EFFECTS OF AN AI-SUPPORTED CONCEPT MAPPING SYSTEM ON ACHIEVEMENT IN ENGLISH SENTENCE CONSTRUCTION AND LEARNING MOTIVATION AMONG THAI FIRST-YEAR UNDERGRADUATE STUDENTS

Meiqin HUANG[1] and Kitipoom VIPAHASNA[1]
1 Faculty of Technical Education, Rajamangala University of Technology Thanyaburi, Thailand; meiqin_h@mail.rmutt.ac.th (M. H.); kitipoom_v@rmutt.ac.th (K. V.)

ABSTRACT
This study developed and evaluated an AI-supported concept mapping system for English sentence construction and examined its effects on achievement and English learning motivation among Thai first-year undergraduate students. A two-phase quantitative design was employed. Phase 1 involved system development, expert evaluation, instrument validation, and pilot usability testing. Phase 2 used a nonequivalent pretest-posttest control-group quasi-experimental design with 60 students from two intact classes (experimental n = 30; control n = 30) over six weeks and twelve instructional sessions. The system was evaluated at the highest level for overall quality (M = 4.68, SD = 0.24) and usability (M = 4.53, SD = 0.42). ANCOVA showed a significant instructional-group effect on sentence-construction achievement, $F(1, 57) = 27.03$, $p < .001$, partial $\eta^2 = .322$, 95% CI [.136, .487], and English learning motivation, $F(1, 57) = 21.44$, $p < .001$, partial $\eta^2 = .273$, 95% CI [.097, .444]. Supplementary post-use evaluation indicated very high overall learner satisfaction and innovation acceptance (M = 4.56, SD = 0.29), with Perceived Learning Usefulness rated highest (M = 4.60, SD = 0.31). The findings indicate that the integrated AI-supported concept mapping environment was associated with stronger achievement, motivation, and favorable learner perceptions.

Keywords: AI-Supported Learning, Concept Mapping, English Sentence Construction, Generative AI, Learning Motivation, Learner Satisfaction, Innovation Acceptance

CITATION INFORMATION: Huang, M., & Vipahasna, K. (2026). Effects of an AI-Supported Concept Mapping System on Achievement in English Sentence Construction and Learning Motivation Among Thai First-Year Undergraduate Students. *Procedia of Multidisciplinary Research, 4*(11), 5.

INTRODUCTION

English sentence construction is a fundamental component of productive language ability because learners must coordinate vocabulary, grammatical relationships, word order, and intended meaning before they can communicate effectively in paragraphs, presentations, digital correspondence, and discipline-specific academic contexts. For first-year undergraduate students, the ability to construct accurate and meaningful sentences provides an essential foundation for subsequent development in writing, speaking, and academic communication. However, years of prior English study do not necessarily ensure that learners can independently produce grammatically acceptable sentences. Thai university students continue to demonstrate difficulties involving sentence structure, articles, subject-verb agreement, verb forms, singular and plural forms, fragments, and word order. Previous research with Thai EFL learners has similarly reported persistent sentence-level and grammatical errors that may interfere with comprehensibility and effective communication (Sermsook et al., 2017; Pawabutra & Sutakote, 2024).

One instructional challenge is that vocabulary knowledge and grammatical knowledge are often acquired as partially separated forms of knowledge. Learners may recognize individual words or recall a rule such as *can* + *base verb*, yet still fail to apply the rule when constructing an unfamiliar sentence. They may also perform successfully on recognition-oriented exercises while experiencing difficulty when required to organize sentence components independently. Such problems suggest that effective sentence instruction should support learners not only in remembering linguistic information but also in perceiving how words, meanings, and grammatical relationships combine to express complete propositions.

Concept mapping offers a potentially useful mechanism for making these relationships visible. A concept map represents concepts as nodes connected through explicitly labelled relationships (Novak & Cañas, 2008). Within the present instructional design, the labelled relationship serves not merely as a semantic connection but also as a grammatical bridge. For example, the relationship *Apple → has → seeds* can be transformed into the sentence *An apple has seeds*, while *Bird → can → fly* produces *A bird can fly*. The central concept provides the subject, the relationship label supplies a grammatical relationship, and the connected concept contributes an object, complement, action, attribute, or location. In this way, visual concept mapping may help learners perceive the internal relationships of a sentence before producing the complete linguistic form.

Recent developments in artificial intelligence further extend the possibilities of technology-supported language learning. Generative AI can provide explanations, examples, corrective feedback, guided questioning, and conversational assistance in real time. Research has reported potential benefits of AI-mediated language learning for achievement, motivation, confidence, and self-regulated learning when artificial intelligence is incorporated within structured pedagogical activities rather than used solely as an unrestricted answer generator (Wei, 2023; Lo et al., 2024). At the same time, generative AI introduces important instructional concerns. Generated responses may be inaccurate, inconsistent, unnecessarily complex, or inappropriate for a learner's proficiency level. Unrestricted use may also encourage students to accept or copy generated sentences without processing the grammatical and semantic relationships underlying them, while raising concerns about reliability, dependency, and responsible use (Lo et al., 2024; Thangthong et al., 2024).

For these reasons, the present study adopts an AI-supported rather than AI-driven instructional architecture. Teacher-validated concept-map content, sentence patterns, model sentences, feedback rules, and learning activities remain the instructional foundation, while generative AI functions as a contextual learning scaffold. The embedded GPT assistant is constrained by the current lesson context and is intended to provide short explanations, examples, clarification, and corrective assistance without replacing the learner's own sentence-construction process.

The system additionally incorporates immediate corrective feedback and transparent rule-based adaptive review so that sentence patterns associated with errors, repeated attempts, hint use, or slower responses can receive further practice. This approach maintains human instructional oversight while using AI to extend individualized learning support (UNESCO, 2023).

Although concept mapping and artificial intelligence have both been investigated in language education, an important integration gap remains. Concept-mapping research has largely emphasized vocabulary organization, reading comprehension, idea generation, or writing planning, whereas AI-supported language research has frequently focused on automated feedback, conversational assistance, or broader writing tasks. However, empirical evidence remains limited regarding a unified, teacher-controlled environment that integrates labelled concept-map relationships, structured sentence-construction activities, immediate corrective feedback, adaptive review, progress tracking, and contextual GPT assistance while simultaneously evaluating objectively measured sentence-construction achievement and English learning motivation among Thai first-year undergraduates.

Beyond measurable learning outcomes, the practical value of a learning innovation also depends on whether learners regard it as useful, compatible with their learning needs, satisfying to use, and worth adopting in future learning. Accordingly, the present study treats student satisfaction and innovation acceptance as supplementary post-use evaluation evidence rather than as additional experimental outcomes (Davis, 1989; Yu et al., 2024).

Accordingly, this study developed an AI-supported concept mapping system for foundational English sentence construction and examined its effectiveness in comparison with conventional English instruction. The study had three objectives: (1) to design, develop, and evaluate the content quality, instructional design quality, technical quality, and usability of the system; (2) to determine its effect on achievement in English sentence construction; and (3) to determine its effect on English learning motivation among Thai first-year undergraduate students.

LITERATURE REVIEWS

Concept Mapping and Visual Relationships in English Sentence Construction

Concept mapping is a graphical knowledge-representation approach in which concepts are represented as nodes and connected through explicitly labelled relationships. A concept, a linking phrase, and another concept form a proposition that communicates a meaningful relationship (Novak & Cañas, 2008). Rather than presenting information as isolated vocabulary items or grammatical rules, concept maps make relationships visible and encourage learners to integrate new information with existing knowledge. This principle is consistent with Meaningful Learning Theory, which proposes that learning becomes more meaningful when new information is connected substantively with knowledge already established in the learner's cognitive structure (Ausubel, 1968).

In English language learning, this relational structure is particularly relevant to sentence construction because learners must coordinate vocabulary, grammatical functions, word order, and meaning. In the present instructional approach, concept-map linking phrases are deliberately designed to function as both semantic and grammatical relationships. For example, the relationship *Bird → can → fly* represents a meaningful proposition and simultaneously provides the grammatical structure for constructing *A bird can fly*. Similarly, *Apple → has → seeds* provides a visible relationship that can be transformed into *An apple has seeds*. Concept maps therefore function not merely as vocabulary-organizing tools but as scaffolds connecting concept → relationship → sentence.

Research increasingly supports the application of concept mapping to productive language learning. Kaveh et al. (2023) found that mobile-assisted concept mapping enhanced EFL learners' writing fluency and motivation. Han (2025) synthesized 68 studies comprising 139

effect sizes and 14,656 participants and reported a large overall positive effect of concept mapping on second- and foreign-language learning, with particularly favorable outcomes for productive language skills. The findings suggest that explicit visual relationships may provide a useful foundation for language production, although previous concept-mapping studies have frequently concentrated on vocabulary, reading comprehension, idea organization, or extended writing rather than foundational sentence construction. This pattern further supports the need for a sentence-focused concept-mapping design.

AI-Supported Language Learning and Teacher-Controlled Generative AI

Artificial intelligence has expanded technology-supported language learning by enabling immediate feedback, individualized practice, adaptive learning, automated explanation, and conversational assistance. AI-supported environments may respond to learner performance more rapidly than conventional whole-class instruction and can provide additional opportunities for practice according to individual needs. Wei (2023), for example, reported significantly stronger English achievement, L2 motivation, and self-regulated learning among university EFL students receiving AI-mediated instruction than among those receiving conventional instruction. Learners associated the intervention with interactivity, individualized learning, rapid feedback, and opportunities for self-paced practice.

Generative AI further extends these possibilities by providing explanations, examples, sentence reformulations, grammatical feedback, and conversational interaction. Mahapatra (2024) reported positive effects of ChatGPT-supported formative feedback in university-level ESL writing, while Polakova and Ivenz (2024) identified improvements in EFL writing following structured ChatGPT feedback. Huang and Mizumoto (2024) also reported positive changes in motivation and writing self-efficacy. Collectively, these studies indicate that generative AI can support cognitive and affective outcomes when integrated into purposeful instructional activities.

Nevertheless, unrestricted generative AI does not automatically constitute effective pedagogy. Potential concerns include inaccurate responses, inappropriate difficulty, academic misconduct, overreliance, privacy, and reduced learner ownership. Lo et al. (2024), in a systematic review of empirical ChatGPT research in ESL/EFL education, identified both opportunities for personalized assistance and significant concerns involving accuracy, privacy, and academic integrity. Thai learners have similarly reported perceived linguistic and affective benefits from AI writing tools while expressing concerns about reliability, dependency, and inappropriate use (Thangthong et al., 2024).

The present system therefore adopts an AI-supported rather than AI-driven architecture. Core learning objectives, concept-map relationships, sentence patterns, examples, and instructional feedback remain teacher controlled. Generative AI functions as a lesson-constrained scaffold that can explain a grammatical rule, provide another example, clarify an error, or pose additional practice questions without replacing the learner's own sentence-production process.

Theoretical Foundations and Instructional Mechanisms

The instructional architecture of the AI-supported concept mapping system draws on five complementary theoretical perspectives. Rather than functioning as competing explanations, each perspective accounts for a distinct mechanism within the learning environment: Meaningful Learning Theory explains the organization of relational knowledge; instructional scaffolding explains the progression from supported to independent performance; the Cognitive Theory of Multimedia Learning explains the coordinated presentation of visual and verbal information; second-language acquisition principles explain the role of language production and the identification of linguistic gaps, while corrective feedback and adaptive retrieval provide mechanisms for subsequent guided practice; and the ARCS Model explains the motivational mechanisms underlying the instructional design.

Meaningful Learning Theory and Relational Knowledge Organization

Meaningful Learning Theory provides the principal rationale for organizing vocabulary, grammatical relationships, and sentence patterns as interconnected propositions rather than isolated linguistic elements. According to Ausubel (1968), new information is learned more meaningfully when it can be related substantively to relevant knowledge already present in the learner's cognitive structure. In the present system, concept maps make these relationships explicit by visually linking concepts through grammatical and semantic relationships. Learners can therefore perceive how vocabulary items and grammatical structures combine to form meaningful propositions before producing complete sentences.

Instructional Scaffolding and Progressive Independence

The instructional sequence was designed to provide progressively reduced support as learners moved toward independent sentence construction. Learners first examined complete concept relationships in Explore activities, subsequently reconstructed missing concepts or links in Complete the Map activities, and finally produced complete sentences in Build a Sentence activities. This progression reflects the principle of instructional scaffolding, in which structured assistance supports performance during early stages of learning and is gradually reduced as learners develop greater competence and independence (Wood et al., 1976).

Cognitive Theory of Multimedia Learning

The Cognitive Theory of Multimedia Learning provides the rationale for coordinating visual concept maps, written words, images, pronunciation support, and instructional feedback within the learning environment. Multimedia learning is most effective when verbal and visual information is organized in ways that support learners in selecting, organizing, and integrating relevant information without introducing unnecessary cognitive processing (Mayer, 2009, 2021). Accordingly, multimedia elements in the system were used to clarify conceptual and grammatical relationships rather than to serve as decorative additions.

Second-Language Production, Corrective Feedback, and Retrieval

Second-language acquisition principles provide the rationale for requiring learners to move beyond passive recognition toward active language production. Producing language can reveal gaps in linguistic knowledge that may remain unnoticed during receptive activities alone (Swain, 1995). The learning sequence therefore requires learners to identify grammatical relationships, attempt sentence production, receive immediate corrective feedback, revise inaccurate responses, and subsequently retrieve difficult patterns through adaptive review. The instructional mechanism can be summarized as: 1) notice the relationship; 2) attempt sentence production; 3) receive corrective feedback; 4) repair the sentence; and 5) retrieve and apply the pattern again. This sequence positions errors as opportunities for guided revision and repeated retrieval rather than as terminal outcomes.

ARCS Motivational Design

The affective dimension of the intervention is informed by Keller's ARCS Model, which conceptualizes motivation through Attention, Relevance, Confidence, and Satisfaction (Keller, 1987, 2010). Interactive concept maps, varied activity formats, and contextual assistance were designed to support Attention; sentence patterns connected to practical communication were intended to strengthen Relevance; progressive scaffolding, repeated attempts, and immediate feedback were designed to enhance Confidence; and successful task completion and visible learning progress were intended to contribute to Satisfaction. The framework does not assume that technology itself produces motivation; rather, motivational effects are expected to depend on the quality of instructional design and the learner's interaction with the learning environment.

Theoretical Rationale for Supplementary Learner Satisfaction and Innovation Acceptance Evaluation

In addition to the two primary experimental outcomes—sentence-construction achievement and English learning motivation—the study included a supplementary post-use evaluation to examine how learners perceived the developed innovation after direct experience with the system. This evaluation was conceptually distinct from the H1-H2 experimental pathway and was not treated as an additional treatment-effect outcome.

Technology-acceptance research identifies perceived usefulness and perceived ease of use as important determinants of users' acceptance of technology (Davis, 1989). In the present study, TAM was used as a conceptual reference rather than as the direct measurement model for the LSIAQ. Technology acceptance was therefore employed as a supplementary evaluative lens rather than as a primary explanatory theory. The 20-item Learner Satisfaction and Innovation Acceptance Questionnaire (LSIAQ) examined four post-use dimensions: Learning Satisfaction, Perceived Learning Usefulness, Learning Compatibility and Integration, and Innovation Acceptance and Intention to Use. These dimensions also correspond with higher-education evidence linking perceived usefulness, compatibility, satisfaction, and continued-use intention within AI-supported learning environments (Yu et al., 2024). Consequently, the LSIAQ provides complementary evidence concerning learners' perceptions of the practical value and acceptability of the developed system without being interpreted as an additional experimental outcome.

Empirical Evidence and Research Gap

Existing empirical evidence provides converging support for visual knowledge organization and structured AI-supported language instruction, but several gaps remain. Thai university learners continue to experience difficulties involving grammatical structure, verb forms, articles, agreement, and word order (Sermsook et al., 2017; Pawabutra & Sutakote, 2024). Previous concept-mapping research has examined receptive skills such as vocabulary and reading comprehension as well as broader productive skills such as writing and speaking (Han, 2025). Thai AI-writing research has also examined learners' perceptions of AI writing assistance (Thangthong et al., 2024), leaving independently measured foundational sentence-construction performance comparatively underexplored.

Evidence from Thai higher education also demonstrates the relevance of AI-supported English learning. Songsiengchai (2025), for example, reported positive learning and affective outcomes from ChatGPT-supported English instruction among Thai first-year undergraduates. However, evidence remains limited regarding a unified system that combines concept mapping, explicit grammatical relationships, progressive sentence construction, corrective feedback, adaptive review, progress tracking, and contextual GPT assistance.

Four gaps can therefore be identified. First, a technological integration gap exists because concept mapping, adaptive review, corrective feedback, learning analytics, and conversational AI have commonly been investigated as separate tools. Second, an instructional-design gap remains because many AI-supported writing studies focus on feedback for completed texts rather than making the relationship between vocabulary, grammar, and sentence construction visually explicit. Third, an outcome gap exists because relatively few studies simultaneously examine objectively assessed sentence-construction achievement and English learning motivation. Fourth, a contextual gap remains because limited experimental research has investigated such an integrated system among Thai first-year undergraduate learners. Together, these gaps indicate the need for an empirically evaluated, teacher-controlled AI-supported learning environment that integrates visual concept relationships, guided sentence production, adaptive review, corrective feedback, and contextual generative-AI assistance within a single instructional design.

Hypothesis Development and Conceptual Framework

Previous research indicates that concept mapping can facilitate productive language learning by making semantic relationships explicit, while structured AI-supported environments may provide individualized feedback, adaptive practice, and additional learning support. Within the present system, these mechanisms are integrated through labelled concept relationships, progressively demanding construction activities, corrective feedback, adaptive retrieval, and contextual GPT assistance. Consequently, students learning through the developed environment are expected to demonstrate stronger sentence-construction performance than students receiving conventional instruction.

H1: After controlling for pretest scores, students who learn through the AI-supported concept mapping system will demonstrate significantly higher posttest achievement in English sentence construction than students who receive conventional instruction.

Research also suggests that interactive, personalized, feedback-rich learning environments may contribute to stronger learner motivation. Within the ARCS framework, the developed system is expected to support attention through interactive activities, relevance through practical sentence use, confidence through progressive scaffolding and corrective feedback, and satisfaction through successful performance and visible progress. Therefore, the following hypothesis is proposed:

H2: After controlling for pre-intervention motivation, students who learn through the AI-supported concept mapping system will demonstrate significantly higher post-intervention English learning motivation than students who receive conventional instruction.

The proposed relationships between the instructional approach, its principal learning mechanisms, and the two dependent variables are summarized in Figure 1, Conceptual Framework of the Study.

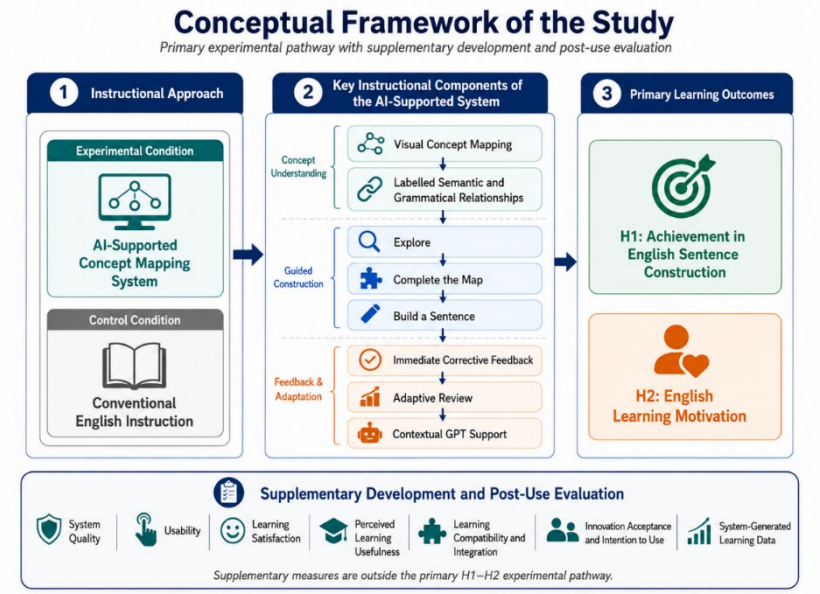


Figure 1 Conceptual Framework of the Study

Student satisfaction and innovation acceptance were evaluated as supplementary post-use measures outside the primary H1-H2 experimental pathway.

RESEARCH METHODOLOGY

Research Design

This study employed a two-phase quantitative research design consisting of system development and quasi-experimental evaluation. Phase 1 used Design and Development Research to design, develop, validate, and evaluate an AI-supported concept mapping system

for foundational English sentence construction. The development process followed the ADDIE model—Analysis, Design, Development, Implementation, and Evaluation—and incorporated iterative expert review and pilot testing before the system was implemented in the main study. Phase 2 employed a nonequivalent pretest-posttest control-group quasi-experimental design. The design was selected because participants were drawn from existing intact classes within the normal university instructional structure. One intact class was assigned by random draw to the experimental condition and the other to the control condition; individual students were not randomly assigned. The independent variable was instructional approach, consisting of AI-supported concept mapping instruction and conventional English instruction. The principal dependent variables were achievement in English sentence construction and English learning motivation.

Participants and Instructional Conditions

The population consisted of Thai first-year undergraduate students enrolled in a foundational English course at Rajamangala University of Technology Thanyaburi, Pathum Thani, Thailand. The sample comprised 60 students from two intact classes, with 30 students assigned to the experimental group and 30 to the control group.

A diagnostic screening procedure was conducted before the main study to identify students requiring foundational sentence-construction support. The screening addressed recognition of acceptable sentences, sentence relationships, word order, subject-verb agreement, verb forms, articles, singular and plural forms, sentence completion, and production of short meaningful sentences. The diagnostic screening was used for participant eligibility and was not treated as the principal pretest or an experimental outcome.

Both groups studied the same target sentence patterns, thematic vocabulary, learning objectives, and total instructional duration. The intervention lasted six weeks and consisted of twelve sessions, with two sessions per week and approximately 40 minutes per session. The experimental group learned through the AI-supported concept mapping system, which incorporated interactive concept maps, labelled semantic and grammatical relationships, pronunciation support, Explore activities, Complete the Map activities, Build a Sentence tasks, immediate corrective feedback, rule-based adaptive review, progress tracking, and contextual GPT assistance. The control group received conventional English instruction through lecturer explanations, presentation materials, worksheets, sentence-completion activities, controlled practice, and regular classroom feedback.

Research Instruments and System Evaluation

The principal achievement instrument was a validated English sentence-construction test administered before and after the intervention. The assessment measured students' abilities to recognize, complete, arrange, correct, and independently produce grammatically acceptable foundational English sentences. English learning motivation was assessed using a validated questionnaire administered before and after the intervention, addressing dimensions including interest, attention, perceived value, confidence, effort, persistence, satisfaction, participation, and willingness to continue learning English.

Additional instruments were used to evaluate the quality and usability of the developed system. Specialists evaluated content quality, instructional design quality, technical quality, and appropriateness for the target learners. A pilot evaluation examined functionality, clarity of instructions, learning difficulty, feedback accuracy, navigation, response time, usability, and the operation of the contextual GPT assistant.

Instrument quality was examined using appropriate validation and reliability procedures. Item-Objective Congruence (IOC) was used for content validation. Achievement-test items were examined for difficulty and discrimination, while internal consistency of objectively scored items was assessed using KR-20. Constructed-response scoring reliability was evaluated using

an intraclass correlation coefficient (ICC), and Cronbach's alpha was used for the motivation and usability questionnaires.

A 20-item LSIAQ was additionally administered to the 30 experimental-group students after the intervention. The five-point Likert instrument assessed Learning Satisfaction, Perceived Learning Usefulness, Learning Compatibility and Integration, and Innovation Acceptance and Intention to Use. Content validity was examined through IOC and internal consistency through Cronbach's alpha.

Data Collection Procedure

Data collection proceeded in five stages. First, institutional procedures, informed consent, participant screening, and class assignment were completed. Second, both groups completed the sentence-construction achievement pretest and pre-intervention motivation questionnaire. Third, the six-week intervention was implemented under equivalent content and instructional-time conditions. During implementation, attendance, activity completion, response accuracy, attempts, hint use, response time, adaptive-review status, and learning progress were recorded as supplementary implementation data for the experimental group. Fourth, both groups completed the achievement posttest and post-intervention motivation questionnaire. Finally, experimental-group students completed both the system-usability evaluation and the LSIAQ because they had direct experience with the developed learning innovation.

Data and Statistical Analysis

Descriptive statistics, including frequency, percentage, mean, and standard deviation, were used to summarize system evaluation, usability, sentence-construction achievement, English learning motivation, and supplementary system-use data. Instrument-quality statistics were also examined where appropriate to establish the validity, reliability, and consistency of the measurements used in the study.

The two research hypotheses were tested using analysis of covariance (ANCOVA). ANCOVA was selected because the study employed intact classes and pre-intervention measurements were available for both principal outcomes. By statistically controlling for baseline scores, ANCOVA reduces the influence of initial individual differences and provides a more precise comparison of post-intervention performance between the experimental and control groups. For H1, posttest sentence-construction achievement was compared while controlling for achievement pretest scores. For H2, post-intervention English learning motivation was compared while controlling for pre-intervention motivation.

Statistical significance was evaluated at the .05 level. In addition to p-values, adjusted means, adjusted mean differences with 95% confidence intervals, F-statistics, and partial eta squared (partial η^2) with 95% confidence intervals were reported. Adjusted means represent estimated post-intervention outcomes after accounting for baseline differences, while the 95% confidence intervals indicate the precision and plausible range of the adjusted group differences. Partial η^2 was used to quantify the magnitude of the instructional-group effect, thereby distinguishing statistical significance from practical or educational importance. Values of approximately .01, .06, and .14 were interpreted as small, moderate, and large effects, respectively.

Prior to conducting ANCOVA, the assumptions of independence of observations, linearity, homogeneity of regression slopes, normality of residuals, homogeneity of error variance, and absence of influential observations were examined. Linearity was assessed through visual inspection of group-specific scatterplots between each covariate and its corresponding post-intervention outcome. Extreme observations were evaluated using standardized residuals, with absolute values exceeding 3.0 considered potentially extreme, while Cook's distance values were examined to identify influential observations, with values below 1.0 indicating no substantial influence. Homogeneity of regression slopes was assessed using the Group × Pretest interaction. The ANCOVA models were interpreted only after these assumptions were considered adequately satisfied.

LSIAQ responses were summarized descriptively at item, dimension, and overall levels using mean and standard deviation, with frequency and percentage used where informative. These post-use results were supplementary and were not subjected to ANCOVA or used to test an additional hypothesis.

RESEARCH RESULTS

System Quality, Usability, and Instrument Quality

The developed AI-supported concept mapping system was evaluated by five specialists in English language teaching, instructional design, educational technology, and software or system development. Overall system quality was rated at the highest level (M = 4.68, SD = 0.24). Content quality obtained the highest rating (M = 4.76, SD = 0.20), followed by appropriateness and responsible AI implementation (M = 4.72, SD = 0.22), instructional design quality (M = 4.68, SD = 0.24), and technical quality (M = 4.56, SD = 0.28).

Pilot evaluation also indicated the highest level of system usability. Overall usability was rated at M = 4.53 (SD = 0.42), with perceived usefulness for sentence learning receiving the highest individual rating (M = 4.67, SD = 0.35). Clarity of instructions and feedback was also rated at the highest level (M = 4.60, SD = 0.39).

The research instruments demonstrated satisfactory validity and reliability. IOC values for the achievement test, motivation questionnaire, usability questionnaire, and LSIAQ ranged from 0.80 to 1.00. The objectively scored achievement items produced difficulty indices of 0.32-0.78 and discrimination indices of 0.27-0.73. Their KR-20 reliability coefficient was 0.86, while constructed-response scoring demonstrated an intraclass correlation coefficient of 0.91. Cronbach's alpha was 0.92 for the English learning motivation questionnaire, 0.90 for the usability questionnaire, and 0.87 for the 20-item LSIAQ.

Table 1 System Evaluation and Measurement Quality

Indicator	Result	Interpretation
Content quality	$M = 4.76$, $SD = 0.20$	Highest
Instructional design quality	$M = 4.68$, $SD = 0.24$	Highest
Technical quality	$M = 4.56$, $SD = 0.28$	Highest
Responsible AI implementation	$M = 4.72$, $SD = 0.22$	Highest
Overall system quality	$M = 4.68$, $SD = 0.24$	Highest
Overall usability	$M = 4.53$, $SD = 0.42$	Highest
Achievement test	KR-20 = 0.86	High reliability
Constructed responses	ICC = 0.91	Excellent agreement
Motivation questionnaire	$\alpha = 0.92$	Excellent reliability
Usability questionnaire	$\alpha = 0.90$	Excellent reliability
LSIAQ IOC	0.80-1.00	Acceptable-Excellent
LSIAQ Cronbach's α	0.87	High reliability

These findings provided satisfactory evidence that the developed learning environment and the principal measurement instruments were appropriate for subsequent experimental evaluation.

Preliminary and Descriptive Results

Complete outcome data were available for 60 students, comprising 30 students in each instructional group. Before the intervention, the two groups demonstrated closely comparable levels of sentence-construction achievement and learning motivation. The standardized mean difference at baseline was 0.07 for achievement and 0.10 for motivation, indicating only small initial differences between the groups.

Both groups demonstrated higher scores following the six-week instructional period, although the descriptive increases were larger in the experimental condition. Achievement in the

experimental group increased from $M = 18.40$ ($SD = 4.10$) to $M = 30.08$ ($SD = 4.23$), an increase of 11.68 points. The control group increased from $M = 18.10$ ($SD = 4.00$) to $M = 25.67$ ($SD = 4.13$), an increase of 7.57 points.
For English learning motivation, the experimental group increased from $M = 3.18$ ($SD = 0.40$) to $M = 3.81$ ($SD = 0.32$), while the control group increased from $M = 3.14$ ($SD = 0.39$) to $M = 3.49$ ($SD = 0.26$).

Table 2 Descriptive Pre- and Post-Intervention Outcomes

Outcome	Group	Pre-intervention M (SD)	Post-intervention M (SD)	Mean Increase
Sentence-construction achievement /40	Experimental	18.40 (4.10)	30.08 (4.23)	11.68
	Control	18.10 (4.00)	25.67 (4.13)	7.57
English learning motivation /5	Experimental	3.18 (0.40)	3.81 (0.32)	0.63
	Control	3.14 (0.39)	3.49 (0.26)	0.35

The descriptive increases were not treated as sufficient evidence of instructional effectiveness because participants came from intact classes and individual baseline differences remained possible. Accordingly, ANCOVA was used to test the two hypotheses after controlling for the corresponding pre-intervention scores.
Before conducting the ANCOVAs, statistical assumptions were examined. Residual normality was satisfactory for achievement, Shapiro-Wilk $W = .980$, $p = .438$, and motivation, $W = .961$, $p = .054$. Levene's tests were nonsignificant for achievement ($p = .985$) and motivation ($p = .139$). The Group × Pretest interactions were also nonsignificant for achievement, $F(1, 56) = 0.14$, $p = .711$, and motivation, $F(1, 56) = 1.65$, $p = .204$. Visual inspection of group-specific scatterplots indicated approximately linear relationships between the covariates and corresponding outcomes, with no pronounced curvilinear patterns. All standardized residuals were within ±3.0, and Cook's distance values were below 1.0, indicating no extreme or unduly influential observations. Consequently, no cases were excluded from the ANCOVA analyses.

Effects on English Sentence-Construction Achievement

After controlling for achievement pretest scores, the experimental group obtained an adjusted posttest mean of 29.98 ($SE = 0.57$, 95% CI [28.83, 31.12]), compared with 25.77 ($SE = 0.57$, 95% CI [24.63, 26.92]) for the control group. The adjusted difference between the groups was 4.20 points, 95% CI [2.58, 5.82].
ANCOVA revealed a statistically significant effect of instructional condition after adjustment for baseline achievement, $F(1, 57) = 27.03$, $p < .001$, partial $\eta^2 = .322$, 95% CI [.136, .487]. The effect size exceeded the conventional criterion for a large effect. The complete model accounted for approximately 57.3% of the variance in posttest achievement ($R^2 = .573$, adjusted $R^2 = .558$).
These results indicate that students who learned through the AI-supported concept mapping system demonstrated significantly stronger adjusted sentence-construction achievement than students receiving conventional instruction. H1 was therefore supported.

Table 3 Adjusted Posttest Means for English Sentence-Construction Achievement

Group	n	Raw Posttest M	Adjusted M	SE	95% CI
Experimental	30	30.08	29.98	0.57	[28.83, 31.12]
Control	30	25.67	25.77	0.57	[24.63, 26.92]

Note. Adjusted means were estimated at the overall achievement-pretest mean of 18.25.

Effects on English Learning Motivation

A similar pattern was observed for English learning motivation. After controlling for pre-intervention motivation, the experimental group obtained an adjusted post-intervention mean of 3.80 (*SE* = 0.05, 95% CI [3.71, 3.90]), while the control group obtained an adjusted mean of 3.50 (*SE* = 0.05, 95% CI [3.40, 3.59]). The adjusted difference was 0.31 points, with a 95% confidence interval of [0.17, 0.44].

ANCOVA indicated a statistically significant effect of instructional condition, $F(1, 57) = 21.44$, $p < .001$, partial $\eta^2 = .273$, 95% CI [.097, .444]. The magnitude of this effect was also large. Therefore, after accounting for students' initial motivation levels, learners using the AI-supported concept mapping system demonstrated significantly higher English learning motivation than learners receiving conventional instruction. H2 was supported.

Table 4 Adjusted Post-Intervention Means for English Learning Motivation

Group	n	Raw Post-intervention M	Adjusted M	SE	95% CI
Experimental	30	3.81	3.80	0.05	[3.71, 3.90]
Control	30	3.49	3.50	0.05	[3.40, 3.59]

Note. Adjusted means were estimated at the overall pre-intervention motivation mean of 3.16.

Table 5 ANCOVA Results for the Primary Learning Outcomes

Outcome	Exp. Adjusted M	Control Adjusted M	Adjusted Difference (95% CI)	$F(1, 57)$	p	Partial η^2 (95% CI)	Decision
Sentence-construction achievement	29.98	25.77	4.20 [2.58, 5.82]	27.03	< .001	.322 [.136, .487]	H1 Supported
English learning motivation	3.80	3.50	0.31 [0.17, 0.44]	21.44	< .001	.273 [.097, .444]	H2 Supported

System-Usage Findings

System-generated data provided supplementary evidence concerning learner engagement with the intervention but were not treated as independent tests of effectiveness. Experimental-group students completed an average of 11.63 of the twelve instructional sessions and approximately 94.80% of assigned learning activities. Overall response accuracy was 84.10%, with an average of 1.38 attempts per activity. Students used approximately 1.72 hints per instructional session, while the adaptive-review mechanism identified an average of 7.43 sentence patterns per learner requiring additional practice. Learners interacted with the contextual GPT assistant an average of 8.10 times during the intervention.

Performance also reflected the progressively demanding instructional structure. Accuracy was highest in Explore activities (92.10%), followed by Complete the Map (84.70%) and Build a Sentence (77.80%). Mean response time increased correspondingly from 22.80 seconds in Explore to 30.60 seconds in Complete the Map and 40.80 seconds in Build a Sentence. These

results should be interpreted descriptively because the three activities constituted successive components of the same intervention rather than independent experimental conditions.

Student Satisfaction and Innovation Acceptance

Supplementary post-use evaluation was completed by the 30 students in the experimental group. Table 6 presents the four LSIAQ dimensions and the overall score. These results are reported descriptively as supplementary evidence of learners' satisfaction with and acceptance of the developed learning innovation.

Table 6 Dimension-Level LSIAQ Results for the Experimental Group (n = 30)

Dimension	M	SD	Interpretation
Learning Satisfaction	4.59	0.33	Very High
Perceived Learning Usefulness	4.60	0.31	Very High
Learning Compatibility and Integration	4.48	0.38	High
Innovation Acceptance and Intention to Use	4.59	0.34	Very High
Overall LSIAQ	4.56	0.29	Very High

Note. Mean scores were interpreted as follows: 4.51-5.00 = Very High, 3.51-4.50 = High, 2.51-3.50 = Moderate, 1.51-2.50 = Low, and 1.00-1.50 = Very Low.

Overall LSIAQ results were very high (M = 4.56, SD = 0.29). Perceived Learning Usefulness was highest (M = 4.60, SD = 0.31), followed by Learning Satisfaction (M = 4.59, SD = 0.33) and Innovation Acceptance and Intention to Use (M = 4.59, SD = 0.34). Learning Compatibility and Integration was comparatively lower (M = 4.48, SD = 0.38) but remained high. These supplementary results describe favorable post-use perceptions and were not treated as an additional between-group hypothesis test.

DISCUSSION AND CONCLUSION

Effects on English Sentence-Construction Achievement

The AI-supported concept mapping system was associated with significantly stronger sentence-construction achievement than conventional instruction after controlling for baseline scores, supporting H1. This finding is educationally important because sentence construction requires learners to coordinate vocabulary, grammar, word order, and meaning rather than merely recognize isolated linguistic elements.

A plausible explanation is the use of explicit visual relationships in which concept-map links functioned as both semantic and grammatical structures. Relationships such as *Apple → has → seeds* and *Bird → can → fly* may have helped learners connect conceptual meaning with sentence form and develop reusable sentence patterns. This interpretation is consistent with Meaningful Learning Theory and with findings from Kaveh et al. (2023) and Han (2025), who reported positive effects of concept mapping on productive language learning.

The progressive sequence from Explore to Complete the Map and Build a Sentence also provided increasing independence through noticing, retrieval, production, corrective feedback, and revision. Immediate feedback, adaptive review, and contextual GPT assistance provided additional support, consistent with AI-mediated language-learning research (Wei, 2023; Mahapatra, 2024). However, the achievement effect should not be attributed to generative AI alone because the intervention integrated concept mapping, structured practice, feedback, adaptive review, progress monitoring, and contextual AI support.

Effects on English Learning Motivation

The findings also supported H2, with students using the AI-supported concept mapping system demonstrating significantly higher English learning motivation after baseline motivation was controlled. This result is consistent with the ARCS Model of Motivational Design, which conceptualizes motivation in terms of Attention, Relevance, Confidence, and Satisfaction.

Attention may have been supported by interactive concept maps, learner-controlled pronunciation, varied activity modes, immediate feedback, and optional conversational assistance. Relevance may have resulted from connecting familiar vocabulary with practical sentence structures that learners needed for university English communication. Confidence may have developed through progressive task difficulty, opportunities to retry incorrect responses, immediate explanations, and adaptive practice. Satisfaction may have arisen from successfully completing activities, recognizing improvement, and observing visible progress.

The motivational effect should therefore not be attributed simply to the novelty of using generative AI. The ARCS perspective explicitly suggests that technology does not become motivational merely because it is new. More plausibly, motivation resulted from the interaction of visual clarity, meaningful tasks, manageable progression, feedback, learner control, and accessible assistance. This interpretation is consistent with Kaveh et al. (2023), Wei (2023), Huang and Mizumoto (2024), and recent evidence from Thai university learners showing that structured AI-supported English learning may enhance engagement, confidence, personalization, and motivation.

At the same time, previous research has raised concerns regarding inaccurate AI responses, overdependence, academic misconduct, and reduced learner ownership. These concerns strengthen the rationale for the teacher-controlled AI architecture adopted in the present system. GPT was positioned as a scaffold for explanation, modelling, prompting, and corrective assistance rather than as an autonomous agent that generated completed language on behalf of the learner.

Student Satisfaction and Innovation Acceptance

The supplementary post-use evaluation showed a very high overall level of learner satisfaction and innovation acceptance ($M = 4.56$, $SD = 0.29$). Perceived Learning Usefulness was highest ($M = 4.60$, $SD = 0.31$); Learning Satisfaction and Innovation Acceptance and Intention to Use were both $M = 4.59$, while Learning Compatibility and Integration remained high ($M = 4.48$, $SD = 0.38$). These findings complement achievement and motivation by showing that students who used the system also regarded it as useful, satisfying, compatible, and acceptable for continued use.

Perceived usefulness is particularly relevant to technology acceptance because learners are more likely to adopt an innovation when they believe that it contributes to meaningful task performance (Davis, 1989). Higher-education evidence likewise links usefulness, compatibility, satisfaction, and behavioral intention with continued use of ChatGPT-supported learning (Yu et al., 2024). These perceptions should nevertheless be interpreted as post-use self-reports rather than evidence of long-term adoption.

Educational Contribution and Implications

The study contributes a pedagogical architecture that integrates visual concept mapping, explicit grammatical relationships, progressive sentence production, corrective feedback, adaptive review, learner analytics, and contextual generative AI. Labelled concept-map relationships function as grammatical structures that support learners in transforming vocabulary knowledge into meaningful sentences. Importantly, instructional objectives, validated content, and learning activities remain teacher controlled, while AI provides contextual support rather than replacing learner thinking. The LSIAQ further extends the evaluation by incorporating learner satisfaction, perceived usefulness, compatibility, and innovation acceptance.

Limitations and Directions for Future Research

The study was conducted at one university with Thai first-year undergraduates and used one intact class per condition, limiting generalizability and preventing complete separation of class and treatment effects despite ANCOVA adjustment. Satisfaction and innovation acceptance were measured only post-intervention in the experimental group and therefore represent

descriptive perceptions rather than comparative effects or long-term adoption. The six-week intervention also reflects short-term outcomes. Future studies should employ delayed posttests, larger multi-class and multi-university samples, and longitudinal or mixed-methods designs. Dismantling or factorial studies could also examine the individual contributions of concept mapping, adaptive review, corrective feedback, and GPT assistance.

Conclusion

The AI-supported concept mapping system was associated with stronger sentence-construction achievement and learning motivation among Thai first-year undergraduates. Its value lies in integrating visual relationships, structured practice, corrective feedback, adaptive review, and contextual AI within a teacher-controlled environment. Experimental-group students also reported very high satisfaction and innovation acceptance (M = 4.56, SD = 0.29), with perceived learning usefulness rated highest. Overall, the findings support AI as a scaffold that supplements, rather than replaces, learner responsibility and teacher-designed instruction.

REFERENCES

Ausubel, D. P. (1968). Educational psychology: A cognitive view. Holt, Rinehart and Winston.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319-340. https://doi.org/10.2307/249008

Han, X. (2025). Effects of concept mapping on L2 learning: A meta-analysis of studies from 2006 to 2024. Asian-Pacific Journal of Second and Foreign Language Education, 10, Article 26. https://doi.org/10.1186/s40862-025-00331-2

Huang, J., & Mizumoto, A. (2024). Examining the effect of generative AI on students' motivation and writing self-efficacy. Digital Applied Linguistics, 1, 102324. https://doi.org/10.29140/dal.v1.102324

Kaveh, A., Hadipourfard, E., & Bavali, M. (2023). The effects of mobile-assisted concept mapping on EFL learners' writing fluency and motivation. The JALT CALL Journal, 19(2), 171-196. https://doi.org/10.29140/jaltcall.v19n2.1037

Keller, J. M. (1987). Development and use of the ARCS model of instructional design. Journal of Instructional Development, 10(3), 2-10. https://doi.org/10.1007/BF02905780

Keller, J. M. (2010). Motivational design for learning and performance: The ARCS model approach. Springer.

Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S. Y. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. Smart Learning Environments, 11, Article 50. https://doi.org/10.1186/s40561-024-00342-5

Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. Smart Learning Environments, 11, Article 9. https://doi.org/10.1186/s40561-024-00295-9

Mayer, R. E. (2009). Multimedia learning (2nd ed.). Cambridge University Press.

Mayer, R. E. (2021). Cognitive theory of multimedia learning. In R. E. Mayer & L. Fiorella (Eds.), The Cambridge handbook of multimedia learning (3rd ed., pp. 57-72). Cambridge University Press. https://doi.org/10.1017/9781108894333.008

Novak, J. D., & Cañas, A. J. (2008). The theory underlying concept maps and how to construct and use them (Technical Report IHMC CmapTools 2006-01, Rev. 01-2008). Florida Institute for Human and Machine Cognition.

Pawabutra, P., & Sutakote, C. (2024). An analysis of grammatical errors in English speaking of Thai EFL learners. Journal of English Language and Linguistics, 5(3), 297-313. https://doi.org/10.62819/jel.2024.625

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

Sermsook, K., Liamnimitr, J., & Pochakorn, R. (2017). An analysis of errors in written English sentences: A case study of Thai EFL students. English Language Teaching, 10(3), 101-110. https://doi.org/10.5539/elt.v10n3p101

Songsiengchai, S. (2025). Implementation of artificial intelligence (AI): Chat GPT for effective English language learning among Thai students in higher education. International Journal of Education and Literacy Studies, 13(1), 302-312.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), Principle and practice in applied linguistics (pp. 125-144). Oxford University Press.

Thangthong, P., Phiromsombut, J., & Imsa-ard, P. (2024). Navigating AI writing assistance tools: Unveiling the insights of Thai EFL learners. THAITESOL Journal, 37(1), 111-131.

UNESCO. (2023). Guidance for generative AI in education and research. United Nations Educational, Scientific and Cultural Organization.

Wei, L. (2023). Artificial intelligence in language instruction: Impact on English learning achievement, L2 motivation, and self-regulated learning. Frontiers in Psychology, 14, 1261955. https://doi.org/10.3389/fpsyg.2023.1261955

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

Yu, C., Yan, J., & Cai, N. (2024). ChatGPT in higher education: Factors influencing ChatGPT user satisfaction and continued use intention. Frontiers in Education, 9, 1354929. https://doi.org/10.3389/feduc.2024.1354929

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.



Copyright: © 2026 by the authors. This is a fully open-access article distributed under the terms of the Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0).