



The Development of Computer-Assisted Instruction (CAI) for Modal Verbs in English Language Learning for Grade 9 Students

Pongsri Sobpol

Bangsaaiwithaya School, Bangsaai Pranakornsri Ayuthaya, THAILAND

*Corresponding Author E-mail: suppholpongsri@gmail.com

(Article history) Received: 2025-08-21, Revised: 2026-08-22, Accepted: 2026-08-24

Abstract

Background: Modal verbs encode functions such as ability, possibility, permission, obligation, prohibition, advice, requests, and suggestions, and their context-dependent meanings can present difficulties in secondary English learning. This study developed an interactive Computer-Assisted Instruction (CAI) package for Grade 9 students and evaluated learning achievement and learner acceptance. Methods: The source research used a one-group pretest–posttest design with 41 Grade 9 students at Bangsaaiwithaya School in the 2025 academic year. The intervention comprised four functionally organized modal-verb units delivered through a Flash-based multimedia environment. Instruments were a 40-item multiple-choice achievement test and a 20-item five-point satisfaction questionnaire. The report also documented one-to-one, small-group, and field try-outs. Results: The reported pretest mean was 12.45 (SD = 3.21) and the posttest mean was 25.87 (SD = 2.98), with $t = 15.62$ and significance reported at the .05 level. A secondary calculation from these reported aggregates yielded a mean gain of 13.42 points (95% CI 11.68–15.16), $d_z = 2.44$, and normalized gain = 0.49. Overall satisfaction was high ($M = 4.35$, $SD = 0.56$). Efficiency results were internally inconsistent: Chapter 4 reported $E1/E2 = 86.75/82.35$, whereas the appendix reported 79.85/81.77 for the field trial; direct recalculation from the 40 displayed field-trial rows yielded 79.50/81.63. Because the report does not provide evidence sufficient to reconcile these values, no single final efficiency estimate is asserted here. Conclusion: The available evidence supports substantial within-group improvement following CAI use and positive learner acceptance, but the absence of a control group prevents causal claims of superiority over conventional instruction. The study contributes a transparent, curriculum-aligned account of interactive digital grammar instruction and identifies reporting requirements for stronger future evaluations.

Keywords: Computer-Assisted Instruction; Modal verbs; English grammar; Grade 9; Technology-enhanced language learning; Learner achievement; Learner satisfaction

1. INTRODUCTION

English modal verbs are a compact grammatical system with substantial communicative load. Forms such as can, could, may, might, must, shall, should, will, and would can express ability, possibility, permission, obligation, advice, prediction, volition, requests, and interpersonal stance. Because the same form may serve more than one semantic or pragmatic function, mastery requires more than memorizing isolated rules. Descriptions of English modality emphasize the interaction among

grammatical form, speaker stance, context, and degrees of commitment or necessity (Palmer, 2001; Biber et al., 1999). For school-age EFL learners, this many-to-many relation between form and function can create a persistent instructional challenge, particularly when grammar is presented as decontextualized rules rather than as resources for communication.

The Thai Basic Education Core Curriculum B.E. 2551 (A.D. 2008) positions foreign-language learning within communicative goals and expects learners to use language to exchange information, express feelings and opinions, and participate appropriately in real situations. The curriculum therefore creates a pedagogical rationale for teaching grammatical structures not as ends in themselves but as means of communication. The source research responded to this curriculum context by organizing modal-verb instruction around four functional units: ability and possibility; permission and obligation; prohibition and advice; and requests and suggestions. This functional organization is pedagogically relevant because contemporary accounts of modality likewise treat modal auxiliaries as form–meaning pairings whose interpretation depends on construction and context (Palmer, 2001; Torres-Martínez, 2019).

Computer-Assisted Language Learning (CALL) has long offered tools for linking form-focused practice with learner control, multimodal presentation, and immediate response. Early and contemporary CALL scholarship emphasizes that technology is educationally useful when it creates conditions for language processing and practice that are difficult to sustain through static materials alone (Warschauer & Healey, 1998; Chapelle, 2001). However, technology should not be treated as intrinsically effective. Meta-analytic and review evidence indicates an overall positive role for technology-supported language instruction, while simultaneously showing considerable heterogeneity in intervention quality, research design, outcome measures, and evidential strength (Grgurović et al., 2013; Golonka et al., 2014). Consequently, the relevant question is not whether computers improve language learning in the abstract, but how a specific instructional design is structured, how it is evaluated, and how confidently the observed outcomes can be attributed to that design.

The source report developed a Flash-based CAI package for Grade 9 modal-verb instruction, incorporating content explanations, examples, exercises, multimedia presentation, self-paced navigation, and feedback. It also documented an iterative development process consisting of one-to-one, small-group, and field try-outs before the main pretest–posttest evaluation. These features create two potentially important contributions. First, the study provides a school-based example of functionally organized digital grammar instruction in a Thai EFL setting. Second, it contains multiple layers of evidence—formative efficiency data, achievement data, and satisfaction data—that can be analyzed together rather than treated as isolated outcomes.

At the same time, converting the full research report into a journal article requires careful treatment of its internal inconsistencies. The report gives more than one final E1/E2 efficiency value, contains an unresolved discrepancy in field-trial sample size, describes three questionnaire domains in the methods but four in the results, and mentions both Flash and more modern platforms in ways that can blur what was actually implemented. A publication-ready article should not silently choose

among conflicting values. The present manuscript therefore follows a source-constrained approach: findings explicitly reported in the research report are distinguished from secondary calculations that can be reproduced from reported aggregates or appendix-level individual scores, and unresolved inconsistencies are stated directly.

The study had three primary objectives: (1) to develop a CAI lesson for English modal verbs and evaluate it against the report's 80/80 efficiency criterion; (2) to compare Grade 9 students' achievement before and after completing the CAI; and (3) to examine learner satisfaction with the CAI. In addition, the present article conducts only those secondary analyses that are mathematically derivable from the retained data: a confidence interval and standardized effect size for the reported pre-post change, a normalized gain estimate, verification of efficiency calculations shown in the appendix, descriptive unit-level field-trial performance, and a process-posttest correlation using the displayed field-trial rows. No new experiment, participant, outcome, or primary measurement is introduced.

2. CRITICAL LITERATURE REVIEW

2.1 Modal Verbs as a Form-function Learning Problem

Modality concerns how speakers qualify propositions in relation to possibility, necessity, commitment, permission, obligation, prediction, and related meanings. Palmer (2001) describes modality as a grammatical category that is realized through multiple systems, including modal verbs. Corpus-based grammar similarly shows that central modals differ in frequency, register distribution, and pragmatic function (Biber et al., 1999). From a pedagogical perspective, these properties mean that a learner cannot reliably map one modal verb onto one fixed meaning. Can and could, for example, participate in ability and permission meanings; may and might are associated with possibility but also interact with interpersonal meanings; must and should can express different strengths and sources of obligation or advice. A functionally organized curriculum can therefore help learners attend to semantic contrasts that are easily obscured when instruction is organized only by lexical item.

Torres-Martínez (2019) argues that English modal auxiliaries are better understood through patterned form-meaning relations than through isolated lexical definitions. Although the source study did not test a construction-grammar model, its four-unit organization is compatible with the broader pedagogical principle that modal meaning should be linked to communicative function and context. This is particularly pertinent in lower-secondary EFL, where students must move from recognition of forms toward contextually appropriate use.

2.2 Computer-assisted Language Learning and Grammar Instruction

CALL has evolved from tutorial and drill-based applications toward networked, adaptive, mobile, and data-rich environments, but tutorial CALL remains relevant for structured grammar practice when it provides meaningful interaction, feedback, and learner control. Warschauer and Healey (1998) characterized the development of CALL as a movement from behavioristic drill toward more integrative uses of

computers. Chapelle (2001) subsequently emphasized that CALL evaluation should connect technological features to language-learning processes rather than assume that digital delivery itself produces learning.

Evidence syntheses provide a cautiously positive picture. Grgurović et al. (2013), reviewing 37 studies and 52 effect sizes, found a small but positive overall advantage for technology-supported second- and foreign-language pedagogy, with stronger results in more rigorous designs. Golonka et al. (2014), reviewing more than 350 studies across technology types, concluded that the volume of technology research exceeds the amount of strong efficacy evidence, highlighting the importance of design quality and appropriate comparison. These reviews support the relevance of CAI while also cautioning against causal overstatement from weak designs.

Research on technology-mediated grammar learning also suggests that specific affordances matter. Sagarra and Zapata (2008) reported positive learner perceptions of an online workbook and improvements in grammar-related assessment, with accessibility, user-friendliness, and immediate error feedback among valued features. Ai (2017) showed that computer-mediated corrective feedback can be designed to support learner uptake and positive perceptions, although the intervention, target language, and feedback architecture differed substantially from those in the present study. At the secondary-school level, Moon and Oh (2018) demonstrated that technology-supported, data-driven grammar work could yield both cognitive and affective benefits. Collectively, these studies suggest that engagement, feedback, and structured practice are plausible mechanisms, but they do not establish that any multimedia grammar package will be effective in every context.

2.3 Multimedia, Feedback, and Learner Control

The source CAI combined text, sound, animation, exercises, and feedback. Multimedia-learning research provides a framework for understanding why such combinations can help when they are designed to guide rather than overload attention. Mayer (2021) argues that people can learn more deeply from appropriately coordinated verbal and visual representations than from words alone, while Clark and Mayer (2016) stress that multimedia elements should serve instructional processing rather than decoration. This distinction is directly relevant to the source report, because early try-outs led to revisions in graphic relevance, animation timing, audio clarity, and explanatory wording. Those revisions are consistent with reducing extraneous processing, although the study did not experimentally isolate or test cognitive-load mechanisms.

Learner control is another important design feature. The report describes the CAI as self-paced, allowing repeated access to explanations and exercises. CALL research has long treated learner control and individualized pacing as potential advantages, but their effectiveness depends on learners' ability to use them productively. Similarly, immediate feedback may support correction and retrieval, but the optimal specificity and sequencing of feedback remain empirical questions. The source study's formative revisions therefore constitute an instructional-development contribution, but not an ablation test of which component was responsible for achievement gains.

2.4 Research Gap and Article Contribution

The gap addressed by this article is narrower than the general question of whether educational technology improves language learning. First, there is value in documenting how a functionally organized modal-verb CAI performs in an authentic Thai Grade 9 context. Second, many school-based reports emphasize statistical significance without reporting effect magnitude, uncertainty, or data-integrity checks. Third, the source report includes formative try-out tables that allow partial verification of the stated efficiency values but were not fully exploited in the original interpretation. The present article therefore contributes a source-faithful synthesis that combines learning outcomes, learner acceptance, iterative design evidence, and transparent secondary verification while explicitly limiting causal claims.

3. MATERIALS AND METHODS

3.1 Study Design and Setting

The research report characterizes the study as experimental research using a One Group Pretest–Posttest Design. Students completed a pretest, learned through the CAI, completed a posttest, and then responded to a satisfaction questionnaire. Because no concurrent control or comparison group was included, the design is more precisely interpreted as a single-group pre–post intervention evaluation. Such a design can estimate within-group change but cannot rule out alternative explanations such as testing effects, maturation, history, or other time-linked influences (Shadish et al., 2002). The study was conducted at Bang Sai Witthaya School, Phra Nakhon Si Ayutthaya, Thailand, during the 2025 academic year.

3.2 Participants and Sample-size Reporting

The main analysis reports 41 Grade 9 students, selected by cluster sampling from a Grade 9/1 classroom. The methodology section also states that the Grade 9 population comprised 41 students, which makes the population–sample relationship unclear: if the entire population was 41, cluster sampling would not reduce the population. The appendix creates a second inconsistency by describing a field try-out with 41 students who were not the main sample, while the displayed field-trial table contains 40 numbered cases. These inconsistencies cannot be resolved from the available report and are not silently corrected in this manuscript. The main achievement and satisfaction analyses are therefore reported with $n = 41$ because that number is explicitly attached to those results; secondary field-trial calculations use $n = 40$ because 40 individual rows are visible in the appendix.

3.3 CAI Intervention and Instructional Content

The CAI focused on nine modal verbs: can, could, may, might, must, shall, should, will, and would. It addressed meaning and function, affirmative/negative/interrogative structures, examples, practice, and assessment. Content was organized into four functional units: Unit 1, ability and possibility; Unit 2, permission and obligation; Unit 3, prohibition and advice; and Unit 4, requests and

suggestions. The report describes interactive, self-paced learning with text, sound, animation, exercises, and immediate feedback. The source definition states that the implemented lesson was delivered through Flash. A later methods passage mentions “Flash (or Adobe Animate/HTML5),” and the discussion recommends migration to HTML5 or web/mobile formats. Because no evidence shows that Adobe Animate or HTML5 was used in the completed experiment, this article treats Flash as the implemented platform and HTML5/Adobe Animate only as future or transferable alternatives.

Instructional development was informed by the general sequence of attention, structured content, guided practice, assessment, and feedback associated with systematic instructional design (Gagné et al., 2005). However, the original report cited an earlier Gagné and Briggs formulation and did not provide a detailed screen-by-screen instructional script. The article therefore describes only the components explicitly reported rather than inferring a complete nine-event implementation.

3.4 Iterative Development and Try-out Phases

Three formative stages were documented. In the one-to-one try-out, three students completed four during-lesson components and a posttest. The report then revised menu instructions, guidance, graphics, animation–audio alignment, motion timing, and explanatory wording. The small-group try-out involved nine students; subsequent revisions targeted graphic–content correspondence, narration clarity, font size, and visual salience. A field try-out followed before the main evaluation. The report’s field-trial narrative states 41 participants, but the appendix table displays 40 cases. These stages were formative evaluations of successive versions rather than randomized comparisons; therefore, changes in efficiency across stages cannot be attributed causally to any single revision.

3.5 Instruments

The achievement test was a 40-item four-option multiple-choice instrument administered before and after the intervention. Content validity was reviewed by three experts using an Item–Objective Congruence criterion of at least 0.50. The methods state that KR-20 reliability was used, but the numerical KR-20 coefficient is not reported in the available results; no reliability value is therefore supplied here.

The satisfaction questionnaire contained 20 items rated on a five-point Likert scale. The methods describe three domains—content, presentation, and usability—whereas the results report four domains: content and accuracy, presentation style, ease of use, and interest/engagement. Because the instrument itself is not reproduced item-by-item in the available report, the domain structure cannot be independently reconciled. The manuscript retains the four-domain result table because those numerical outcomes are explicitly reported, while noting the discrepancy.

3.6 Data Collection Procedure

For the main evaluation, students first completed the 40-item pretest, then learned with the CAI, completed the 40-item posttest, and finally completed the

satisfaction questionnaire. The report does not state the intervention duration, number of sessions, device specifications, operating system, or whether students worked individually or in a supervised laboratory. These details are therefore not inferred.

3.7 Primary Analysis and Evaluation Metrics

The report specified three primary analytic procedures: E1/E2 efficiency analysis, a dependent-samples t-test for pretest–posttest achievement, and descriptive means and standard deviations for satisfaction. E1 represents average process performance during the CAI relative to the maximum available process score, and E2 represents average posttest performance relative to the maximum posttest score. The target criterion was stated as 80/80. The report did not define an explicit tolerance rule for values slightly below 80, which is relevant because one appendix value was reported as 79.85 for E1 but described as meeting the criterion.

3.8 Secondary Analyses

No new experiment was conducted. Secondary analyses were restricted to quantities derivable from the report. From the main pretest/posttest aggregates, the mean gain was calculated as posttest minus pretest. The standard error of the paired difference was recovered from the reported t statistic, allowing a 95% confidence interval for the mean gain. The standardized paired effect size was calculated as $d_z = t/\sqrt{n}$, following standard effect-size reporting principles (Lakens, 2013). A normalized gain was calculated as $(\text{posttest mean} - \text{pretest mean})/(\text{maximum possible score} - \text{pretest mean})$, using the reported maximum of 40.

The appendix-level try-out tables were independently recomputed from displayed scores. For the 40 visible field-trial cases, unit means, total process mean, posttest mean, E1, and E2 were recalculated. Pearson's correlation was also computed between total process score and posttest score. These analyses are explicitly labeled “secondary” because they were not reported as inferential results in the original research report. No secondary analysis was performed where the necessary individual-level data were unavailable; in particular, the main 41-student pretest and posttest raw scores were not present in the reviewed materials, preventing distributional diagnostics or non-parametric paired sensitivity testing.

3.9 Machine-learning-specific Reporting Items and Reproducibility

The study was an educational intervention, not a machine-learning classification or prediction study. Consequently, there were no class labels or class distributions, data preprocessing pipeline, predictive model architecture, training/validation/test split, cross-validation, ablation study, calibration analysis, or unseen-dataset evaluation. No computational hardware specifications were reported. These items are therefore not applicable and are not fabricated to satisfy a generic computational template. Reproducibility is instead supported by reporting the available scoring rules, sample sizes, aggregate statistics, and secondary formulas, and by separating source-reported values from values recalculated directly from appendix rows.

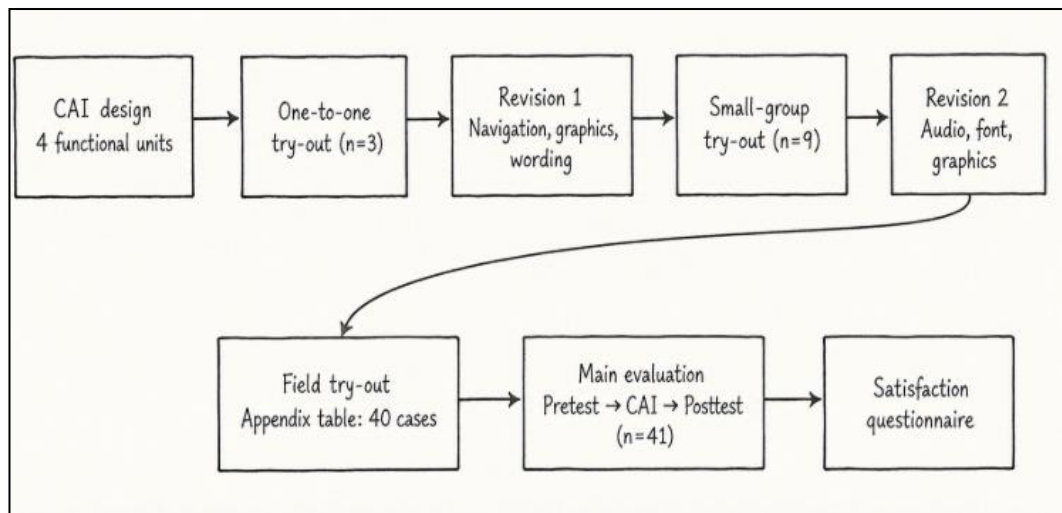


Figure 1: CAI Development and Evaluation Workflow

4. RESULTS

4.1 Data-stream Verification and Unresolved Inconsistencies

The report contains five analytically distinct data streams: one-to-one try-out scores, small-group try-out scores, field-trial process/posttest scores, main pretest/posttest achievement aggregates, and satisfaction aggregates. Recalculation shows that some efficiency summaries do not exactly match the displayed appendix rows. Table 1 therefore distinguishes source-reported values from independently reproducible values before substantive interpretation.

Table 1: Data Streams Available

Data stream	Cases shown/reported	Variables available	Role in this article
One-to-one try-out	3 displayed cases	Four process components + posttest	Formative development; E1/E2 verification
Small-group try-out	9 displayed cases	Four process components + posttest	Formative development; E1/E2 verification
Field try-out	40 displayed cases; narrative states 41	Four process components + posttest	Secondary row-level reanalysis
Main achievement study	41 reported participants	Pretest and posttest aggregate mean/SD; t	Primary learning-outcome analysis
Satisfaction study	41 implied by main sample	Overall and domain means/SDs	Primary learner-acceptance analysis

Note. "Displayed cases" refers to individual rows visible in the appendix. The main-study raw pretest/posttest scores were not available in the reviewed report.

Table 2: Verification of Reported CAI Efficiency Values

Stage/source	Source-reported E1/E2	Recalculated from displayed rows	Interpretation
One-to-one try-out	65.83 / 75.00	65.83 / 75.00	Exact agreement.
Small-group try-out	72.22 / 77.44	72.22 / 77.50	E2 differs by 0.06 percentage points.
Field try-out appendix	79.85 / 81.77	79.50 / 81.63	Both E1 and E2 differ from displayed-row recalculation; 40 rows are shown.
Chapter 4 final efficiency	86.75 / 82.35	Not independently verifiable from an accompanying row-level table	Conflicts with the appendix field-trial efficiency; no evidence in the report resolves the difference.

Note: Recalculated E1 and E2 use the displayed integer scores and a maximum of 40. Because the report contains incompatible final-efficiency values

The one-to-one value is internally consistent. The small-group E2 summary differs slightly from the displayed data. More importantly, the 40 field-trial rows yield an average process score of 31.80/40 (79.50%) and an average posttest score of 32.65/40 (81.63%), whereas the appendix states 79.85/81.77. Chapter 4 separately states 86.75/82.35. The latter value cannot be reconstructed from the appendix field table and is therefore retained only as a source-reported figure, not as an independently verified estimate. This handling follows the user-specified requirement to identify unresolved inconsistencies rather than choose an unsupported value.

4.2 Main Pretest–Posttest Achievement

The main study reported a pretest mean of 12.45 (SD = 3.21) and a posttest mean of 25.87 (SD = 2.98) for 41 students. The reported dependent-samples *t* statistic was 15.62, with significance reported as .000 and interpreted at the .05 level. The raw mean increase was therefore 13.42 points on a 40-point scale.

Table 3: Main Achievement Results and Source-derivable Secondary Statistics

Statistic	Pretest	Posttest / change	Status
Mean score	12.45	25.87	Reported
Standard deviation	3.21	2.98	Reported
Scale percentage	31.13%	64.68%	Secondary descriptive calculation
Mean paired gain	—	13.42 points	Secondary calculation from reported means
95% CI for mean gain	—	11.68 to 15.16	Secondary calculation from reported <i>t</i> and <i>n</i>
Dependent-samples <i>t</i>	—	<i>t</i> (40) = 15.62	Reported <i>t</i> ; <i>df</i> derived from <i>n</i> = 41
Standardized paired effect	—	<i>d</i> _z = 2.44	Secondary calculation: $t/\sqrt{n}$
Normalized gain	—	<i>g</i> = 0.49	Secondary calculation using maximum score = 40

Note: Secondary quantities were calculated only from reported aggregates; no participant-level main-study scores were available. The large standardized effect describes within-group pre–post change and must not be interpreted as a controlled treatment effect.

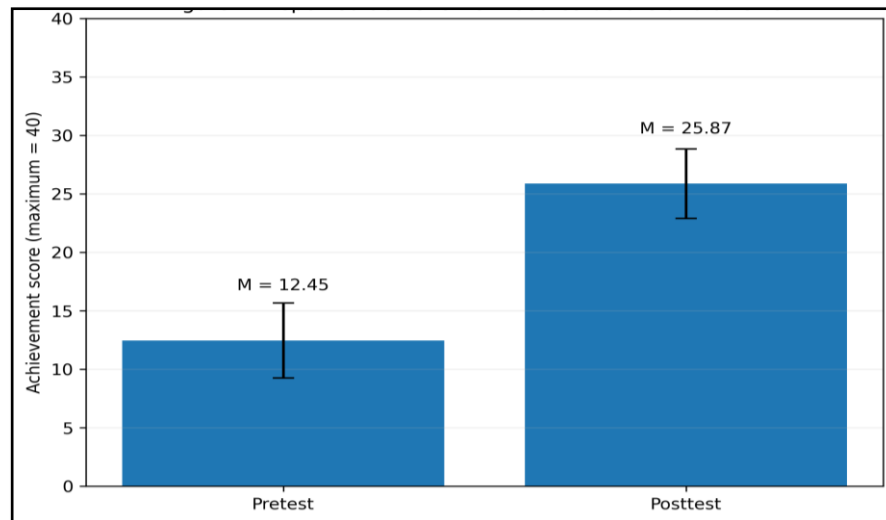


Figure 2: The Achievement Scores Before & After Using CAI

The secondary effect-size calculation indicates that the pre–post shift was large relative to the standard deviation of the paired differences. However, because the study had no control group, this effect size quantifies change within the observed cohort rather than the causal effect of CAI relative to alternative instruction. The normalized gain of 0.49 likewise describes the proportion of the available score improvement realized at the group-mean level; it is not a comparative treatment effect.

4.3 Student Satisfaction

Overall satisfaction was reported as high, with a mean of 4.35 and SD of 0.56 on a five-point scale. Domain-level results were also uniformly high. Content and accuracy received the highest mean (4.40), followed by interest and engagement (4.38), presentation style (4.33), and ease of use (4.28).

Table 4: Reported learner satisfaction with the CAI

Dimension	Mean	SD	Reported level
Content and accuracy	4.40	0.52	High
Presentation style	4.33	0.60	High
Ease of use	4.28	0.57	High
Interest and engagement	4.38	0.55	High
Overall	4.35	0.56	High

Note: The methods describe three questionnaire domains, whereas the results report four. The numerical results above reproduce the four-domain outcome table without inferring the underlying item allocation.

4.4 Secondary Reanalysis of the Field Try-out

The appendix provides sufficient individual-level field-trial data for a limited secondary analysis of 40 displayed cases. Mean unit performance decreased across the four units: Unit 1 = 8.50/10 (85.00%), Unit 2 = 8.20/10 (82.00%), Unit 3 = 7.73/10 (77.25%), and Unit 4 = 7.38/10 (73.75%). Because these were repeated components administered within the same formative trial and no inferential unit-comparison procedure was specified in the report, the pattern is interpreted descriptively only. It

suggests that later or functionally different content areas may have been more demanding, but the available data do not identify why.

Table 5: Secondary Descriptive Analysis of the 40 Displayed Field-trial Cases

Measure	Mean	SD	Percentage of maximum / association
Unit 1 process score	8.50	0.72	85.00%
Unit 2 process score	8.20	0.72	82.00%
Unit 3 process score	7.73	0.72	77.25%
Unit 4 process score	7.38	0.49	73.75%
Total process score	31.80	1.51	79.50% (= recalculated E1)
Field posttest score	32.65	1.17	81.63% (= recalculated E2)
Process–posttest correlation	—	—	$r = .616$; 95% CI [.377, .778]; $p < .001$

Note: All values in this table are secondary calculations from the 40 individual field-trial rows displayed in the appendix. They are not new primary data. The correlation is descriptive of the formative field sample and does not establish causality.

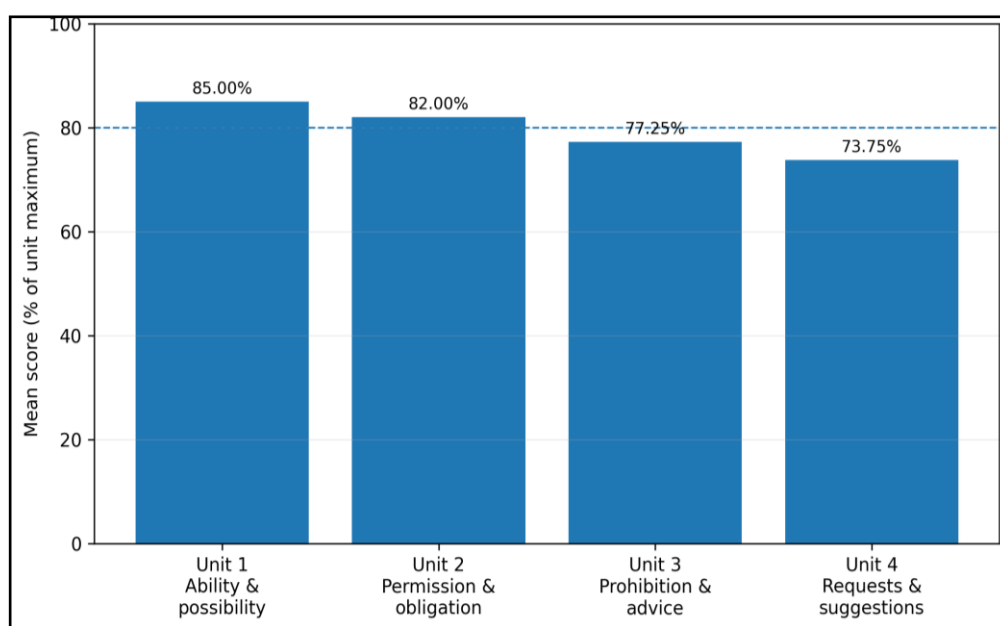


Figure 3: Secondary Re-analysis of Field-try-out Unit Performance

Process performance and field-trial posttest performance were moderately to strongly positively associated ($r = .616$, $p < .001$; secondary analysis), with approximately 38% shared variance ($r^2 = .379$). This association is compatible with the idea that students who performed better during CAI practice also tended to perform better on the immediate posttest. It does not determine whether process performance caused posttest performance, because both measures may reflect common underlying proficiency or engagement.

4.5 Comparisons, Ablation, Cross-validation, and Unseen-data Evaluation

The only comparison supported by the main study is the within-group pretest–posttest comparison. The report contains no conventional-teaching control group,

randomized comparator, alternative CAI version, or between-group benchmark. Likewise, no ablation analysis was performed to isolate the contribution of animation, audio, self-pacing, feedback, or any individual modal-verb unit. Because the intervention was not a predictive computational model, cross-validation, calibration, and unseen-dataset evaluation are not applicable. Reporting such analyses as though they existed would introduce fabricated evidence; accordingly, they are explicitly absent from the results.

5. DISCUSSION

5.1 Interpretation of Learning Achievement

The clearest empirical finding is the large increase in the cohort's mean achievement score after CAI use. The mean rose from 12.45 to 25.87 out of 40, and the reported paired *t* statistic was 15.62. The secondary *d_z* estimate of 2.44 indicates that the observed mean change was large relative to the variability of within-person differences. This is substantively meaningful as a description of learning change, but it should not be interpreted as evidence that CAI was superior to conventional instruction. In a one-group pre–post design, the intervention is temporally associated with the improvement, yet other factors cannot be excluded. A high-impact journal presentation therefore benefits from replacing causal language such as “CAI caused superior achievement” with the more defensible conclusion that achievement improved substantially following the CAI experience.

The magnitude of the score change is consistent with the pedagogical architecture of the intervention. Students received focused exposure to a limited grammatical domain, repeated practice, and immediate post-instruction assessment. Such conditions can support short-term acquisition through repeated retrieval, immediate feedback, and multiple representations. CALL literature supports these mechanisms in principle: technology-mediated environments can increase opportunities for practice and feedback, while the effectiveness of those opportunities depends on how they are designed and integrated (Chapelle, 2001; Sagarra & Zapata, 2008; Ai, 2017). Nevertheless, because the study did not include delayed retention measures, the results concern immediate learning achievement rather than durable acquisition.

5.2 Learner Acceptance and the Role of Multimedia Design

The satisfaction findings complement the achievement results. Students evaluated all reported dimensions positively, with content/accuracy and interest/engagement receiving the highest means. The pattern is compatible with the study's design emphasis on multimedia presentation and interaction. Importantly, the formative try-outs did not simply add more multimedia; they led to refinement of navigation, graphic relevance, audio clarity, text concision, and font readability. This distinction matters because multimedia-learning theory does not predict that more visual or auditory material is always beneficial. Learning is more likely to benefit when representational elements are coordinated and when unnecessary processing is reduced (Mayer, 2021; Clark & Mayer, 2016).

The satisfaction pattern also resembles findings from computer-assisted L2 learning studies in which learners value accessibility, usability, and rapid feedback (Sagarra & Zapata, 2008). Yet satisfaction should not be treated as a proxy for learning. A lesson can be enjoyable without producing achievement, and a demanding lesson can produce learning despite lower satisfaction. The present study is stronger because both outcome types moved in a favorable direction, but the correlation between main-study satisfaction and individual learning gains cannot be tested because participant-level linked data were not available.

5.3 Iterative Development as an Underused Contribution

One of the more distinctive features of the report is its iterative development process. The initial one-to-one try-out fell clearly below the 80/80 target. Revisions were made, followed by a higher small-group result and then a higher field-trial result. Although the stages used different participants and therefore do not constitute a controlled longitudinal improvement test, the sequence documents a practical cycle of prototype evaluation, diagnosis, revision, and retesting. In design terms, this is more informative than presenting only the final software package because it identifies concrete usability and instructional issues that emerged during development.

The secondary field-trial reanalysis adds a further layer. Process scores were positively correlated with field posttest scores, and performance was descriptively lower in Units 3 and 4 than in Units 1 and 2. This pattern could inform future revision priorities, particularly for prohibition/advice and requests/suggestions. However, the current evidence cannot determine whether the lower unit means reflect greater conceptual difficulty, item difficulty, instructional sequencing, or differences in scoring. Item-level psychometric analysis would be useful if the original response data remain available.

5.4 Relationship to Prior CALL Evidence

The present findings fit within, but should not be overstated relative to, the broader CALL evidence base. Grgurović et al. (2013) found an overall positive effect of technology-supported pedagogy but emphasized differences in research design and instructional conditions. Golonka et al. (2014) similarly warned that a large literature does not automatically translate into strong causal evidence. The current study exemplifies both sides of that literature: it provides encouraging achievement and acceptance evidence in a real school context, but its single-group design and incomplete reporting of reliability and implementation details limit causal and generalizable conclusions.

Grammar-specific technology research further supports the plausibility of the observed pattern. Sagarra and Zapata (2008) linked positive attitudes toward online language work with grammar gains and valued immediate feedback. Moon and Oh (2018) reported cognitive and affective benefits from technology-supported grammar work among secondary EFL learners, though their data-driven approach differed markedly from tutorial CAI. Heift and Vyatkina (2017) review multiple technology-based approaches to L2 grammar and underscore that tutorial CALL, intelligent CALL, data-driven learning, and computer-mediated communication offer different affordances. The present CAI belongs most clearly to the tutorial/multimedia tradition

and should be evaluated on that basis rather than compared directly with adaptive or intelligent systems that were not implemented here.

5.5 Modal-verb Pedagogy and Curriculum Alignment

The study's organization around communicative functions is theoretically important. Modal verbs are not simply a list of irregular auxiliaries; they encode stance and interpersonal relations. Organizing practice around ability/possibility, permission/obligation, prohibition/advice, and requests/suggestions creates a bridge between grammar and communicative intention. This organization aligns with the Thai curriculum's emphasis on using language in context and with linguistic accounts of modality as a system of meaning rather than a set of isolated word forms (Ministry of Education, 2008; Palmer, 2001). Future versions could strengthen this contribution by adding contextualized production tasks and by separating epistemic, deontic, and dynamic meanings where appropriate for the learners' level.

5.6 Data Integrity and Research-reporting Lessons

A central methodological finding of the present reconstruction is that the report's efficiency statistics are not fully internally consistent. Transparent reporting requires acknowledging this rather than selecting the most favorable value. The inconsistency does not invalidate the independently reported main pre-post achievement and satisfaction results, but it weakens any strong claim that a particular final E1/E2 value has been verified. For future studies, raw scoring tables, exact sample membership for each development stage, instrument reliability coefficients, intervention duration, and a clearly defined 80/80 decision rule should be preserved alongside the manuscript. Such practices would improve auditability and facilitate cumulative evidence synthesis.

6. CONTRIBUTIONS

6.1 Theoretical Contribution

The study contributes to grammar pedagogy by operationalizing English modality through communicative functions rather than treating modal verbs only as isolated forms. The results are consistent with the proposition that technology-supported grammar practice can connect explicit form-focused learning with contextualized meaning, although the study did not experimentally compare alternative theoretical organizations of the content.

6.2 Methodological contribution

Methodologically, the article demonstrates how a school-based CAI report can be strengthened without collecting new data: effect magnitude and confidence intervals can be derived from reported paired statistics; appendix tables can be recalculated for integrity checks; formative-development data can be analyzed separately from main-study outcomes; and causal language can be aligned with the actual design. The explicit

separation of reported findings and secondary calculations improves reproducibility and reduces the risk of retrospective overclaiming.

Computational and Digital-Instruction Contribution

The contribution is digital-instructional rather than machine-learning computational. The CAI implemented an interactive multimedia lesson with self-paced navigation, practice, feedback, and assessment. No predictive model, adaptive algorithm, artificial-intelligence component, or automated decision-support engine was reported. The computational contribution therefore lies in the structured delivery of instructional content through an interactive computer environment and in the transferable design logic that could be reimplemented using current web technologies. Any claim of AI, personalization, or intelligent adaptation would exceed the evidence.

Practical Contribution

Practically, the CAI offers a supplementary route for repeated grammar explanation and practice, potentially allowing classroom time to be used for discussion, group work, or communicative application. The report itself recommends combining the CAI with group activities, discussion, and conversational practice. Its iterative revisions also provide concrete design lessons: instructions should be unambiguous, graphics should correspond to content, narration should be clear, text should be concise, and typography should support readability.

Limitations and Future Research

The most important limitation is the one-group pretest–posttest design. Without a control or comparison group, the observed learning gain cannot be attributed exclusively to CAI. Future evaluations should use a randomized or well-matched quasi-experimental comparison where feasible, with identical pretest/posttest timing and transparent reporting of instructional exposure.

Second, the study was conducted in a single school with a small Grade 9 cohort and one grammar topic. Generalization to other schools, proficiency levels, age groups, or grammar domains is therefore uncertain. Replication across classes and schools would improve external validity.

Third, the main-study raw pretest/posttest data were not available in the reviewed report. Retaining participant-level scores would permit distributional checks, robust/non-parametric sensitivity analyses, learner-level gain analysis, and exploration of whether lower- and higher-proficiency learners benefit differently.

Fourth, instrument reporting is incomplete. Although IOC procedures, KR-20, and Cronbach's alpha are mentioned, numerical reliability coefficients are not reported. The satisfaction domain structure also differs between the methods and results. Future work should provide the complete instrument blueprint, item-to-objective mapping, item difficulty and discrimination where available, reliability estimates with confidence intervals, and the exact allocation of questionnaire items to domains.

Fifth, the technology platform is obsolete. Flash is no longer a practical deployment environment for contemporary browsers. The report itself recommends migration to HTML5 or mobile/web applications. Such migration should preserve the tested instructional structure while allowing accessibility, device compatibility, analytics, and security to be evaluated afresh. A new platform implementation should not be assumed equivalent to the original CAI without usability and learning validation.

Sixth, the study measured immediate post-instruction achievement and satisfaction but did not report delayed retention or transfer to communicative production. Future research should add delayed posttests and performance tasks requiring students to choose and use modal verbs in contextualized speaking or writing.

Finally, the iterative development stages should be documented with stable version numbers and separate participant rosters so that efficiency changes can be traced accurately and the sample-size inconsistencies identified in the present report do not recur.

7. CONCLUSION

This study developed a functionally organized, interactive CAI package for Grade 9 English modal verbs and evaluated it in a Thai secondary-school context. The main evidence shows a substantial increase in achievement from pretest to posttest and high learner satisfaction. Secondary analysis of the reported aggregates yielded a mean gain of 13.42 points, a 95% confidence interval of 11.68–15.16, $d_z = 2.44$, and normalized gain = 0.49. The appendix also documents a meaningful iterative design process and, in a secondary reanalysis of 40 displayed field-trial cases, a positive association between during-lesson process performance and posttest performance.

The study's contribution is strongest when framed as evidence of substantial within-group improvement and positive learner acceptance following a carefully refined CAI intervention. It should not be framed as proof that CAI is superior to conventional teaching because no control group was included. Equally important, the report contains unresolved inconsistencies in efficiency values and field-trial sample reporting. By making those discrepancies explicit rather than selecting an unsupported value, the present manuscript preserves the evidential boundary of the completed research. The resulting account supports CAI as a promising supplementary approach to modal-verb instruction while identifying the design, reporting, and replication steps needed for stronger future evidence.

REFERENCES

- Ai, H. (2017). Providing graduated corrective feedback in an intelligent computer-assisted language learning environment. *ReCALL*, 29(3), 313–334. <https://doi.org/10.1017/S095834401700012X>
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Longman.

- Chapelle, C. A. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing and research*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139524681>
- Clark, R. C., & Mayer, R. E. (2016). *e-Learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning* (4th ed.). Wiley. <https://doi.org/10.1002/9781119239086>
- Gagné, R. M., Wager, W. W., Golas, K. C., & Keller, J. M. (2005). *Principles of instructional design* (5th ed.). Wadsworth/Thomson Learning.
- Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning*, 27(1), 70–105. <https://doi.org/10.1080/09588221.2012.700315>
- Grgurović, M., Chapelle, C. A., & Shelley, M. C. (2013). A meta-analysis of effectiveness studies on computer technology-supported language learning. *ReCALL*, 25(2), 165–198. <https://doi.org/10.1017/S0958344013000013>
- Heift, T., & Vyatkina, N. (2017). Technologies for teaching and learning L2 grammar. In C. A. Chapelle & S. Sauro (Eds.), *The handbook of technology and second language teaching and learning* (pp. 26–44). Wiley Blackwell. <https://doi.org/10.1002/9781118914069.ch3>
- Hockly, N. (2013). Mobile learning. *ELT Journal*, 67(1), 80–84. <https://doi.org/10.1093/elt/ccs064>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 863. <https://doi.org/10.3389/fpsyg.2013.00863>
- Levy, M. (1997). *Computer-assisted language learning: Context and conceptualization*. Oxford University Press.
- Mayer, R. E. (2021). *Multimedia learning* (3rd ed.). Cambridge University Press.
- Ministry of Education, Thailand. (2008). *Basic Education Core Curriculum B.E. 2551 (A.D. 2008)*. Office of the Basic Education Commission.
- Moon, S., & Oh, S.-Y. (2018). Unlearning overgenerated be through data-driven learning in the secondary EFL classroom. *ReCALL*, 30(1), 48–67. <https://doi.org/10.1017/S0958344017000246>
- Nagata, N. (1996). Computer vs. workbook instruction in second language acquisition. *CALICO Journal*, 14(1), 53–75.
- Palmer, F. R. (2001). *Mood and modality* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139167178>

- Sagarra, N., & Zapata, G. C. (2008). Blending classroom instruction with online homework: A study of student perceptions of computer-assisted L2 learning. *ReCALL*, 20(2), 208–224. <https://doi.org/10.1017/S0958344008000621>
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Torres-Martínez, S. (2019). Taming English modals: How a Construction Grammar approach helps to understand modal verbs. *English Today*, 35(2), 50–57. <https://doi.org/10.1017/S0266078418000081>
- Trapsaphon, P. (2025). The development of computer-assisted instruction for Modal Verbs for Grade 9 students [Unpublished research report]. Bang Sai Witthaya School.
- Warschauer, M., & Healey, D. (1998). Computers and language learning: An overview. *Language Teaching*, 31(2), 57–71. <https://doi.org/10.1017/S0261444800012970>