proving it against the corpus, above the downstream revision. This ordering is pedagogically motivated: since the module holds that natural academic English follows from diagnosis grounded in data, the assessment must not allow a fluent-sounding revision to compensate for an ungrounded one. The annotation rubric applies the same principle, foregrounding accuracy of identification (35%) and corpus-supported justification (30%) while reserving a smaller share for the metalinguistic explanation and the concluding mini-thesis, so that naturalness cannot be credited on assertion alone.

Two design features warrant explicit justification. First, both new rubrics treat corpus verification as an independent, heavily weighted criterion rather than folding it into a general "accuracy" band. This guards against the failure mode the module is built to counter—students, and AI systems alike, producing plausible prose that no data support—and gives tutors a clear warrant for distinguishing a well-evidenced answer from a merely confident one. Second, the rubrics are criterion-referenced rather than norm-referenced, an approach suited to a small elective whose aim is mastery of a transferable diagnostic method rather than the ranking of students against one another.

Two points bear on implementation. The weightings suggested for the AI-critique log and the annotation task are provisional, since the presentation already accounts for 30% of the course mark; the final assessment scheme will need to specify how the three instruments combine into a coherent whole. In addition, the presentation rubric refers to a separate ten-point peer-evaluation scale, which the department maintains as a standalone instrument and which is therefore not reproduced here.

5. Methodology: An Action Research Design

5.1. Rationale for Action Research

The development of the course is framed not as an act of curriculum design alone but as a piece of practitioner inquiry, and the methodological tradition that most precisely fits this orientation is action research. Action research is distinguished from other forms of educational inquiry by its commitment to the improvement of practice from within, by the practitioner, in the practitioner's own setting. Lewin (1946), to whom the term is conventionally traced, conceived of action research as a spiral of planning, action, observation, and reflection through which practitioners progressively refine their interventions in light of evidence gathered in the course

of acting. Carr and Kemmis (1986), in their influential elaboration of the tradition, emphasise its emancipatory and self-critical character: action research is not merely problem-solving but a systematic, reflexive examination of one's own practice with a view to making it more rational, more just, and more educationally defensible.

This orientation is particularly well suited to the present project for three reasons. First, the integration of AI into vocabulary teaching is a genuinely open problem to which no settled pedagogical answer yet exists; it therefore calls for an exploratory, iterative methodology rather than the testing of a predetermined hypothesis. Second, action research foregrounds precisely the educator perspective whose absence Zawacki-Richter et al. (2019) lamented in the wider AIEd literature, positioning the teacher as the agent of inquiry rather than as the object of study or the implementer of others' designs. Third, the tradition's insistence on the unity of action and reflection mirrors the course's own pedagogical commitment to reflective, self-regulating learning, lending the project a methodological coherence in which the manner of inquiry echoes the substance of what is taught.

5.2. The Action Research Cycle

The project adopts the cyclical model elaborated by Kemmis, McTaggart, and Nixon (2014), in which each cycle comprises four interrelated moments: planning a change, acting and observing the process and consequences of the change, reflecting on these processes and consequences, and then re-planning in light of that reflection. Burns (2010), in her practical guide to action research specifically for English language teaching, stresses that these moments are not rigidly sequential but recursive and overlapping, and that the value of the model lies less in procedural fidelity than in its disciplining of the practitioner's attention toward systematic observation and honest reflection. The present paper documents the first, design-oriented phase of this larger cycle: the planning that precedes implementation. Subsequent cycles, to be reported in future work, will encompass the enactment of the course, the gathering of evidence during teaching, and the reflective revision of the design for its second iteration.

The decision to report the planning phase as a discrete contribution is itself methodologically motivated. The design rationale of an educational intervention is frequently rendered invisible in published accounts, which tend to report outcomes while leaving the reasoning that produced the intervention tacit. By making the planning phase explicit and theoretically grounded, the paper renders the design accountable to scrutiny and provides a baseline against which the lessons of subsequent cycles can be measured. This is consistent with Norton's (2009) argument that the credibility of pedagogical action research depends heavily on the transparency with which the practitioner articulates the assumptions and intentions underpinning the intervention.

5.3. Setting, Participants, and the Researcher's Role

The setting is an elective course at a research-focused liberal arts university in Hong Kong, where students are typically multilingual users of English studying it as an additional language at an advanced level. The participants in subsequent cycles will be the students who enrol in the elective, together with the teacher-researcher who designed and will teach it. The dual role of teacher and researcher is a defining and well-recognised feature of action research, and it carries both advantages and risks. Its advantage is intimate, contextualised knowledge of the learning environment; its risk is the threat to objectivity that arises when the designer of an intervention also evaluates it. Anderson and Herr (2015) address this tension through the concept of validity criteria specific to practitioner research, among them democratic validity, which concerns the extent to which inquiry is conducted collaboratively with stakeholders, and catalytic validity, which concerns the degree to which the inquiry energises participants toward transformation. The course design attends to these criteria by embedding student reflection and feedback as data sources, thereby positioning students not merely as subjects of the research but as participants in the reflective cycle.

5.4. Anticipated Data Sources and Analysis

Although the present phase precedes data collection, the design specifies in advance the evidence that subsequent cycles will gather, in keeping with the planning discipline of action research. These sources are deliberately multiple, so that triangulation can guard against the partiality of any single instrument. They will include the students' reflective writing, in particular the five-hundred-word reflection embedded in the final paper, which offers direct insight into learners' developing metacognition; documentary records of student–AI interaction, in which students log the AI suggestions they receive and the decisions they make in response; pre- and post-course measures of vocabulary depth, designed to capture not merely the size of students' lexicons but their command of collocation, register, and discrimination among near-synonyms; the teacher-researcher's own reflective journal, maintained throughout the teaching cycle in the manner Burns (2010) recommends; and samples of student writing across drafts, which permit the tracing of lexical decision-making over the recursive cycles of composition. Analysis will be predominantly qualitative and interpretive, employing thematic analysis of reflective and documentary data, supplemented by descriptive analysis of the vocabulary measures. The aim is not statistical generalisation—an aspiration foreign to the action research tradition—but the generation of contextually rich, transferable insight into how AI can be integrated into vocabulary pedagogy in a manner faithful to the principles set out above.

6. Course Design: An AI-Mediated Vocabulary Elective

The preceding sections have established the historical, theoretical, and disciplinary grounds for treating AI as a continuation of established pedagogical traditions. This section translates those grounds into a concrete course design: a new elective on vocabulary development currently under construction at a research-focused liberal arts university in Hong Kong. The course is presented as an action research project, and what follows documents its rationale, learning outcomes, content, pedagogical architecture, and methods.

6.1. Context and Rationale

The course is being developed within a liberal arts institution that places a high value on critical thinking, close reading, and the cultivation of articulate, independent learners. This institutional context is significant: it means the course can be designed not merely to expand students' word stocks but to develop their analytical and communicative capacities through engagement with vocabulary. The course thus aligns vocabulary development with the broader liberal arts mission, treating lexical growth as inseparable from critical thinking and meaningful communication.

The rationale for the course rests on the gap identified in Section 4 between traditional, memorization-based vocabulary instruction and the kind of deep, transferable lexical knowledge that research shows is required for genuine competence. The course is conceived as a pedagogically grounded, educator-led response to this gap—and, in its integration of AI, as a deliberate attempt to address the absence of educator perspectives that Zawacki-Richter et al. (2019) identified in the AIEd literature. Rather than adopting AI because it is available or fashionable, the course design begins from pedagogical principles and asks how AI can be enlisted in their service.

6.2. Learning Outcomes

The course is built around six learning outcomes that together express a progression from foundational strategy use to autonomous, reflective practice. Upon successful completion, students are expected to be able to: first, effectively utilize strategies for vocabulary acquisition, such as context clues and word analysis, in varied contexts, demonstrating an ability to apply these principles in English writing; second, identify and articulate the subtle differences in meaning and usage of similar vocabulary terms, thereby avoiding common usage errors in writing and speech; third, proficiently incorporate newly learned vocabulary into discussions and presentations to enhance communication skills; fourth, analyze and interpret various texts, demonstrating the ability to discuss the content, mood, and style with appropriate vocabulary; fifth, apply appropriate vocabulary techniques tailored to different genres, such as academic, literary, and technical, when engaging with texts and during discussions; and sixth, engage in

reflective practices to assess their vocabulary growth and usage, setting personal goals for improvement.

This progression is purposeful. The first two outcomes establish the foundational capacities of strategy use and lexical discrimination, reflecting the depth-over-size principle drawn from Nation (2001) and Schmitt (2008). The third and fourth move toward communicative and analytical application, embodying the shift from acquisition to expansion. The fifth introduces genre-sensitivity, recognizing that appropriate lexical choice is context-dependent. The sixth and final outcome foregrounds reflection and self-directed goal-setting, embodying the commitment to learner autonomy that runs through the historical trajectory and that self-determination theory (Deci & Ryan, 2000) identifies as a driver of deep engagement.

6.3. Indicative Content and Pedagogical Architecture

The course content is organized to operationalize the theoretical framework. It begins with structured vocabulary expansion, helping students build their lexical resources through morphological analysis, the study of synonyms and antonyms, and the exploration of word families. This explicit strategy instruction equips students with tools they can apply independently, consistent with the autonomy-fostering aim of the course.

The content then moves to contextual and textual analysis, in which students engage with a variety of texts to infer word meanings through context clues and to explore connotation, denotation, and the subtle nuances that distinguish near-synonyms. This component directly enacts the depth-over-size principle and provides the multiple meaningful encounters across contexts that Webb (2007) identifies as essential to durable retention.

Practical application follows, integrating vocabulary into discussions, presentations, and textual analysis tasks. Here the emphasis falls on use rather than recognition, embodying the communicative and task-based commitments inherited from Littlewood (1981) and Ellis (2003). The course then advances toward more sophisticated work on lexical choice through text analysis and the synthesis of knowledge into well-structured responses—the point at which vocabulary learning becomes most fully a matter of knowledge-transforming in Bereiter and Scardamalia's (1987) sense, as students reorganize and evaluate their lexical resources in the service of articulate expression.

6.4. The Blended Learning Framework

The course adopts a blended learning framework that integrates face-to-face instruction with structured digital engagement. Graham (2006), in his foundational definition of blended learning, emphasizes that the value of blending lies in the *purposeful combination* of modalities rather than the simple addition of technology to existing practice. This principle is central to

the course: digital and AI-supported components are not bolted on but designed to complement and extend in-class work.

Mobile and AI-supported tools play a particular role in extending practice beyond classroom time. Godwin-Jones (2018), in his discussion of mobile technology for language learning, highlights how such tools can promote sustained engagement by enabling learners to practice and explore language in the interstices of daily life. In the course, AI-supported vocabulary refinement allows students to continue working with target vocabulary outside class, encountering it in the varied contexts that build depth.

Yet the course also takes seriously the caution that technology alone guarantees nothing. Stockwell (2013), examining the relationship between technology and motivation in language learning, stresses that it is pedagogical design, not the technology itself, that determines effectiveness. Accordingly, the AI-supported components of the course are tightly integrated with in-class critique: students document the feedback they receive from AI tools and bring it into the classroom, where it is treated as provisional input to be examined rather than as authoritative output to be accepted. This pairing of out-of-class AI engagement with in-class critical evaluation is the mechanism by which the course keeps the evaluative authority with the learner.

6.5. The Reading-into-Writing Cycle

At the heart of the course lies a reading-into-writing cycle that links the analysis of texts to the production of writing, with vocabulary development as the connecting thread. The design of this cycle draws on writing research that conceives of writing not as the linear transcription of pre-formed thought but as a recursive, problem-solving process. Flower and Hayes (1981), in their cognitive process theory of writing, describe writing as an ongoing cycle of planning, drafting, and revising, in which writers continually return to and rework their developing texts. Writing, in this account, is itself a mode of thinking—a process through which ideas are generated and transformed.

In second language contexts, this recursive process is further complicated by the need to attend to discourse conventions and discipline-specific language. Hyland (2019), in his account of second language writing, emphasizes that L2 writers must navigate not only the cognitive demands of composition but also the genre expectations and lexical conventions of their target discourse communities. This makes vocabulary an especially salient concern in L2 writing, since appropriate lexical choice is integral to meeting genre expectations.

Vocabulary development, in turn, is strengthened precisely by the kind of iterative engagement that the writing process entails. Nation (2013), in the second edition of his foundational work, reaffirms the importance of encountering and using words repeatedly across contexts and across cycles of engagement. The reading-into-writing cycle thus serves a dual purpose: it

develops writing as a recursive process while simultaneously providing the repeated, meaningful encounters with vocabulary that deep lexical learning requires.

Concretely, the cycle proceeds through analysis, drafting, AI refinement, cross-checking, and redrafting. Students first analyze a text, attending closely to its lexical choices; they then draft a written response that deploys target vocabulary; they consult AI tools to refine their lexical choices; they cross-check the AI's suggestions against established resources and their own judgement; and they redraft accordingly. Critically, AI is positioned within this cycle not as a generator of text but as a feedback partner—one input among several that the student must evaluate. This iterative process deepens lexical processing and fosters the metacognitive awareness that the sixth learning outcome targets.

6.6. Teaching Methods

The course employs a range of teaching methods designed to enact its pedagogical commitments. Engaging teaching and learning activities, including lectures and guided exercises, introduce key vocabulary concepts, word formation, and contextual analysis techniques, establishing the strong foundation that subsequent independent work requires. Multimodal materials—videos, infographics, articles, and digital tools—are integrated to support diverse learning preferences and to enrich vocabulary acquisition through varied modes of encounter.

Interactive and gamified activities enhanced by AI tools form a distinctive component. Students participate in vocabulary games, quizzes, and collaborative problem-solving tasks in which AI tools provide personalized feedback and adaptive support. This use of AI for individualized, responsive practice resonates with the broader vision of AI in education articulated by Luckin, Holmes, Griffiths, and Forcier (2016), who argue that AI's promise lies in its capacity to provide the kind of personalized support that human teachers, constrained by time and scale, often cannot. Yet in keeping with the course's overarching principle, this personalization is framed as support for the learner's own work rather than as a replacement for it.

Finally, student-centred interaction—through weekly presentations, peer evaluations, and class discussions—encourages students to practice new vocabulary, articulate ideas precisely, and analyze word meanings in real-world contexts. This emphasis on interaction reflects the communicative and social conception of learning that has animated ELT since the communicative turn, and it ensures that the relatedness dimension of self-determination theory (Deci & Ryan, 2000), alongside autonomy and competence, is addressed.

6.7. The Week-by-Week Architecture

The course unfolds across fourteen weeks, structured to enact the deliberate progression from guided instruction to independent practice that the scaffolding metaphor demands. The fading

of support is not incidental but is built into the temporal architecture of the semester: the density of explicit instruction is greatest at the outset and diminishes as students assume increasing responsibility for their own lexical decision-making, in direct expression of the principle articulated by Wood, Bruner, and Ross (1976) that the scaffold exists to be withdrawn.

The opening three weeks establish foundations. Week one introduces the course's central premise—the distinction between vocabulary size and vocabulary depth (Nation, 2001)—and orients students to the role AI will play, explicitly framing it from the first day as a scaffold to be interrogated rather than an authority to be obeyed. Week two develops morphological analysis, equipping students to decompose unfamiliar words into roots, prefixes, and suffixes, while week three addresses word families, synonymy, and antonymy. Throughout this opening phase the cognitive load is carefully managed in the manner Sweller (1988) prescribes, with explicit instruction and worked examples predominating so that students can build robust schemata before being asked to apply them independently.

Weeks four through seven turn to contextual and textual analysis. Here students learn to infer meaning from context clues, to distinguish connotation from denotation, and to perceive the fine discriminations that separate near-synonyms. These weeks introduce the literary short texts—Jackson's "The Lottery" and O'Connor's "A Good Man is Hard to Find"—as sites for the close reading of lexical choice, and they inaugurate the first full passes through the reading-into-writing cycle. It is in this phase that AI is first systematically integrated into the cycle's refinement and cross-checking stages, with substantial in-class time devoted to the collective critique of AI suggestions so that students develop a shared vocabulary of evaluation before working independently.

Weeks eight through eleven shift the centre of gravity toward genre and application. Students engage successively with academic prose drawn from journal articles and IELTS reading passages, with literary registers drawn from the longer novels on the reading list, and with the demands of formal and technical writing, learning to calibrate lexical choice to the conventions of each discourse community (Hyland, 2019). The weekly presentations intensify during this phase, and the gamified, AI-supported practice activities are increasingly directed toward genre-specific lexical discrimination. Crucially, the proportion of in-class instruction devoted to AI critique diminishes here, as students are expected to conduct cross-checking with growing autonomy.

The final three weeks consolidate and individualize. Weeks twelve and thirteen are given over largely to the drafting, AI-refinement, cross-checking, and redrafting of the final paper's textual analysis, with the teacher functioning as consultant rather than instructor—a deliberate enactment of the scaffold's withdrawal. Week fourteen foregrounds the reflective component, guiding students through the metacognitive self-assessment and goal-setting that the sixth

learning outcome targets and that constitutes the culmination of a semester designed throughout to produce self-regulating learners.

7. AI as Scaffold in Practice

The course design presented above is an attempt to embody, in concrete pedagogical practice, the argument that AI is a continuation rather than a rupture in the history of ELT. This framing is itself consonant with Williamson and Eynon's (2020) historical account, which traces the long lineage of AI in education and cautions against the recurrent tendency to greet each new technology as an unprecedented break with what came before; seen against this longer arc, generative AI appears less as a revolution than as the latest instance of a familiar pattern in which novel tools are absorbed into, and reshaped by, enduring pedagogical purposes. This section steps back to consider the broader implications of that attempt, drawing on three bodies of work in turn. It draws, first, on frameworks for technology integration, and in particular on the technological, pedagogical and content knowledge framework (TPACK) of Mishra and Koehler (2006), which insists that the value of any technology is realized only at its intersection with pedagogy and content rather than in the technology itself.

It draws, second, on the wider literature concerning the purpose of education, and especially on Biesta's (2009) argument that educational decisions must be governed by a prior reckoning with what education is for, rather than by the seductions of measurable efficiency—a caution acutely relevant to a technology that arrives bearing precisely the promise of quantifiable gains. And it draws, third, on the state of AI research in higher education, where the systematic review of Zawacki-Richter et al. (2019) and the more recent survey of Crompton and Burke (2023) together document a field still weighted toward technical and efficacy-oriented studies and still markedly short of educator-led, theoretically grounded, and ethically reflective work. Read together, these three literatures supply the criteria against which the course's design choices can be evaluated, and they clarify the sense in which an educator-led, theory-first project of this kind constitutes a deliberate response to the gaps the field has acknowledged in itself.

7.1. Positioning AI Pedagogically

The theoretical case for positioning AI as scaffold, developed in Section 3.4, translates into a concrete set of practices. The decisive move is to require that students treat AI feedback as provisional input rather than as unquestioned authority. In practical terms, this means that when students consult AI tools to refine their vocabulary choices, they are required to document the feedback they receive, to evaluate it against other sources, and to justify their final decisions. The AI's suggestion is never the end of the process; it is the beginning of an act of evaluation.

This practice directly enacts the scaffolding principle articulated by Wood, Bruner, and Ross (1976): the AI provides temporary support that helps the learner accomplish more than they

could unaided, but the support is structured so that the learner retains and develops their own evaluative capacity. It also manages cognitive load in the manner described by Sweller (1988), since the AI can handle the generative burden of surfacing options, freeing the learner to concentrate on the higher-order task of evaluation and justification. And crucially, by requiring students to interrogate, verify, and justify AI suggestions, the practice ensures that the evaluative authority remains with the learner—preventing the collapse into dependency that would occur if AI were treated as an authority to be obeyed.

The design also anticipates the gradual withdrawal of the scaffold that defines genuine scaffolding. As students develop competence over the course of the semester, their reliance on AI for surfacing lexical options is expected to diminish, and their capacity to make and justify lexical choices independently is expected to grow. The aim, consistent with the scaffolding metaphor, is for the support eventually to become unnecessary—for students to learn to think with AI and, ultimately, beyond it.

7.2. Anticipated Scenarios of Student–AI Interaction

The abstract principle of AI-as-scaffold acquires concrete meaning only when illustrated in the texture of practice. Because the course awaits its first implementation, the following scenarios are offered not as collected data but as hypothetical illustrations of the kinds of interaction the course is designed to elicit. They are imagined cases, constructed to demonstrate how the cross-checking stage of the reading-into-writing cycle is intended to keep the evaluative authority with the learner. Their purpose is to make the pedagogical design legible in advance of teaching, and they should be read as projected possibilities rather than as evidence of what has occurred.

Consider a hypothetical student writing a textual analysis of Toni Morrison's *Beloved* who wishes to characterize the novel's atmosphere and drafts the sentence, "The house has a scary presence." Sensing that "scary" is tonally inadequate to Morrison's prose, such a student might ask an AI tool for more sophisticated alternatives and receive a list: eerie, sinister, menacing, haunting, ominous, foreboding. Under the course's design, this list would not be an answer but the beginning of inquiry. The student would be required to cross-check each candidate, and in doing so would confront genuine discriminations of meaning. Consulting a collocation resource and a learner's dictionary, the student might observe that "haunting" carries a connotation of lingering, melancholic beauty that resonates with the novel's preoccupation with memory, whereas "menacing" foregrounds active threat, and "sinister" implies concealed evil. The student would then need to reason about which connotation serves the analytical claim being made.

If the argument concerned the way the past refuses to depart, "haunting" would be not merely a more elevated synonym but the lexically precise choice—and, in a felicitous discovery, one that puns productively on the novel's literal haunting. The student might annotate the redraft

accordingly: "Chose haunting over menacing because the analysis concerns the persistence of memory rather than active threat; haunting also echoes the novel's literal ghost." In such a scenario the AI would have widened the field of options, but the act of discrimination, the reasoning that connects word to argument, and the justification would all reside with the learner—precisely the knowledge-transforming work that Bereiter and Scardamalia (1987) describe.

A second scenario illustrates the equally important lesson that AI can mislead. Imagine a student analyzing the formal register of an academic article who drafts the phrase "this proves that the hypothesis is correct" and asks an AI tool whether the phrasing is appropriately academic. Suppose the tool were to affirm that the sentence is "clear and effective." A student trained only to accept AI output would stop here. But the course's cross-checking discipline would direct the student to verify the suggestion against the conventions of academic discourse, where the verb "proves" expresses a degree of certainty that empirical research conventions typically avoid, preferring hedged formulations such as "suggests," "indicates," or "provides evidence that." Recognizing the mismatch between the AI's blanket approval and the genre's hedging conventions (Hyland, 2019), the student would revise to "this provides strong evidence that the hypothesis is supported" and record the reasoning: "AI approved the original, but academic register requires hedging; proves overstates the epistemic claim." Such a scenario would be pedagogically valuable precisely because the AI was unhelpful; it would teach the limitation that Ng (2012) identifies as central to genuine digital literacy, namely the capacity to recognize where a tool falls short.

A third scenario concerns the danger of plausible fabrication. Suppose a student exploring collocations for the word "consequences" were to ask an AI tool for examples and receive, among accurate collocations such as "far-reaching consequences" and "unintended consequences," a confidently asserted but dubious one. Directed by the course to cross-check collocational claims against a corpus tool, the student would discover that the suggested collocation is vanishingly rare in actual usage. Such an episode would teach, in a manner no abstract warning could, that fluent and confident AI output is not the same as accurate output—a lesson directly pertinent to the broader cultivation of critical AI literacy discussed below.

What unites these anticipated scenarios is the structural feature that the AI's contribution is always to be made subject to an evaluative act that the student must perform and document. The documentation requirement is not bureaucratic but pedagogical: by externalizing the reasoning, it is intended to render the evaluative process visible to the student, to peers in class critique, and to the teacher, and to furnish the reflective and documentary data on which the action research design depends.

7.3. Cultivating Critical AI Literacy

Beyond vocabulary development, the course pursues a second, broader aim: the cultivation of critical AI literacy. This aim responds to the recognition that students entering an AI-saturated world need not merely to use AI tools but to understand and critically evaluate them. Selwyn (2019), in his examination of AI and the future of education, argues forcefully that AI in education must be approached critically rather than with either uncritical enthusiasm or reflexive fear—that educators and learners alike must move beyond the binary of celebration and resistance toward informed, critical engagement.

This critical stance requires the development of the evaluative dimension of digital literacy. Ng (2012), in his analysis of digital literacy, emphasizes that genuine digital literacy includes the capacity to evaluate digital tools and to understand their limitations—a capacity that cannot be assumed even among students who are fluent users of technology. The course therefore explicitly teaches students to interrogate AI output, to verify it through established resources, to justify their decisions with evidence, and to recognize the limitations and potential errors of the tools they use.

Underlying this approach is a particular vision of the relationship between human and machine intelligence. Luckin et al. (2016) maintain that AI in education should augment human intelligence rather than replace it—that the goal is to enhance human capability, not to substitute for it. This vision informs the course's consistent positioning of AI as a partner in, rather than a replacement for, the learner's intellectual work. The ultimate goal of the course's AI literacy strand is discernment, not efficiency: students are not being trained to extract answers from AI as quickly as possible, but to exercise informed judgement about when, how, and whether to rely on it.

8. Assessment and Materials

A course's assessment regime reveals its true priorities, and the assessment design of this course is carefully aligned with its learning outcomes and pedagogical commitments. This section details the assessment methods, their alignment with the learning outcomes, and the materials that support the course.

8.1. Mixed Assessment Methods

The course employs a mixed assessment regime comprising four components. The first is classwork in the form of in-class presentations, weighted at thirty per cent of the overall grade. Each student delivers ten three-minute presentations across the fourteen-week course, with the presentation of new vocabulary contributing to the assessment of each. This component privileges the communicative and applied dimensions of vocabulary learning, directly targeting the third learning outcome (incorporating new vocabulary into discussions and presentations)

and the fifth (applying genre-appropriate vocabulary techniques), while also engaging the first, fourth, and others.

The second component is a mid-term test, weighted at twenty per cent, consisting of multiple-choice and open-ended questions on vocabulary and textual analysis. This component assesses the foundational capacities of strategy use, lexical discrimination, and textual analysis—the first, second, fourth, and fifth learning outcomes—providing a checkpoint on the knowledge and skills that the latter part of the course builds upon.

The third and most heavily weighted component, at forty per cent, is the final paper, comprising a seven-hundred-word textual analysis and a five-hundred-word reflection on vocabulary building. The textual analysis assesses the higher-order capacities of analysis, interpretation, and genre-appropriate lexical choice (the first, second, fourth, and fifth outcomes), while the reflection assesses the metacognitive and self-directed capacities targeted by the sixth outcome.

The fourth component, attendance and participation, is weighted at ten per cent, recognizing the importance of the sustained, interactive engagement through which much of the course's learning occurs.

8.2. Reflection and Metacognition

The five-hundred-word reflection within the final paper merits particular attention, because it embodies a central commitment of the course. By requiring students to assess their own vocabulary growth and usage and to set personal goals for improvement, the reflection operationalizes the sixth learning outcome and enacts the course's broader commitment to learner autonomy and self-direction. This commitment connects directly to self-determination theory (Deci & Ryan, 2000), since reflective self-assessment supports the learner's sense of autonomy and competence, and to the historical trajectory toward autonomy traced in Section 2. The reflection is not an afterthought but the culmination of a course designed throughout to develop self-regulating learners.

8.3. Materials and Texts

The course draws on a carefully curated range of materials selected to provide rich, contextualised lexical encounters across genres. The required readings include established vocabulary-building guides that furnish explicit instruction in word-building strategies. The recommended readings, organized by weekly theme, span a deliberately diverse range: vocabulary-building guides for foundational strategy instruction; textual analysis guides to support analytical work; international short stories such as Shirley Jackson's "The Lottery" and Flannery O'Connor's "A Good Man is Hard to Find" for the study of vocabulary in literary context; sample IELTS reading passages for the development of academic reading strategies; academic journal articles for engagement with scholarly vocabulary; novels such as Chinua

Achebe's *Things Fall Apart*, Amy Tan's *The Joy Luck Club*, Oscar Wilde's *The Picture of Dorian Gray*, Toni Morrison's *Beloved*, J.D. Salinger's *The Catcher in the Rye*, Cormac McCarthy's *The Road*, and Alice Walker's *The Color Purple* for thematic and tonal exploration; and writing guides to support the composition of the final paper.

This diversity is purposeful. By exposing students to vocabulary in literary, journalistic, academic, and technical contexts, the materials provide precisely the multiple meaningful encounters across varied contexts that Webb (2007) identifies as essential to deep, durable vocabulary learning, and they support the genre-sensitivity that the fifth learning outcome targets. The selection of literary works from a range of national and cultural traditions also reflects the liberal arts mission of the host institution, treating vocabulary development as a pathway into the broader intellectual and cultural engagement that close reading affords.

8.4. Instruments for Measuring Vocabulary Depth and Analytic Procedures

The module's principal depth instrument is the thirty-item Vocabulary Baseline Test administered under identical conditions in Week 2 and Week 13 (**Appendix D**). Its design draws directly on the logic of Wesche and Paribakht's Vocabulary Knowledge Scale (VKS), which distinguishes self-reported familiarity from demonstrable knowledge of meaning. Each item elicits two responses: a binary self-rating ("I know this word"), which indexes confidence, and a forced-choice definition-matching task, which indexes accuracy. Items are stratified across the CEFR range (A2–C2) and sampled from every module, so that the instrument tracks depth across the full curricular vocabulary rather than a single register.

The critical measure is not either score in isolation but the gap between them. Words that a student defines correctly without having claimed to "know" mark the transition from passive recognition toward active control—the very shift the module theorises. This paired design thus operationalizes depth as movement along a receptive–productive continuum rather than as a simple count of known items. Because the module's central construct is collocational competence, the baseline is complemented by the collocation-focused items embedded in the assessed tasks (Section 4.6), where corpus verification requires students to demonstrate knowledge of habitual word partnerships rather than isolated meanings.

Analysis proceeds by comparing paired Week 2 and Week 13 sheets for each student. For the small elective cohort, the primary procedure is descriptive: gain scores for both the "known" and "correct-definition" totals, and change in the recognition–accuracy gap.

9. Anticipated Challenges and Ethical Safeguards

A course that places AI at the centre of its pedagogy cannot responsibly proceed without confronting, in advance, the challenges and ethical hazards that such integration entails. This section identifies the principal risks the design anticipates and the safeguards it builds in to

address them, in the conviction that ethical reflection belongs at the design stage rather than as a retrospective concern.

9.1. Academic Integrity and the Risk of Outsourced Thinking

The most immediate challenge is that of academic integrity. The same tools that can scaffold lexical reasoning can, if misused, generate entire textual analyzes, reducing the student to a transcriber of machine output. Cotton, Cotton, and Shipway (2024), examining the implications of generative AI for academic integrity, observe that conventional plagiarism detection is poorly equipped to address text that is generated rather than copied, and they argue that the durable response lies less in detection than in the redesign of assessment so that the process of learning, and not merely its product, is made visible and valued. The present course embodies precisely this strategy.

By assessing the reading-into-writing cycle as a documented process—requiring annotated drafts, logs of AI interaction, and justifications of lexical choice—the design makes the student's reasoning, rather than the finished text alone, the object of assessment. A student who outsourced the analysis to AI would be unable to produce the trail of evaluative reasoning the assessment demands. The weekly in-person presentations and the reflective component further anchor assessment in capacities that cannot be delegated to a machine, since they require the student to articulate and defend their thinking in real time. The aim is not to police AI use but to design assessment such that genuine engagement is the path of least resistance.

9.2. Over-Reliance and the Erosion of Competence

A subtler challenge is that of over-reliance. There is a genuine risk that scaffolding intended to be temporary becomes a permanent crutch, and that students who lean habitually on AI to surface lexical options fail to develop the independent capacity the course exists to cultivate. This is the very dependency that the scaffolding framework warns against (Wood, Bruner, & Ross, 1976), and it cannot be assumed away simply because the design intends the scaffold to fade. The course therefore builds the fading of support into its temporal architecture, as described in Section 6.7, deliberately reducing the role of AI and of explicit instruction across the semester. Assessment components such as the mid-term test and the in-class presentations require performance without AI assistance, providing checkpoints on independent competence. Most importantly, the reflective component asks students to evaluate their own evolving relationship with the tools, prompting the metacognitive awareness of dependency that is the precondition for overcoming it. Whether the scaffold does in fact fade as intended is an empirical question that subsequent action research cycles are designed to answer.

9.3. Data Governance, Consent, and Privacy

The integration of AI tools raises significant questions of privacy and data protection that an educator-led design must take seriously rather than delegate to vendors. When students submit their writing to commercial AI systems, that text may be stored, processed, and used to train future models, often under terms of service that students have neither read nor meaningfully consented to. Slade and Prinsloo (2013), in their analysis of the ethics of learning analytics, articulate principles—transparency, student agency over one's own data, and the framing of students as participants rather than as mere data sources—that apply with equal force to AI-mediated instruction.

The module adopts several safeguards in this spirit. Students are informed clearly about what data the tools collect and how it may be used, enabling genuine rather than nominal consent. Where feasible, the module favours tools with stronger privacy protections and counsels students against submitting sensitive personal information. Crucially, no student is required to submit graded work to third-party systems without an alternative, so that participation is not made contingent on the surrender of privacy. These measures align with the broader principles for the ethical use of AI in education articulated by Holmes et al. (2022), who call for explicit attention to data governance, transparency, and learner rights in any educational AI deployment.

Because the module is also conceived as an action-research site, a second layer of governance applies to any data gathered for evaluative or scholarly purposes. Before implementation, the study will seek approval from the institutional research ethics committee, and no baseline-test scores, critique logs, or presentation artefacts will be analyzed or reported for research unless the student has given separate, informed, written consent. That consent is distinct from, and does not condition, the mark: a student may complete every assessed task while declining to have any data used for research. All research data will be pseudonymized at the point of collection, stored on institutionally managed and access-restricted systems rather than on vendor platforms, and retained only for the period specified in the approved protocol. Consent will be revocable without penalty at any point during the cycle.

9.4. Workload Feasibility and Class-Size Contingency

The summative word presentation is the most workload-sensitive component, since a cohort in which every student presents individually generates as many as ten presentations to be scheduled, delivered, and assessed within the Week 13 window. Two features keep this manageable. First, the presentation is an existing LUE3008 assessment with an established rubric and weighting (Section 4.6), so it imposes no new marking architecture on tutors. Second, the criterion-referenced design allows assessment during delivery rather than after it, avoiding a post-hoc marking backlog.

For contingency, class size should govern presentation format rather than the assessment logic. In a small elective of up to roughly eighteen students, individual presentations remain feasible within a 180-minute session. The two novel instruments—the AI-critique log and the annotation task—scale linearly and asynchronously, and so absorb variation in class size without additional live contact time. Since the university has adopted a small-class teaching mode for all language courses, there should be no sudden surge of student enrollment.

9.5. Timeline and Success Criteria

The first action-research cycle maps onto a single semester: baseline test in Week 2, Pillar One activities and formative critique logs in Weeks 3–10, annotation task in Weeks 9–11, presentations spanned across Weeks 2 to 13, and the repeat baseline test in Week 13. Analysis and reflective write-up follow in the subsequent inter-semester period, feeding revisions into the next iteration.

While empirical outcomes are rightly reserved for future cycles, three provisional success criteria will anchor evaluation. First, measurable growth in the recognition–accuracy gap on the paired baseline test (Section 8.4), indicating movement toward active vocabulary control. Second, evidence in the critique logs that students verify claims against corpus data rather than intuition, judged against the corpus-verification criterion. Third, sustained feasibility, operationalized as completion rates and a manageable tutor workload across the cycle. These criteria are deliberately modest and diagnostic, consistent with an action-research trajectory whose purpose is iterative refinement rather than summative proof.

9.6. Bias, Equity, and Access

Two further challenges concern bias and equity. Because large language models reproduce patterns latent in their training data, their outputs may encode and propagate social and linguistic biases, including, pertinently for a language course, a tendency to privilege particular varieties of English and to treat the norms of native-speaker, and often specifically Anglo-American, usage as the unmarked standard (Bender et al., 2021). For a course taught to multilingual users of English in Hong Kong, this is not a remote concern: an AI tool may, for instance, flag as "incorrect" perfectly legitimate features of the students' own richly developed multilingual repertoires.

The course turns this hazard into pedagogical material, teaching students to recognize that AI judgements about "good" English reflect particular norms rather than neutral truths—an insight that itself deepens critical AI literacy and connects to broader debates about linguistic ownership and the global ecology of English. Equity of access poses a related challenge: if effective participation depended on paid AI subscriptions, the course would risk entrenching

the very inequalities education should mitigate. The design therefore relies on freely available tools and ensures that no graded outcome is contingent on access to premium services.

9.7. The Limits of Personalization and the Irreducible Role of the Teacher

A final challenge is conceptual as much as practical: the temptation to overstate what AI can do and, correspondingly, to understate the role of the teacher. Williamson and Eynon (2020), surveying the history and trajectory of AI in education, caution against the recurrent tendency to invest educational technology with transformative powers it does not possess, and against the quiet displacement of pedagogical judgement by automated systems. The safeguard here is the course's foundational stance, maintained throughout, that AI augments rather than replaces the teacher (Luckin et al., 2016). The personalization AI offers is real but limited; it cannot supply the relational, motivational, and judgmental dimensions of teaching that the relatedness need of self-determination theory (Deci & Ryan, 2000) identifies as essential. By keeping the teacher central—as designer of tasks, facilitator of critique, and cultivator of judgement—the course guards against the deskilling that uncritical automation invites.

10. Discussion: Designing for Reasoning over Convenience

The course design presented above is an attempt to embody, in concrete pedagogical practice, the argument that AI is a continuation rather than a rupture in the history of ELT. This section steps back to consider the broader implications of that attempt, drawing on frameworks for technology integration and on the wider literature concerning the purpose of education and the state of AI research in higher education.

10.1 Aligning Technology, Pedagogy, and Content

A recurring theme of this paper is that the value of AI depends not on the technology itself but on how it is integrated with pedagogy and content. This insight is given systematic expression in the TPACK framework. Mishra and Koehler (2006), in their influential account of technological pedagogical content knowledge, argue that effective technology integration requires the alignment of three forms of knowledge: knowledge of the technology, knowledge of pedagogy, and knowledge of the subject matter. Technology deployed without regard to pedagogy or content is unlikely to produce learning gains; it is only at the intersection of all three that meaningful integration occurs.

The course described here can be understood as an attempt to work at this intersection. The integration of AI tools is consistently aligned with a pedagogy of scaffolding, constructivism, and learner autonomy, and with the specific content of vocabulary development understood as depth, expansion, and informed lexical choice. The TPACK framework thus provides a useful

lens for evaluating the course: its design strives to ensure that the technological, pedagogical, and content dimensions are mutually reinforcing rather than independently pursued.

10.2 Foregrounding Educational Purpose

Beneath the question of technology integration lies a deeper question about the purpose of education itself. Biesta (2009), in his influential critique of the contemporary preoccupation with measurement, warns against allowing education to be driven by the measurement and comparison of outcomes at the expense of a genuine reckoning with the question of purpose—the question of what constitutes good education. In an age increasingly enamoured of metrics, dashboards, and quantifiable performance, Biesta insists that educators must reconnect with the prior question of what education is *for*.

This warning is acutely relevant to the integration of AI, which arrives bearing precisely the promise of measurable efficiency: faster feedback, automated assessment, quantified progress. The course described here is conceived as an instance of foregrounding purpose over measurement—of asking, in the spirit of Biesta's argument and echoing the broader caution that one should not strive for what is technically possible but for what makes pedagogical sense, what is pedagogically meaningful rather than what is technologically possible. The course's insistence that AI function as a scaffold for reasoning rather than a generator of answers, and its emphasis on discernment over efficiency, reflect a deliberate prioritization of educational purpose over the seductive metrics of technological convenience.

10.3 The Ethical Dimension and the State of the Field

The ethical safeguards detailed in Section 9 are not peripheral additions but integral to the paper's central claim, and situating them within the wider literature clarifies the project's contribution. Zawacki-Richter et al. (2019), in their systematic review, documented two findings of particular relevance. The first is the dramatic lack of critical reflection on the pedagogical and ethical implications of AI in higher education: of the 146 studies they reviewed, only a tiny fraction critically engaged with the challenges and risks of AI, and only a handful even defined the term. The second is the marked underrepresentation of educators in the field: the overwhelming majority of research came from computer science and STEM, with very few first authors affiliated with education departments. Their review thus issued a clear call for greater educational perspective, theoretical grounding, and ethical reflection in education action research.

More recent surveys suggest that this gap, while narrowing, has not closed. Crompton and Burke (2023), reviewing the state of the field of AI in higher education, find a continuing predominance of efficacy-oriented and technical studies and a relative paucity of work grounded in educational theory or attentive to ethical complexity. The persistence of this

imbalance lends the present project a continuing relevance, and the course can be understood as a deliberate response to the call these reviews issue. It is an educator-led project that begins from pedagogical theory rather than technological capability, that foregrounds critical and ethical reflection through both its cultivation of critical AI literacy and its explicit ethical safeguards, and that is candid about the theoretical traditions underpinning its design. In these respects it attempts to model the kind of pedagogically grounded, theoretically explicit engagement with AI that the literature has found so conspicuously absent.

This ethical orientation can be situated within emerging frameworks for the responsible use of AI more broadly. Floridi and Cowls (2019), synthesizing prior statements of AI ethics, articulate a set of unifying principles—beneficence, non-maleficence, autonomy, justice, and explicability—that map readily onto the concerns the course addresses: beneficence in its aim to deepen reasoning, non-maleficence in its safeguards against over-reliance and privacy harm, autonomy in its preservation of the learner's evaluative authority, justice in its attention to equity of access and linguistic bias, and explicability in its teaching of how AI tools function and fail. The convergence between these general principles and the specific safeguards of a vocabulary course suggests that pedagogically grounded design and ethically responsible design are not separate undertakings but two aspects of the same commitment. In modelling this convergence, the project attempts to demonstrate that the educator perspective whose absence the field has lamented is not merely a missing voice but a necessary one—necessary precisely because it is educators who are positioned to hold the technical, the pedagogical, and the ethical together in a single, coherent practice.

11. Conclusion

This paper has argued that the entry of artificial intelligence into English language teaching should be understood not as a pedagogical rupture but as a continuation of a long-standing trajectory. Historically, ELT has moved steadily from the rote memorization and teacher authority of the grammar-translation era, through the meaning-focused interaction of the communicative turn and task-based learning, to the autonomy-fostering tools of the CALL era—a consistent movement from passive reception toward engagement, agency, and reasoning. Theoretically, the principles of constructivism, self-determination theory, the knowledge-telling/knowledge-transforming distinction, and scaffolding and cognitive load theory all converge on the conclusion that AI, properly framed, can support rather than subvert the deep learning these traditions describe. In both respects, AI represents continuity rather than rupture—provided it is designed for reasoning over convenience.

11.1 A Practical Model

The course described here offers a practical model of what such a design can look like. By treating vocabulary as a test case, by adopting a scaffolded, blended, reading-into-writing architecture, and by consistently positioning AI as a scaffold to be interrogated rather than an authority to be obeyed, the course demonstrates how AI-mediated learning can be grounded in sound pedagogical principles. In doing so it enacts, at the level of classroom practice, the central claim of Mishra and Koehler's (2006) framework: that a technology yields educational value not in isolation but only where it is deliberately integrated with content and pedagogy, so that the affordances of generative AI are made to serve the specific demands of vocabulary development rather than dictating them. Its insistence that the cognitive work of evaluation and justification remain with the learner, and its cultivation of critical AI literacy, embody the conviction that the goal of AI integration is discernment, not efficiency.

This is, in effect, a practical answer to Biesta's (2009) question of what education is for: the course refuses to let the measurable efficiency the technology so readily promises stand in for the slower, more demanding cultivation of judgement, and it treats the learner's capacity to interrogate, weigh, and justify as the proper end of instruction rather than an obstacle to be automated away. In this respect the design also responds to the gaps that the field has documented in its own scholarship. Where the reviews of Zawacki-Richter et al. (2019) and Crompton and Burke (2023) find a literature dominated by technical and efficacy-oriented studies and conspicuously short of educator-led, theoretically grounded, and ethically reflective work, the present course models precisely the kind of intervention they call for—one designed from a pedagogical premise rather than a technological one.

And by positioning AI as the latest tool to be absorbed into enduring educational purposes rather than as a force that overturns them, it remains faithful to the historical perspective of Williamson and Eynon (2020), demonstrating that the continuity thesis advanced throughout this discussion is not merely an argumentative stance but a workable foundation for design.

11.2 The Future of ELT

The future of English language teaching is not, on this account, one in which AI replaces teachers. It is one in which teachers design environments where students learn to think with AI—and, as their competence grows and the scaffold is withdrawn, to think beyond it. The teacher's role, far from being diminished, becomes more important than ever: it is the teacher who frames the technology pedagogically, who designs the tasks that turn AI from a shortcut into a scaffold, and who cultivates in students the critical judgement to use AI wisely. This is precisely the work that Mishra and Koehler's (2006) framework reserves for the educator, for if technology acquires its value only at the point where it is integrated with pedagogy and content, then it is the teacher's design decisions, and not the capabilities of the tool, that

determine whether AI becomes a prop or a crutch. It is also work that cannot be delegated to the technology itself, since the question of what the cultivation of judgement is for—Biesta's (2009) question—is an educational rather than a technical one, and answering it well demands a human educator willing to subordinate the lure of efficiency to the harder purpose of formation.

In this sense, the integration of AI does not displace the teacher's real task but clarifies it. And it is a clarification the field urgently needs, for if the research surveyed by Zawacki-Richter et al. (2019) and Crompton and Burke (2023) has too often left educators on the margins of a technology-driven conversation, then a future built on this account restores them to its centre—not as operators of new machinery, but as the designers of the environments in which that machinery is made to serve learning. Understood this way, the arrival of generative AI is, as Williamson and Eynon (2020) would remind us, less a rupture than the latest occasion for teachers to do what they have always done: to absorb a new tool into the enduring purpose of helping students learn to think for themselves.

11.3 Limitations and Future Directions

Several limitations must be acknowledged. Most importantly, the course described here is currently under construction; this paper documents its design rationale rather than its outcomes. As an action research project, the course will require systematic evaluation of its actual pedagogical impact once implemented—evaluation that should attend not only to measurable vocabulary gains but to the deeper capacities of reasoning, lexical decision-making, and critical AI literacy that the course aspires to develop.

The AIEd literature is notable for its lack of longitudinal and impact studies (Zawacki-Richter et al., 2019), and future work on this course should aim to address that gap by tracking student development over time and, ideally, beyond the duration of the course itself. Such evaluation will be essential to determining whether the pedagogical aspirations articulated here are in fact realized in practice. Until then, this paper offers the design as a contribution to the ongoing conversation about how educators might integrate AI in ways that honour, rather than abandon, the purposes of education.

Acknowledgment

This paper is an output of TDG and TDLEG project supported by CEAL, Lingnan University. The project is entitled: AI-Mediated Language Classrooms and Vocabulary Retention: A Comparative Study of Student Responses in Core and Elective University English Courses (Project code: 102770).

References

Anderson, G. L., & Herr, K. (2015). *The action research dissertation: A guide for students and faculty* (2nd ed.). SAGE.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Benson, P. (2011). *Teaching and researching autonomy* (2nd ed.). Routledge.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Lawrence Erlbaum Associates.

Biesta, G. (2009). Good education in an age of measurement: On the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation and Accountability, 21*(1), 33–46. https://doi.org/10.1007/s11092-008-9064-9

Burns, A. (2010). *Doing action research in English language teaching: A guide for practitioners*. Routledge.

Carr, W., & Kemmis, S. (1986). *Becoming critical: Education, knowledge and action research*. Falmer Press.

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 61*(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education, 20*, Article 22. https://doi.org/10.1186/s41239-023-00392-8

Davies, M. (2008). *The Corpus of Contemporary American English (COCA)*. https://www.english-corpora.org/coca/

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

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1). https://doi.org/10.1162/99608f92.8cd550d1

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication, 32*(4), 365–387. https://doi.org/10.58680/ccc198115885

Godwin-Jones, R. (2018). Using mobile technology to develop language skills and cultural understanding. *Language Learning & Technology, 22*(3), 1–17.

Graham, C. R. (2006). Blended learning systems: Definition, current trends, and future directions. In C. J. Bonk & C. R. Graham (Eds.), *The handbook of blended learning: Global perspectives, local designs* (pp. 3–21). Pfeiffer.

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., Santos, O. C., Rodrigo, M. T., Cukurova, M., Bittencourt, I. I., & Koedinger, K. R. (2022). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education, 32*(3), 504–526. https://doi.org/10.1007/s40593-021-00239-1

Hyland, K. (2019). *Second language writing* (2nd ed.). Cambridge University Press.

Information Technology Services Centre. (n.d.). *Guidelines on the use of generative AI tools and privacy-compliant configuration*. Lingnan University. Available: https://www.ln.edu.hk/itsc/services/teaching-learning-research-services/generative-ai-service

Jonassen, D. H. (1999). Designing constructivist learning environments. In C. M. Reigeluth (Ed.), *Instructional-design theories and models: A new paradigm of instructional theory* (Vol. II, pp. 215–239). Lawrence Erlbaum Associates.

Kemmis, S., McTaggart, R., & Nixon, R. (2014). *The action research planner: Doing critical participatory action research*. Springer.

Levy, M. (1997). *Computer-assisted language learning: Context and conceptualization*. Oxford University Press.

Lewin, K. (1946). Action research and minority problems. *Journal of Social Issues, 2*(4), 34–46. https://doi.org/10.1111/j.1540-4560.1946.tb02295.x

Littlewood, W. (1981). *Communicative language teaching: An introduction*. Cambridge University Press.

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson.

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

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.

Ng, W. (2012). Can we teach digital natives digital literacy? *Computers & Education, 59*(3), 1065–1078. https://doi.org/10.1016/j.compedu.2012.04.016

Norton, L. S. (2009). *Action research in teaching and learning: A practical guide to conducting pedagogical research in universities*. Routledge.

Richards, J. C., & Rodgers, T. S. (2014). *Approaches and methods in language teaching* (3rd ed.). Cambridge University Press.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329–363. https://doi.org/10.1177/1362168808089921

Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Polity Press.

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethics and dilemmas. *American Behavioral Scientist, 57*(10), 1510–1529. https://doi.org/10.1177/0002764213479366

Stockwell, G. (2013). Technology and motivation in English-language teaching and learning. In E. Ushioda (Ed.), *International perspectives on motivation* (pp. 156–175). Palgrave Macmillan.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

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

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28*(1), 46–65. https://doi.org/10.1093/applin/aml048

Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology, 45*(3), 223–235. https://doi.org/10.1080/17439884.2020.1798995

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

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education, 16*, Article 39. https://doi.org/10.1186/s41239-019-0171-0