

Integrating Enthymemic Pedagogy to Enhance Critical Thinking and Oral Communication Skills in Bachelor of Elementary Education at Suan Dusit University, Suphanburi Campus

Rechele Ballovar Ella

Suan Dusit University, Suphanburi Campus, Thailand

Corresponding Author Email: rechele_ela@dusit.ac.th

DOI Link: <http://dx.doi.org/10.6007/IJARPED/v15-i4/29284>

Published Online: 26 October 2026

Abstract

This study examined whether integrating enthymemic pedagogy into an English oral communication course could strengthen critical thinking and oral communication among pre-service elementary teachers at a Thai university. An embedded mixed-methods, one-group pretest–posttest design was used with one intact class of approximately 50 students. Data came from a 30-item knowledge test (RI-1), an oral communication rubric (RI-2), a classroom observation checklist (RI-3), and weekly reflective journals (RI-4). Knowledge scores did not improve significantly, $t(36) = -0.12$, $p = .902$, and observed engagement was not positively related to score change. End-of-cycle oral performance was rated Good by a single colleague-rater ($M = 3.95/5.00$), while journal analysis indicated prompted reflective uptake among submitters. However, the four instruments captured different, partially overlapping subsets of the class: only 4 of the 37 students in the matched panel appeared in all four. Given the one-group design, partially overlapping analytic samples, and measurement limitations, the findings do not support a causal interpretation of the intervention. Its principal contribution is methodological. It shows how instrument-specific visibility shapes what teacher-led classroom research can claim, and it offers a simple presence/absence audit and design recommendations for classroom researchers.

Keywords: Enthymemic Pedagogy, Critical Thinking, Oral Communication, Teacher Education, Classroom Research, Instrument Visibility

Introduction

English communication remains an important concern in Thai higher education, particularly in teacher education, where future teachers are expected not only to speak English clearly but also to explain ideas, evaluate classroom situations, and justify professional decisions. For pre-service teachers, oral communication is therefore both a language outcome and part of professional preparation. This challenge is especially relevant in Thailand, where limited confidence can constrain oral English performance; Saengbanchong and Saito (2020),

for example, report that Thai pre-service teachers often experience difficulty in oral communication because they lack confidence.

Beyond confidence, the instructional problem is also one of reasoning. Students may participate in oral communication tasks yet give responses that remain descriptive, hesitant, or weakly supported by reasons and evidence. They may state opinions without making the assumptions behind those opinions visible. This gap matters in teacher education because future teachers must be able to explain and defend ideas, examine alternatives, and make reasoned decisions in educational situations. It therefore creates a need for an approach that connects spoken communication with explicit reasoning rather than treating speaking and critical thinking as separate classroom outcomes.

Enthymemic pedagogy was selected to address this need because it connects reasoning directly to communication. In Aristotelian rhetoric, an enthymeme is a rhetorical syllogism in which one premise is left unstated and must be supplied by the audience (Aristotle, 2007; Tindale, 2021). In the classroom, this structure can help students identify claims, reasons, hidden assumptions, and evidence, encouraging them to move from simply giving answers to examining the reasoning that supports those answers. The approach is consistent with constructivist learning theory, which treats learners as active makers of meaning, and with critical pedagogy, which values dialogue and questioning (Freire, 2020). However, direct applications of enthymemic pedagogy in Thai pre-service teacher education remain limited, leaving both a pedagogical and an empirical gap in how reasoning-focused speaking instruction is understood in this context.

This article addresses that gap through a classroom study of enthymemic pedagogy in the course English Oral Communication in Educational Contexts in a Bachelor of Elementary Education program at Suan Dusit University, Suphanburi Campus, Thailand. The study was conducted during the first semester of Academic Year 2025, from June to October 2025. Its novelty lies in examining enthymemic reasoning not only as a speaking pedagogy but also through multiple classroom evidence sources, coupled with an explicit audit of whether those instruments captured the same students. The study contributes to the social sciences by connecting language education, critical-thinking pedagogy, and classroom research methodology, and by demonstrating how participant visibility can shape the validity and scope of claims drawn from teacher-led mixed-methods research. Five analytical questions guided the planned analysis: (a) What were students' pretest and posttest knowledge scores? (b) Was there a measurable difference between them? (c) How were observed classroom behaviors related to knowledge gain? (d) What level of oral communication performance did the rubric show? and (e) What themes appeared in students' reflective journals, and what interpretations can those themes support? A sixth question arose only after analysis began, when discrepancies in instrument participation were discovered. It was not part of the original proposal and is identified here as post hoc: To what extent did the four instruments capture overlapping or different subsets of the enrolled class?

The intervention did not produce the measurable gains anticipated at the outset; instead, the analysis revealed a methodological issue with broader relevance. Each instrument made a different group of students visible, and only 4 of 37 matched students appeared in all four. The article therefore moves from the instructional problem to the

intervention evidence and then to the limits of that evidence, using a participation audit that other classroom researchers can replicate. In this way, the study contributes both a cautious evaluation of enthymemic pedagogy and a practical methodological lesson for social-science research conducted in authentic educational settings.

Literature Review

Enthymemic Reasoning and Critical Thinking in Teacher Education

Because an enthymeme leaves a premise unstated, it is pedagogically useful: it asks students to notice what is assumed but not said. Students learn not only to identify claims and reasons but also to ask what hidden belief makes an argument work (Tindale, 2021). Vlah et al. (2024) examined video-based instruction on enthymemes with adult learners and measured critical thinking and clickbait detection. Their work points to both the potential and the limits of brief explicit instruction in enthymemic reasoning, which supports a cautious use of the enthymeme as a tool for analytical awareness.

Critical thinking is central to teacher education because teachers interpret classroom situations, weigh alternatives, and decide under uncertainty. Ennis (2018) argues that analytical thinking should be taught explicitly, and Paul and Elder (2019) treat critical thinking as a practical tool for professional and personal judgment. Bozan et al. (2023) investigated primary school students' critical thinking dispositions and decision-making skills. In Thai higher education, Namsaeng and Ambele (2024) showed how intercultural competence can foster open-minded, analytical approaches, and Sadoo et al. (2023) argued that social studies education can support questioning and the evaluation of diverse perspectives. Reflective writing has also been used to develop critical thinking among Thai EFL learners (Kawinkoonlasate, 2021), which fits an enthymemic approach because both ask learners to make reasoning visible.

Oral Communication in Thai EFL and Teacher Education Contexts

Simpson (2021) argues that EFL instruction becomes more meaningful when students are also trained to evaluate arguments and reason logically. Matyakhan et al. (2024) reported that learning-oriented assessment improved fluency, coherence, interaction, pronunciation, and confidence among Thai pre-service teachers. Narueprempree et al. (2023) used dilemma scenarios to enhance undergraduate students' English speaking skills and perceptions, and other Thai studies support dialogic teaching, active learning, and online collaborative discussion for speaking development (Chaiyaratana & Ekkayokkaya, 2022; Songpitak & Ekkayokkaya, 2022; Tananuraksakul, 2023). These approaches resemble enthymemic pedagogy because they require students to justify opinions and examine assumptions. Related work on task-based learning (Rachayon & Soontornwipast, 2019), inquiry-based learning (Vongsila & Reinders, 2022), and the identification of assumptions and fallacies (Moore & Parker, 2021) points in the same direction. Direct implementations of enthymemic pedagogy in Thai teacher education remain limited.

Missingness and Instrument Visibility in Classroom Research

The methodological problem identified in this study is not new. Attrition, nonresponse, and missing data have long been treated as threats to validity. The What Works Clearinghouse (2022) treats attrition as a validity concern because the final analytic sample may differ from the original sample, and Groves and Peytcheva (2008) show that nonresponse

bias depends not only on how much nonresponse there is but also on whether respondents and nonrespondents differ in ways related to the outcome. Nissen et al. (2019) caution that complete-case analysis assumes data are missing completely at random, an assumption that is often untested, and they recommend multiple imputation where appropriate. The contribution of the present study is not the identification of missingness itself, but a classroom-based demonstration of how a simple presence/absence audit can reveal that multiple instruments in one teacher-led study capture different subsets of the same class.

Method

Design and Context

The study used an embedded mixed-methods classroom intervention. The quantitative strand followed a one-group pretest–posttest quasi-experimental design, with RI-1 providing pre–post evidence and RI-2 describing end-of-cycle oral performance. The second strand used observation (RI-3) and weekly reflective journals (RI-4) to provide process and experience data. There was no control group and no random assignment, because the intervention was embedded in an existing class. A one-group design cannot rule out testing effects, maturation, or concurrent history (Creswell & Creswell, 2018), so the findings are read as classroom-based evidence and not as generalizable experimental evidence.

RI-2 deserves a specific note. No pre-intervention RI-2 score was documented, so RI-2 speaks to a performance level at the end of the cycle and not to change.

Participants

The target group was an intact class of approximately 50 undergraduates enrolled in *English Oral Communication in Educational Contexts* in the Bachelor of Elementary Education program. All enrolled students were invited to take part in the pretest and posttest and in classroom observation. Participation varied across instruments because of attendance, task completion, journal submission, and data-matching conditions. Table 1 shows what was collected and what remained analyzable for each instrument.

Table 1
Instruments, Records Collected, and Analyzable Groups

Instrument	Collected	Analyzable group	Main caution
RI-1 Knowledge test	50 pretest and posttest records	37 matched pairs	Matching required name-based reconciliation because coded IDs were unavailable.
RI-2 Oral performance	42 records	40 unique records	Single colleague-rater; no inter-rater reliability; no pre-intervention score.
RI-3 Observation checklist	50 student-level records	30 linked to RI-1 (29 in later audit)	Some records could not be confidently or ethically linked; one-case discrepancy unresolved.
RI-4 Reflective journals	About 107 entries	104 verified entries from about 27 students	Represents journal submitters only, not the full class.

Note. RI = research instrument.

The Intervention

Enthymemic pedagogy was operationalized as an instructional approach that trains students to construct, analyze, and communicate arguments by identifying claims, reasons, evidence, and unstated assumptions. It was embedded in speaking lessons and not taught as a separate unit of formal logic. Each lesson cycle combined five elements: a mini-lesson with modeling of claims, reasons, evidence, and hidden assumptions; guided analysis of sample statements, short arguments, and classroom scenarios; collaborative reasoning in pairs or small groups to complete missing assumptions and strengthen evidence; an oral task in which students presented ideas, answered questions, and defended conclusions; and reflection through journal prompts about how identifying assumptions affected their thinking and speaking. The intervention ran from August to September 2025, after instrument validation and piloting in July 2025.

Instruments

RI-1 was a 30-item multiple-choice test (five options per item, scored 0 or 1, total 0 to 30) covering critical thinking, enthymeme reasoning, and oral communication knowledge. It was given before and after the intervention. RI-2 was an analytic rubric with five dimensions (fluency; coherence and organization; accuracy; persuasiveness and engagement; pronunciation and intonation), each rated from 1 to 5 for a maximum of 25 points. A single colleague-rater, not the researcher, scored it. RI-3 was a structured observation checklist of binary indicators of engagement, critical thinking, enthymeme reasoning, and oral communication, with observer remarks as supplementary context. RI-4 was a weekly reflective journal with structured prompts, collected over four weeks.

Three experts, one each in education, English communication, and curriculum and instruction, reviewed every instrument. Their ratings produced item-objective congruence values from 0.67 to 1.00, and items that drew concerns were revised before piloting. A pilot with 30 students gave a Kuder-Richardson 20 coefficient of 0.87 for RI-1. That coefficient describes internal consistency within one administration, not stability across the intervention period. Formal inter-rater statistics were not computed for RI-2 or RI-3, and both are therefore interpreted as supportive evidence and not as precise measures.

Data Analysis

Descriptive statistics summarized performance. A paired-samples t-test compared pretest and posttest totals, with Cohen's d_z as the effect size. Normality of score differences was examined with the Shapiro-Wilk test and a Q-Q plot (Ary et al., 2019), and a Wilcoxon signed-rank test served as a robustness check. Because RI-3 totals were non-normal, a Spearman correlation related observed behavior to RI-1 gain. Cronbach's (1951) alpha described the internal consistency of the RI-2 dimensions. Journals were analyzed thematically (Braun & Clarke, 2006). One verified entry was defined as one distinct student submission per week, with repeated signatures, page headers, and continuation pages merged. Entries were coded for spontaneous uptake of enthymemic concepts beyond direct prompt echo, ambiguity, disconfirming content, and suspected AI-assisted or highly formulaic composition. Ambiguous cases were resolved as no uptake, and Week 1 was treated descriptively because it mixed retrospective questions with an early prompt for enthymemic vocabulary. A pre-interpretation rule, documented before the final RI-4 results were

interpreted, required that at least 25% of non-flagged Weeks 2–4 entries show spontaneous uptake before the journals could be read as partial support.

After the main analyses, a presence/absence audit recorded which students in the RI-1 matched panel also appeared in each other instrument. The audit recorded only presence or absence and did not link any reflective content to assessment performance.

Ethical Considerations and Positionality

The study was conducted with permission from the appropriate university authorities. Students were told the purpose of the study, the classroom activities involved, and how their responses would be used. Bilingual English–Thai consent forms were used and are held with the study files. Students were informed, verbally and through the consent process, that participation in data collection was independent of course grading and that critical responses would not affect their standing. No identifying information is reported. Anonymization was applied to reporting and not consistently before initial analysis, because I was also the instructor and some data were handled in identified form during analysis. I did not link individual journal entries to test records, because I judged that linkage to exceed the consent scope for a classroom reflection exercise. Formal approval from a research ethics committee was not obtained before data collection. Accordingly, the study is reported transparently as classroom research conducted under institutional permission and informed consent, without implying retrospective ethics approval.

I was the instructor, the designer of the intervention, the data manager, and the analyst. Students may have written journals and taken part in class knowing their instructor would read them, and my investment in the intervention may have shaped what I noticed and coded. The findings are therefore evidence produced within a teacher–researcher relationship and not by a detached observer (Cohen et al., 2018).

Results

Instrument Denominators

Each instrument produced a different analyzable group (Table 1). RI-1 yielded 37 matched pairs from 50 collected records after name-based reconciliation, because the raw export did not contain consistent coded identifiers. Duplicates were handled by a documented timestamp rule, and sensitivity checks did not change the substantive result. RI-2 yielded 40 unique records from 42 collected after duplicate handling. RI-3 had 50 student-level records, of which 30 were confidently linked to RI-1 for the primary correlation. A later visibility artifact showed 29 RI-3 cases in the RI-1 panel, and I could not reconcile this one-case difference because the row-level file was unavailable. The correlation below therefore keeps $N = 30$, and the 29-case figure is reported separately as an unresolved audit discrepancy. RI-4 comprised 104 verified entries from approximately 27 students, so it represents the journal-submitting subset and not the full cohort.

RI-1 Knowledge Test

The RI-1 total score did not improve from pretest to posttest, $t(36) = -0.12$, $p = .902$, 95% CI for the mean gain $[-3.29, 2.91]$, and the Wilcoxon test agreed, $W = 314$, $p = .765$. The effect was negligible ($d_z = -0.02$), and the domain scores followed the same pattern (Table 2). The upper bound of the 95% confidence interval was 2.91 points, approximately 13% of the pretest mean. The design had about 80% power to detect a paired effect near $d_z = 0.47$ at N

= 37, so the result is best read as no detected moderate or larger effect, with limited power to rule out small ones.

Table 2

RI-1 Pretest and Posttest Scores (N = 37 Matched Pairs)

Score area	Pretest <i>M</i> (<i>SD</i>)	Posttest (<i>SD</i>)	<i>M</i> Gain <i>M</i> (<i>SD</i>)	<i>d_z</i>	Reading
Total score	21.86 (4.86)	21.68 (8.03)	−0.19 (9.29)	−0.02	No measurable gain
Critical thinking	8.05 (1.84)	8.00 (2.69)	−0.05 (3.35)	−0.01	No measurable gain
Enthymeme reasoning	6.43 (2.52)	6.22 (3.17)	−0.22 (3.89)	−0.06	No measurable gain
Oral communication knowledge	7.38 (2.22)	7.46 (2.96)	0.08 (3.61)	0.02	No measurable gain

Note. Cohen's d_z = mean gain divided by the *SD* of gain. Total score range 0–30.

The variance structure calls for caution. Posttest variance was about 2.7 times pretest variance (*SD* = 8.03 vs. 4.86), and the *SD* of gains (9.29) exceeded both. These values imply a pre–post correlation of about .02, meaning posttest scores bore almost no linear relation to pretest scores for the same students. Several conditions could produce this pattern: score instability, mismatches during name-based reconciliation, differences between the two sittings, data-entry or scoring inconsistencies, ceiling effects, heterogeneous responses to the intervention, or limits in how the test captured the construct. The available evidence cannot say which was responsible. The mean gain of −0.19 therefore summarizes substantial bidirectional movement rather than an absence of individual change, and the null result should be interpreted neither as proof that no learning occurred nor as proof that the test failed.

RI-3 Classroom Observation

Among the 30 linked students, the relationship between RI-1 gain and total observation score was weak, negative, and non-significant, Spearman $\rho = -.231$, $p = .220$, 95% CI [−.546, .141]. Domain-level relationships were also weak and non-significant. The observation data therefore did not show a positive association between higher observed engagement and greater knowledge gain.

RI-2 Oral Communication Performance

The 40 rated students had an overall mean of 3.95 out of 5.00 (79.0%), which falls in the Good band (3.50–4.49) of the reporting scale used here and just below the pre-interpretation strong-evidence threshold of 4.00. The scale was an applied reporting convention and not a pre-registered feature of the instrument. Table 3 summarizes the distribution. Internal consistency across the five dimensions was very high, Cronbach's $\alpha = .985$, but with one rater this describes consistency within that rater's scoring and not agreement between raters. Dimension-level scores are therefore not interpreted separately.

The audit also showed that only 20 of the 37 RI-1 matched students were RI-2 rated, so roughly half of the RI-2 records fall outside the RI-1 panel.

Table 3

RI-2 Oral Communication Performance Indicators (40 Unique Records)

Indicator	Result	Reading
Overall mean	3.95 / 5.00	Good overall performance (3.50–4.49 band)
Percentage equivalent	79.0%	Just below the pre-interpretation 80% (4.00) threshold
Good or higher (≥ 3.5)	67.5%	Most rated students reached at least Good
Score of 4.0 or higher	55.0%	Slightly more than half
Below satisfactory (< 3.0)	12.5%	A minority remained below satisfactory

Note. Scored by a single colleague-rater. Cronbach's $\alpha = .985$ across five dimensions.

RI-4 Reflective Journals

The journals were prompted, identified, instructor-facing, and partly affected by suspected AI-assisted or highly formulaic composition, so they were read as evidence of described perceptions and prompted reflective uptake and not as independent proof of growth. Table 4 shows the weekly coding. Across Weeks 2–4, 27 of 62 non-flagged entries showed spontaneous uptake beyond prompt echo (43.5%), which passed the 25% rule. This figure is a conservative lower bound, because 16 ambiguous cases were resolved against uptake. If all ambiguous cases had counted as uptake, the rate would have been 69.4%, but I did not use that figure as the result.

Table 4

RI-4 Weekly Coding Summary

Week	Verified entries	Suspected AI or formulaic	Non-flagged	Uptake	Ambiguous	Disconfirming
Week 1	27	5	22	8	6	10
Week 2	26	5	21	9	5	2
Week 3	26	4	22	10	6	4
Week 4	25	6	19	8	5	5
Weeks 2–4 total	77	15	62	27	16	11

Note. Uptake, ambiguous, and disconfirming counts refer to non-flagged entries. Week 1 was treated descriptively.

Two cautions follow. First, entries suspected of AI-assisted or formulaic composition made up 15 of 77 verified Weeks 2–4 entries (19.5%), at the edge of the pre-interpretation 20% concern threshold, with Week 4 highest at 6 of 25 (24.0%). These flags came from documented human-review criteria and not from a validated detector, so they are cautionary and not confirmed determinations. Second, disconfirming content appeared in 11 of 62 non-flagged entries (17.7%), more than the 10% I had predicted, although it was generally brief and not central to the entries containing it. Week 1 showed spontaneous uptake in 8 of 22

non-flagged entries (36.4%), above the 25% threshold even though it preceded the intervention window, so the threshold does not fully separate intervention uptake from earlier mixed reflection. A blind re-coding pass reproduced the earlier counts, but because the earlier counts stayed visible in the same workbook, I report it as a procedural safeguard and not as independent replication.

Participation Visibility Across Instruments

Table 5 anchors the audit on the 37 matched RI-1 students. Only 4 appeared in all four instruments, 19 appeared in exactly three, 13 in exactly two, and 1 in RI-1 only. Only 14 of the 37 submitted RI-4 journals, while RI-4 itself came from approximately 27 submitters, so about half of the journal corpus came from students outside the RI-1 panel. The same asymmetry applies to RI-2. The two instruments providing the most supportive evidence therefore included substantial numbers of students whose knowledge scores were not represented in the RI-1 matched analysis. Because the audit was anchored on the RI-1 panel, it could not assess the 13 students excluded during matching, so visibility among the least consistently participating students remains unknown.

Table 5

Participation Visibility Anchored on the RI-1 Matched Panel

Visibility measure	Count
RI-1 matched panel	37
Also RI-2 rated	20 / 37
Also RI-3 linked (current auditable artifact)	29 / 37
Also RI-4 submitted	14 / 37
Visible in all four instruments	4 / 37
Visible in exactly three instruments	19 / 37
Visible in exactly two instruments	13 / 37
Visible in RI-1 only	1 / 37

Note. The audit recorded presence or absence only. It could not assess the 13 students excluded during RI-1 matching.

Discussion

What the Evidence Can and Cannot Support

The study did not show that enthymemic pedagogy produced measurable gains in critical thinking or oral communication as captured by the quantitative instruments. The knowledge test showed no meaningful mean change, observed engagement was not positively associated with knowledge gain, and the Good oral-performance ratings came from a single rater and a group only partly overlapping with the tested students. The journals suggest that some students who submitted them could describe reasoning and communication gains and sometimes extended taught concepts beyond the prompt. That is partial and uneven support for the intervention as a reflective classroom experience. It does not establish developmental growth, and it does not show that the students who described gains were the ones whose scores were measured.

The contrast between described gains and unchanged scores is therefore a contrast between two instruments' respondent groups, not between one group's perceptions and its own performance. Only 14 of the 37 students in the test panel submitted journals. The instruments were initially assumed to represent the same class, but the audit showed that they captured materially different subsets. Consequently, before evaluating whether the intervention worked, the study first had to establish whose learning the classroom research actually measured.

Methodological Implications for Classroom Researchers

The audit that exposed this problem is methodologically simple. It records only presence or absence for each instrument, can be implemented with minimal analytic burden, and is consistent with the methods literature on attrition, nonresponse, and complete-case analysis (Groves & Peytcheva, 2008; Nissen et al., 2019; What Works Clearinghouse, 2022). Several practical recommendations follow from this study.

First, assign student codes before data collection and use them across every instrument, which prevents post hoc name matching and the exclusion of students because of spelling or transliteration variation. Second, report the pre-post correlation, internal consistency, and variance pattern alongside mean change. A null mean from a test with no demonstrated stability cannot be read as clear evidence about an intervention, and multiple imputation would address missing cases but not an unstable construct. Third, use at least two trained raters, with separate rater columns and a scoring reconciliation rule, whenever performance ratings carry interpretive weight. Fourth, anonymize reflective data before analysis where ethically and practically possible, and include at least one unprompted transfer task that lets students show no enthymemic reasoning if the concept has not transferred. Fifth, run a roster-level visibility audit at the start and end of analysis. Finally, keep perceived learning, prompted reflective uptake, performance judgment, observation, and measured knowledge gain distinct in reporting. Treating them as interchangeable produces stronger-looking conclusions than the data can bear.

Limitations

This was one intact class with no control group, so testing, maturation, and concurrent coursework cannot be ruled out. Names, not codes, linked pretests and posttests, and the missingness mechanism was not formally tested. A single colleague-rater scored RI-2 without a pre-intervention baseline, and no inter-rater statistics exist for RI-2 or RI-3. The journals were instructor-facing and partly flagged for formulaic composition. One unresolved one-case discrepancy remains in the RI-3 counts. Finally, I was both instructor and researcher, and some data were handled in identified form during analysis. The audit shows that the instruments made different groups visible. It does not show that the instructor-researcher relationship alone caused the difference, and ordinary classroom logistics will explain some of it.

Conclusion

In this study of enthymemic pedagogy with pre-service elementary teachers, knowledge scores did not improve significantly, observed engagement was not positively associated with knowledge gain, oral performance was rated Good by a single colleague-rater, and journal analysis indicated prompted reflective uptake among a subset of submitters. The

instruments, however, did not see the same class: only 4 of 37 matched students appeared in all four. The study's principal contribution is therefore methodological, a classroom-based demonstration of how instrument-specific visibility shapes what teacher-led research can claim. The broader question for teachers and researchers is whether the instruments they use adequately represent the students whose learning they seek to understand.

Declarations

Funding. This work was supported by a Suan Dusit University research grant, fiscal year 2026 (Contract No. มสศ. 694006).

Ethics. The study was conducted with permission from the appropriate university authorities and used informed consent. Formal approval from a research ethics committee was not obtained before data collection. The study is therefore reported transparently as classroom research conducted under institutional permission and informed consent, without implying retrospective ethics approval (see Ethical Considerations and Positionality).

Use of AI tools. OpenAI was used as an editorial and drafting aid to condense, reorganize, and language-edit material from the author's institutional research report and to support consistency checking during manuscript preparation. It was not used to fabricate or alter participant data. The author independently reviewed and verified the final text, numerical results, interpretations, and references and assumes full responsibility for the manuscript.

References

- Aristotle. (2007). *On rhetoric: A theory of civic discourse* (G. A. Kennedy, Trans.; 2nd ed.). Oxford University Press.
- Ary, D., Jacobs, L. C., Irvine, C. K. S., & Walker, D. (2019). *Introduction to research in education* (10th ed.). Cengage Learning.
- Bozan, M. A., Karahan, E., Akçay, A. O., & Akçay, İ. M. (2023). An investigation of primary school students' critical thinking dispositions and decision-making skills. *International Journal of Educational Research Review*, 8(2), 137–150. <https://doi.org/10.24331/ijere.1205285>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Chaiyaratana, C., & Ekkayokkaya, S. (2022). Effects of dialogic teaching in a CLIL science subject on English oral communication ability of primary students. *LEARN Journal: Language Education and Acquisition Research Network*, 15(1), 523–544. <https://so04.tci-thaijo.org/index.php/LEARN/article/view/259199>
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.
- Ennis, R. H. (2018). *Critical thinking across the curriculum: A brief edition of the thinker's guide to analytic thinking*. Rowman & Littlefield.
- Freire, P. (2020). *Pedagogy of the oppressed* (50th anniversary ed.). Bloomsbury Publishing.

- Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly*, 72(2), 167–189. <https://doi.org/10.1093/poq/nfn011>
- Kawinkoonlasate, P. (2021). Developing critical thinking through reflective writing in EFL classrooms in Thailand. *English Language Teaching*, 14(6), 42–49. <https://doi.org/10.5539/elt.v14n6p42>
- Matyakhan, R., Sukseemuang, P., & Ngonkum, S. (2024). Developing English oral communication ability of Thai pre-service teachers through learning-oriented assessments. *LEARN Journal: Language Education and Acquisition Research Network*, 17(1), 60–88. <https://so04.tci-thaijo.org/index.php/LEARN/article/view/267979>
- Moore, B. N., & Parker, R. (2021). *Critical thinking* (13th ed.). McGraw-Hill Education.
- Namsaeng, P., & Ambele, E. A. (2024). Integrating intercultural competence into the curriculum: A catalyst for critical thinking in Thai tertiary education. *THAITESOL Journal*, 37(2), 74–95. <https://so05.tci-thaijo.org/index.php/thaitesoljournal/article/view/268999>
- Narueprempree, P., Jansawang, N., & Pongpairaj, N. (2023). The effects of dilemma scenarios on undergraduate students' English speaking skills and perceptions. *Journal of Asia TEFL*, 20(3), 756–770. <https://doi.org/10.18823/asiatefl.2023.20.3.17.756>
- Nissen, J. M., Donatello, R., & Van Dusen, B. (2019). Missing data and bias in physics education research: A case for using multiple imputation. *Physical Review Physics Education Research*, 15(2), 020106. <https://doi.org/10.1103/PhysRevPhysEducRes.15.020106>
- Paul, R., & Elder, L. (2019). *Critical thinking: Tools for taking charge of your professional and personal life* (2nd ed.). Pearson Education.
- Rachayon, P., & Soontornwipast, K. (2019). Improving Thai students' speaking skills through flipped classroom and task-based learning. *Journal of Language Teaching and Research*, 10(2), 364–370.
- Sadoo, S., Nandapañño (Sriyort), P. S., Theerangkuro (Pukham), P. W., Uthaisa, P., & Tenissara, S. (2023). Developing critical thinking skills through social studies education in Thai classrooms. *Journal of Buddhist Education and Research*, 9(1), 375–383. <https://so06.tci-thaijo.org/index.php/jber/article/view/278749>
- Saengbanchong, A., & Saito, E. (2020). An exploration of Thai pre-service teachers' perceptions of oral communication challenges. *Kasetsart Journal of Social Sciences*, 41(3), 548–553. <https://doi.org/10.1016/j.kjss.2019.07.001>
- Simpson, J. (2021). *Teaching critical thinking in EFL classrooms: From theory to practice*. Routledge.
- Songpitak, S., & Ekkayokkaya, M. (2022). The effects of using online collaborative learning on English oral communication ability of EFL lower secondary students. *An Online Journal of Education*, 17(1), OJED1701019. <https://doi.org/10.14456/ojed.2022.18>
- Tananuraksakul, N. (2023). Building Thai students' resilience and confidence in speaking English through active learning. *Journal of English Language Teaching and Linguistics*, 8(2), 145–158. <https://doi.org/10.21462/jeltl.v8i2.978>
- Tindale, C. W. (2021). *The power of rhetoric: Rhetorical theory and practice*. Routledge.
- Vlah, A., Wijnia, L., Lutz, C., Burke, M., & Loyens, S. M. M. (2024). Can video lectures on enthymemes improve adult learners' critical thinking and clickbait detection skills? *Education Sciences*, 14(12), Article 1284. <https://doi.org/10.3390/educsci14121284>
- Vongsila, V., & Reinders, H. (2022). Inquiry-based learning and its effects on language acquisition and critical thinking among Thai learners. *Asian-Pacific Journal of Second and*

Foreign Language Education, 7, Article 18. <https://doi.org/10.1186/s40862-022-00145-9>

What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0* (WWC 2022008REV). Institute of Education Sciences, U.S. Department of Education. <https://ies.ed.gov/ncee/wwc/handbooks>