Arab World English Journal
INTERNATIONAL PEER REVIEWED JOURNAL ISSN: 2229-9327

Arab World English Journal (AWEJ) Volume 17. Number 2, June 2026 Pp.356-373
DOI: https://dx.doi.org/10.24093/awej/vol17no2.23

The Impact of Content and Language Integrated Learning on English Speaking Proficiency: A Mixed-Methods Study in Chinese Tertiary Education

Hengzhi Hu
School of Foreign Languages, North China Electric Power University, Beijing, China
Faculty of Education, Universiti Kebangsaan Malaysia, Bangi, Malaysia
Email: hengzhi.hu@ukm.edu.my
Harwati Hashim
Faculty of Education, Universiti Kebangsaan Malaysia, Bangi, Malaysia
Corresponding Author: harwati@ukm.edu.my

Received: 02/29/2026 Accepted: 05/30/2026 Published:06/15/2026

Abstract
Content and Language Integrated Learning is recognized for its potential to enhance language proficiency while fostering content mastery, particularly in multilingual and globalized education systems. However, its impact on English-speaking proficiency, particularly in the Chinese tertiary education context, remains underexplored. This study addresses this gap by examining the effects of a 12-week Content and Language Integrated Learning intervention on the speaking complexity, accuracy, fluency, and pronunciation of Chinese university students. The study is significant because it extends current understanding of how CLIL influences different dimensions of speaking proficiency in higher education and provides evidence-based insights for improving English language pedagogy in the Chinese tertiary context. Employing a mixed-methods approach, quantitative data were gathered through pre- and post-tests, complemented by qualitative insights from classroom observations and student interviews. The findings revealed that the intervention significantly improved fluency and pronunciation, yet had limited effects on speaking complexity and accuracy compared to traditional instruction. Qualitative data highlighted that meaningful communication tasks fostered confidence and spontaneity in speaking but lacked a focus on explicit grammar and vocabulary instruction, leading to cognitive overload in balancing content comprehension with linguistic refinement. The study underscores the need for integrating form-focused instruction into Content and Language Integrated Learning frameworks to achieve balanced language development. These findings provide valuable implications for teachers and curriculum planners, emphasizing the importance of harmonizing content and linguistic objectives to optimize the dual focus of Content and Language Integrated Learning in English language education.

Keywords: Content and Language Integrated Learning, China, English speaking, English language education, higher education
How to cite: Hu, H. & Hashim, H. (2026). The Impact of Content and Language Integrated Learning on English Speaking Proficiency: A Mixed-Methods Study in Chinese Tertiary Education. *Arab World English Journal, 17*(2): 356-373.
https://dx.doi.org/10.24093/awej/vol17no2.23

Introduction

Content and Language Integrated Learning (CLIL) has emerged as a transformative educational approach that combines subject content teaching with the simultaneous learning of a foreign language. Originating in Europe, CLIL emphasizes a dual-focus methodology that allows students to acquire content knowledge while developing language skills in a meaningful context (Coyle & Meyer, 2021). This integration has been widely recognized for promoting linguistic competence alongside cognitive and cultural development, making it a favored approach in multilingual and multicultural education systems (Hidalgo & Ortega-Sánchez, 2023; Sylvén & Tsuchiya, 2023). By fostering real-world language use and active engagement with content, CLIL effectively bridges the gap between theoretical knowledge and practical application.

In China, the integration of CLIL into education is a relatively recent development (Li, 2021), reflecting the country's evolving focus on global competitiveness and the cultivation of bilingual or multilingual citizens. English, as a global lingua franca, occupies a central role in China's educational reforms (Li, 2020), particularly at the tertiary level. National policies, such as the *Guidelines for College English Teaching* (College Foreign Language Teaching Steering Committee of the Ministry of Education, 2020) and the *National Standards for Teaching Quality in Foreign Languages and Literature* (Teaching Advisory Board in Tertiary Education of the Ministry of Education, 2018), underscore the importance of innovative teaching methods to enhance students' English proficiency while ensuring subject mastery and cultural development. These policies encourage the adoption of content-based approaches to produce graduates who are both academically and linguistically competent and capable of navigating international academic and professional environments, thereby providing top-down support for the application of CLIL (Cao, 2021).

While CLIL has been applied in various educational contexts in China, its implementation in English language education at the tertiary level has largely focused on improving academic reading and writing skills (Cao, 2021; Zhang, 2022), with less emphasis on speaking proficiency (Hu et al., 2023). This relative neglect is problematic because many Chinese university students continue to demonstrate comparatively low levels of English-speaking proficiency, particularly in spontaneous oral communication, despite years of formal English learning (Chen, 2024). English-speaking proficiency is crucial for effective communication in academic and professional domains, yet it remains a challenging area for many Chinese students due to limited opportunities for authentic oral language use in traditional classrooms (Zhang, 2024). Moreover, studies exploring the impact of CLIL on speaking proficiency in China's higher education system are scarce, and most existing research has been underpinned by the positivist paradigm and does not address the nuanced oral language development that involves multiple skills (Hu et al., 2023), such as complexity, accuracy, fluency, and pronunciation (CAFP), which are critical indicators of comprehensive speaking proficiency.

Therefore, this study aims to fill these gaps by examining the effect of CLIL on the English-speaking proficiency of Chinese university students. Employing a mixed-methods approach, the research investigates both quantitative improvements in students' speaking abilities measured by analytic CAFP indices and qualitative insights into how CLIL creates an environment that supports the development of English speaking proficiency. The following research questions are proposed: (1) How does CLIL affect Chinese university students' English-speaking proficiency in terms of CAFP compared with regular instruction? and (2) How do Chinese university students perceive the ways in which CLIL supports or constrains the development of their English-speaking

proficiency? The significance of this research lies in its contribution to both theory and practice. By focusing on an underexplored area—speaking proficiency—it enriches the body of literature on CLIL and offers actionable insights for educators and policymakers. The findings will help optimize CLIL implementation strategies to ensure they align with the linguistic and academic goals of Chinese universities and support broader national aspirations for internationalization and educational innovation.

Literature Review

CLIL is an educational approach that simultaneously fosters subject knowledge and language proficiency, offering a dual-focused methodology that enhances both areas in a single instructional context. Defined as a teaching practice where content is delivered through a foreign language (Coyle & Meyer, 2021), CLIL prioritizes contextualized and practical language use, enabling learners to acquire linguistic skills while engaging deeply with subject matter (Peérez & Unamuno, 2024). This approach differs from traditional language teaching by emphasizing real-world application and promoting active engagement, critical thinking, and problem-solving (Costa, 2025), making it especially suitable for multilingual and globalized educational settings.

In China, CLIL has gained prominence in recent years, with educational authorities, driven by national policies emphasizing bilingual education and global competencies, explicitly encouraging integrating subject content with language learning to improve students' English proficiency and subject-specific expertise (Zheng, 2023). According to systematic reviews of CLIL literature, CLIL has been predominantly implemented in China's tertiary education system, where it has been incorporated into various disciplines such as science, technology, and business studies (Hu et al., 2023). At the tertiary level, CLIL is particularly significant in the context of College English, a compulsory course for non-English major students. Designed to enhance general English proficiency, College English provides an ideal platform for integrating subject-specific content with language learning (Liao, 2024). Through CLIL, students engage with materials that not only develop their linguistic competence but also expose them to content relevant to their fields of study, fostering a more holistic educational experience. For English majors, CLIL serves as an advanced pedagogical approach to deepen their mastery of the language while applying it to specialized areas such as linguistics, literature, and translation studies (Zhang, 2023). This dual focus aligns with the goal of producing graduates who are both linguistically proficient and academically versatile.

The benefits of CLIL are varied, including enhancing content knowledge, fostering intercultural competence, and improving overall language proficiency (Zanoni, 2021). Among these, its effect on English speaking proficiency has garnered significant attention from educators and researchers. Underpinned by second language acquisition theories, such as the Input Hypothesis (Krashen, 1987), Output Hypothesis (Swain, 1985), and Sociocultural Theory (Vygotsky, 1978), CLIL is believed to create an immersive environment conducive to developing foreign language speaking skills. While CLIL has demonstrated potential in improving speaking skills internationally, its specific impact on English speaking proficiency within China remains underexplored. Research from global contexts suggests that CLIL can enhance dimensions of speaking proficiency. For instance, studies indicate that CLIL students often demonstrate greater lexical diversity (Zhang et al., 2024) and fewer grammatical errors (Ikeda et al., 2022) compared to their peers in traditional language programs. Additionally, the frequent oral interactions inherent

in CLIL classrooms significantly enhance fluency (Hu, 2022) and pronunciation clarity (Uhl et al., 2023).

Emerging research from China supports these global findings. In addition to research that highlights the benefits of CLIL on general English speaking proficiency. Ding (2022) and Hu et al. (2022) found that CLIL students showed increased confidence and fluency during English presentations, while Zhang (2022) reported significant improvements in pronunciation accuracy among engineering students. As such, current research tends to emphasize general language outcomes or specific sub-skills in isolation, offering a fragmented picture of how CLIL impacts speaking ability in content-driven learning environments. Additionally, most studies adopt quantitative methodologies (Ding, 2022; Hu et al., 2022; Zhang, 2022), focusing on measurable outcomes while neglecting qualitative insights into students' experiences and perceptions.

For a comprehensive understanding of how CLIL facilitates oral language development, this study adopts a pragmatic research paradigm underpinned by the CAFP framework to analyze English speaking proficiency. The CAFP framework provides a structured lens for evaluating multidimensional speaking skills, which are critical for effective communication in academic and professional contexts (Joo, 2022; Luoma, 2004). Complexity captures the richness and variety of linguistic structures; accuracy focuses on grammatical correctness; fluency addresses the smoothness and speed of speech delivery; and pronunciation emphasizes the clarity and intelligibility of spoken language (Alonso, 2018). The interactive and contextualized nature of CLIL is hypothesized to encourage learners to engage actively in oral communication, thereby fostering improvements across these dimensions (Klewitz, 2021). Through this lens, the study aims to generate nuanced insights into the mechanisms by which CLIL contributes to oral language development, offering a deeper understanding of its potential to enhance English speaking skills in tertiary education. This approach not only highlights the multifaceted impact of CLIL on speaking proficiency but also provides a foundation for targeted instructional strategies and pedagogical innovations.

Methodology

Research Design

This study employed a mixed-methods sequential explanatory design, integrating quantitative and qualitative approaches to provide a comprehensive understanding of the effects of CLIL on English speaking proficiency. Quantitative findings were used to measure the impact of CLIL on students' CAFP dimensions, while qualitative findings explored students' perceptions and experiences to contextualize and deepen the interpretation of the quantitative results. The mixed-methods approach was justified as it enabled the triangulation of data (Thomas, 2021), which offered nuanced insights into both the outcomes and processes associated with CLIL implementation.

Participants

The study was conducted at a Chinese university renowned for its educational innovations. The CLIL approach was applied to selected College English courses, which are compulsory for non-English majors. These courses were chosen for their well-established and authentic curriculum, which had been rigorously developed and studied, ensuring a high degree of relevance to the research objectives. Purposive sampling was adopted because the study required participants who met specific criteria closely aligned with the research focus, thereby enabling the collection

of relevant and contextually appropriate data. This sampling method was particularly suitable for ensuring that the selected participants shared comparable educational experiences and proficiency levels, which strengthened the validity of comparisons across the intervention. The inclusion criteria were: (1) students enrolled in College English courses, (2) comparable language proficiency levels as measured by pre-study English placement tests, and (3) willingness to participate in both quantitative assessments and qualitative interviews. Two year-three College English classes were then selected, with one randomly assigned as the experimental group (EG) (*N* = 40), receiving CLIL instruction, and the other as the control group (CG) (*N* = 42), receiving regular instruction akin to the grammar-translation method. This quasi-experiment spanned 12 consecutive weeks within an academic semester, a duration considered sufficient to observe any potential effects of the CLIL approach. As shown in Table 1, the two classes, comprising students aged 20 to 21 years, had similar gender proportions. The students were enrolled in various majors and had 13–15 years of English learning experience. Both groups were taught by the same teacher, who held a master's degree in English education and had over 10 years of teaching experience, including 3 years implementing CLIL methodologies. The teacher's expertise ensured consistency in instruction quality across groups, minimizing extraneous variables related to teacher effects.

Table 1. *Participants' demographic profiles*

Group	Age	Gender	Majors	Years of English Learning
EG	20-21	Male: N = 24 Female: N = 16	Education (N = 8) Business/Finance (N = 16) Information Technology (N = 7) Engineering (N = 9)	13-15
CG	20-21	Male: N = 22 Female: N = 20	Education (N = 7) Business/Finance (N = 18) Information Technology (N = 7) Engineering (N = 10)	13-15

Data Collection and Analysis

Quantitative data were collected through a mock standardized English speaking test, namely College English Test Band-6, which was organized both before and after the study to measure student progress under different instructions. Previous research had mostly used human ratings to quantify speaking CAFP, namely using established marking rubrics and give scores. However, this approach was very subjective. Therefore, the analytic CAFP indices, as shown in Table 2, were used to ensure that students' English speaking proficiency could be accurately measured. Qualitative data, on the other hand, were collected from classroom observations of the EG during the intervention. The observation protocol developed by Zheng (2023) was used, aimed at noting down key moments wherein CLIL was deemed to be influential in English speaking proficiency. Also, at the end of the intervention, semi-structured interviews were conducted with nine students, recruited through voluntary sampling. The interview guideline developed by Fan and Zhang (2024) was adapted to focus on students' perceptions of CLIL, especially its effects on English speaking proficiency and the challenges they faced in developing speaking proficiency.

Table 2. *CAFP indices*

Skill	Dimension	Index	Justification
Speaking	Grammatical Complexity	Mean Length of AS-unit (MLAS) = Number of Tokens divided by Number of AS-units	Norris and Ortega (2009)
		Mean Length of Clause (MLC) = Number of Tokens divided by Number of Clauses	
		Ratio of Sub-clauses (RS) = Number of Clauses divided Number of AS-units	
	Lexical Complexity	D-score calculated in the Computerized Language Analysis program	Siskova (2012)
	Global Accuracy	Percentage of Errorless Clauses (PEC) = Number of Errorless Clauses divided by Number of Clauses	Ellis and Barkhuizen (2005)
		Percentage of Errorless AS-unit (PEA) = Number of Errorless AS-units divided by Number of AS-units	
	Fluency	Number of Pauses (NP) = Number of Pauses divided by Speaking Time in Seconds	Skehan (2009)
		Number of Repairs (NR) = Number of Repairs divided by Speaking Time in Seconds	
	Pronunciation	Intelligibility = Number of Accurately Transcribed Keywords divided by Number of Keywords in Each Sentence	Luoma (2004)

To examine the effect of CLIL and compare it with regular instruction, the collected quantitative data were processed using the Statistical Package for Social Sciences to perform both between-group and within-group comparisons. Since not all data for the variables were normally distributed, non-parametric tests (the Mann-Whitney U Test and the Wilcoxon Signed-Rank Test) were employed to ensure accurate analysis. The qualitative data were analyzed thematically, following the steps of familiarization with data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and producing the report (Braun & Clarke, 2021).

Validity and Reliability

Throughout the study, extensive measures were undertaken to ensure its validity and reliability. For instance, the speaking test used in the study was reviewed by a panel of experts for face validity. These experts concluded that the test was well-designed to elicit meaningful and authentic responses from students, aligning with the study's objectives. Similarly, the qualitative instruments, including interview guides and observation protocols, were reviewed and validated by the experts, who confirmed their relevance and potential to generate insightful data. Importantly, the quantification of speaking proficiency based on the CAFP framework required meticulous identification of linguistic components, such as AS-units, clauses, errors, pauses, and

repairs. To ensure accuracy, a group of professional interpreters was engaged in the process. These interpreters underwent specialized training and participated in a pilot study, where they independently coded the same five samples of speaking test recordings. Their transcripts were then compared, resulting in a high inter-rater agreement rate of 95% to 97%. This level of agreement indicates a strong consistency in coding (Bulté & Housen, 2012), reinforcing the reliability of the analytical process.

Findings

Quantitative Findings

Firstly, a between-group analysis was conducted to compare the pre-test and post-test scores of the EG and the CG. As shown in Appendix A, there was no statistically significant difference ($p > .05$) in the pre-test scores for the CAFP indices, indicating that both groups had similar levels of English-speaking proficiency before the study. However, the post-test revealed significant differences between the groups. The EG scored lower in speaking complexity, including MLAS ($p < .001$), MLC ($p = .001$), RS ($p < .001$), and D-score ($p < .001$), indicating that the CG demonstrated a higher level of syntactic and lexical complexity in English speaking after the study. Similarly, the CG showed higher levels of speaking accuracy, as reflected by their higher PEC ($p < .001$) and PEA ($p < .001$) scores compared to the EG in the post-test. Conversely, the EG had significantly lower NP ($p < .001$) and NR ($p < .001$) scores than the CG, suggesting that the EG exhibited greater speaking fluency, with fewer pauses and repairs. Additionally, in terms of intelligibility, the EG scored significantly higher than the CG ($p < .001$), indicating better pronunciation at the end of the study.

Then, a within-group analysis was conducted. According to Appendix B, no statistically significant differences were found for the EG in terms of MLAS, MLC, and RS (indicators of syntactic complexity in English speaking), D-score (lexical complexity), and PEC and PEA (speaking accuracy) ($p > .05$). This indicates that the EG did not make noticeable progress in speaking complexity or accuracy during the study. However, significant differences were observed in NP and NR ($p < .001$), with pre-test scores higher than post-test scores. Given the nature of these indices, this suggests that the EG improved speaking fluency, with fewer pauses and repairs by the end of the intervention. Similarly, intelligibility showed a significantly higher post-test score ($p < .001$), indicating improved pronunciation. In contrast, the CG demonstrated significant improvement in speaking complexity and accuracy, with statistically significant differences observed in MLAS, MLC, RS, D-score, PEC, and PEA ($p \leq .001$). However, for fluency and pronunciation indices, namely NP, NR, and intelligibility, no significant differences were found, suggesting that the CG did not improve in these areas.

Combining the quantitative findings, it was observed that the intervention, when compared to regular instruction focused on linguistic forms such as grammar and vocabulary, did not enhance students' speaking complexity or accuracy. In other words, the traditional approach had a more pronounced effect on students' language use in speaking. However, the intervention demonstrated a significant impact on speaking fluency and pronunciation—skills that regular instruction failed to improve. The combined results from the between-group and within-group analyses highlighted both the strengths and weaknesses of the intervention.

Qualitative Findings

Emphasis on Meaningful Communication

The findings revealed that the CLIL intervention's emphasis on content-based tasks encouraged students to engage in meaningful communication, focusing on practical language use rather than strict adherence to grammatical structures. During classroom observations, students actively participated in group discussions, presentations, and collaborative problem-solving, which required them to express ideas clearly and effectively. Similarly, interviews with students highlighted that they felt more confident speaking in class as the tasks prioritized conveying meaning over form. Many students mentioned that the real-world relevance of these activities motivated them to focus on delivering their ideas fluently and clearly. This practical engagement, supported by meaningful communication tasks, likely contributed to improvements in speaking fluency and pronunciation. By prioritizing communication over linguistic precision, the CLIL approach provided students with opportunities to build confidence in speaking, reduce hesitations, and produce speech with fewer pauses and clearer articulation.

> "Students engaged actively in group discussions, expressing their ideas with noticeable fluency. While minor grammatical errors occurred, they did not hinder communication or participation... The focus was clearly on sharing content and maintaining the flow of conversation." (Observation Note)
>
> "I liked that we were encouraged to focus on what we wanted to say, not just how to say it perfectly. It helped me feel more confident when speaking. I wasn't afraid to make mistakes because the goal was to communicate effectively." (Interview)

Development of Confidence in Fluency

The findings indicated that frequent interactive activities within the CLIL intervention played a significant role in fostering students' confidence and spontaneity in speaking. Classroom observations showed that students were actively engaged in activities such as role-plays, debates, and collaborative tasks, which required them to speak without overthinking linguistic accuracy. This focus on active participation encouraged students to take risks and communicate their ideas more freely, contributing to noticeable improvements in fluency. Interviews supported these observations, as students reported feeling more comfortable speaking in class over time. They attributed this increased confidence to the interactive nature of the tasks, which allowed them to practice speaking in a supportive and engaging environment. However, despite the progress in fluency, students and observations also highlighted that linguistic precision, such as grammatical accuracy and vocabulary complexity, remained underdeveloped due to the limited focus on form during these activities.

> "During role-plays and debates, students spoke with noticeable fluency and confidence, engaging actively with their peers. While some grammatical errors occurred, these did not deter their willingness to express ideas freely." (Observation Note)
>
> "The interactive activities helped me feel more relaxed when speaking, as the focus was on communicating [my ideas]... I became less afraid of making mistakes because the tasks encouraged active participation." (Interview)

Limited Focus on Linguistic Forms

The findings revealed that the CLIL intervention's limited emphasis on explicit grammar and vocabulary instruction hindered students' ability to develop speaking complexity and

accuracy. Classroom observations highlighted that while activities encouraged communication and fluency, there was little direct focus on linguistic structures or advanced vocabulary. Students often relied on simpler sentence constructions and familiar words, which restricted the overall syntactic and lexical complexity of their speech. Interviews further emphasized this issue, with students expressing that they felt less equipped to construct grammatically accurate or complex sentences during speaking tasks. Some students noted that the lack of targeted feedback on language forms made it difficult for them to identify and correct recurring errors, limiting their progress in speaking accuracy and complexity.

> "Students communicated effectively during tasks but often used basic sentence structures and repeated vocabulary, with little focus from the teacher on improving grammar." (Observation Note)
>
> "I could speak fluently, [but] I didn't learn how to make my sentences more complex or fix my grammar mistakes during the activities." (Interview)

Content Over Language Precision

The findings indicated that the integration of subject-specific content in the CLIL intervention shifted students' focus toward understanding and conveying content, often at the expense of grammatical accuracy and syntactic complexity. Classroom observations showed that students prioritized delivering subject-related ideas effectively, which sometimes led to oversimplified sentence structures and recurring grammatical errors. The emphasis on content comprehension and communication left little room for deliberate attention to language precision. Interviews reinforced this observation, with students mentioning that their primary focus during tasks was on mastering the content rather than on constructing accurate or complex sentences. Some students noted that they found it challenging to balance content learning with linguistic refinement, resulting in less progress in grammatical and syntactic development.

> "Students focused on explaining subject-related ideas, often using simple sentence structures and overlooking grammatical accuracy." (Observation Note)
>
> "I concentrated more on getting the content right, so I didn't think much about making my sentences complex or grammatically correct." (Interview)

Cognitive Overload from Dual Focus

The simultaneous focus on learning subject-specific content and using English as the medium of instruction in the CLIL approach created cognitive overload for students. Observations revealed that students often struggled to balance understanding complex subject material while constructing grammatically correct sentences. This dual demand led to a prioritization of content communication over linguistic accuracy and complexity. Interviews supported this finding, as students expressed difficulty managing both tasks simultaneously. Many reported that their mental effort was directed toward grasping the subject matter, leaving less capacity to focus on refining grammar or expanding vocabulary during speaking tasks. This cognitive strain likely contributed to the limited improvements in complexity and accuracy observed in the quantitative results.

> "Students appeared to focus heavily on understanding the subject content, often sacrificing grammatical accuracy and sentence complexity in their spoken responses." (Observation Note)

> "It was hard to think about both the content and how to say it correctly at the same time. Most of my attention went to understanding the material, so I didn't focus much on grammar." (Interview)

Summary of Findings

Overall, the findings suggest that the CLIL intervention produced a selective rather than uniform effect on students' English-speaking proficiency. Quantitatively, both groups were comparable at the pre-test stage, but by the post-test, the EG outperformed the CG in fluency and pronunciation, as reflected in significantly fewer pauses and repairs and higher intelligibility, whereas the CG achieved greater gains in speaking complexity and accuracy, including syntactic, lexical, and error-related measures. Within-group analysis further confirmed this pattern: the EG improved significantly only in fluency and pronunciation, while the CG improved significantly only in complexity and accuracy. The qualitative findings help explain these results by showing that CLIL created a communicative and confidence-building environment through meaningful, interactive, and content-based tasks, which encouraged spontaneity and reduced hesitation in speaking. At the same time, the limited focus on explicit grammar and vocabulary instruction, students' prioritization of content over language precision, and the cognitive overload arising from the dual demands of content learning and language use appeared to constrain the development of complexity and accuracy. Taken together, the findings indicate that CLIL was effective in promoting more fluent and intelligible oral communication, but less effective in fostering linguistically complex and accurate speech without stronger form-focused support.

Discussion

Unexpectedly, the present research revealed that while the CLIL intervention significantly improved speaking fluency and pronunciation, it had limited impact on speaking complexity and accuracy, answering the first research question on the effect of CLIL on English speaking proficiency. These findings diverge from studies such as Hu et al. (2022) and Zanoni (2021), which argue that the dual focus of CLIL on content and language inherently supports both communicative and linguistic development. However, the lack of explicit instruction in grammar and vocabulary observed in the present study highlights a key limitation in CLIL's design, supporting Sysoyev's (2021) assertion that CLIL often prioritizes communicative fluency over linguistic precision.

The improvement in fluency and pronunciation can be attributed to the meaningful communication tasks embedded in the CLIL approach, as part of the findings for the second research question. Students were frequently engaged in discussions, presentations, and collaborative tasks, creating opportunities for spontaneous language use. This aligns with Ikeda et al. (2022), who emphasized the role of interactive and authentic tasks in fostering confidence and fluency. The requirement to communicate effectively in content-related contexts likely motivated students to articulate their ideas clearly, which may explain the significant gains in pronunciation. However, the pronounced improvement in pronunciation in this study compared to previous CLIL research may also reflect task-specific factors (Zhang et al., 2024; Zheng, 2023), such as the need for accurate articulation to convey complex subject content, an aspect often underemphasized in other studies.

The limited development in speaking complexity and accuracy can be interpreted through the lens of cognitive load theory. Coyle and Meyer (2021) highlighted that the dual demands of mastering subject content and producing language in CLIL contexts can overwhelm students'

cognitive capacity. In the current study, students reported focusing their mental resources on understanding and conveying content, leaving little attention for refining grammatical accuracy or expanding syntactic structures. This dual focus likely contributed to their reliance on simpler sentence constructions and familiar vocabulary (Ikeda et al., 2022), as observed in both classroom interactions and interview responses.

Moreover, the absence of targeted feedback on linguistic forms emerged as a critical barrier to improving accuracy and complexity. Cao (2021) noted that while CLIL promotes naturalistic language use, the lack of systematic grammar instruction can lead to persistent errors. The findings from this study echo this limitation, as students expressed feeling less equipped to address grammatical issues due to the minimal emphasis on corrective feedback during the intervention. This gap highlights a potential mismatch between the communicative focus of CLIL (Coyle & Meyer, 2021) and the structured linguistic support required for achieving balanced language development (Mahan, 2023).

These findings suggest that while CLIL effectively fosters fluency and pronunciation, its impact on other aspects of speaking proficiency is limited without deliberate integration of form-focused instruction. This aligns with Coyle and Meyer (2021), who argued that balancing content and language goals requires more than just exposure to communicative tasks; it necessitates explicit attention to linguistic forms within the CLIL framework. The cognitive and pedagogical challenges observed in this study underscore the need for a more integrative approach to CLIL (Hussain, 2022), where communicative and linguistic objectives are harmonized to support comprehensive language development.

Overall, the findings from this study underscore important implications for teachers and curriculum planners involved in CLIL implementation. Teachers need to strike a balance between fostering meaningful communication and addressing linguistic precision. While the emphasis on fluency and pronunciation in content-driven tasks proved effective, the lack of explicit grammar and vocabulary instruction highlighted a critical gap. Teachers should incorporate form-focused activities into CLIL lessons (Klewitz, 2021), such as brief, targeted grammar and vocabulary sessions that complement content learning without disrupting the flow of communication tasks. Curriculum planners should design CLIL programs that explicitly integrate linguistic objectives with content goals (Fan & Zhang, 2024). This includes creating lesson frameworks that allocate time for both subject knowledge and language development. Incorporating linguistic outcomes as part of the broader learning objectives can ensure students receive a more holistic language education. Curriculum planners should also consider developing materials that provide clear guidelines for addressing both content and language, enabling teachers to deliver lessons that support balanced linguistic and cognitive growth. Furthermore, teacher training programs should equip educators with strategies to harmonize content delivery with language instruction. This involves training teachers to identify and address linguistic gaps during content tasks and providing them with resources to scaffold language development effectively (Hussain, 2022). Strengthening this integration will enhance the dual focus of CLIL, ensuring students gain both communicative fluency and linguistic accuracy while mastering subject-specific content.

Conclusion

This study examined the effects of a CLIL intervention on various aspects of English-speaking proficiency, combining quantitative and qualitative findings to provide a comprehensive understanding of its impact. Quantitative results revealed that while the intervention significantly

improved students' speaking fluency and pronunciation, it had limited effects on speaking complexity and accuracy. Students instructed by CLIL demonstrated fewer pauses and repairs in their speech and higher intelligibility, while students following a traditional approach achieved greater gains in syntactic and lexical complexity as well as accuracy. Qualitative findings from classroom observations and interviews highlighted the strengths of CLIL in fostering meaningful communication and building confidence in fluency. However, they also revealed challenges such as the lack of explicit focus on grammar and vocabulary and the cognitive overload experienced by students balancing content comprehension with linguistic precision.

The study makes several important contributions. Theoretically, it extends CLIL scholarship by demonstrating that the effects of CLIL on speaking proficiency are uneven across different dimensions, showing that fluency and pronunciation may develop more readily than complexity and accuracy in content-driven learning environments. In doing so, it offers a more nuanced understanding of CLIL in Chinese tertiary education and highlights the need to conceptualize speaking development as multidimensional rather than as a single overall outcome. Methodologically, the study contributes by adopting a mixed-methods design that combines analytic measures of complexity, accuracy, fluency, and pronunciation with qualitative evidence from classroom observations and interviews, thereby providing both measurable outcomes and explanatory insights into how and why these outcomes emerged. Pedagogically, the findings offer practical guidance for teachers and curriculum planners by showing that CLIL can be effective in promoting confidence, meaningful communication, and oral fluency, but that its benefits can be strengthened through the deliberate integration of form-focused instruction, corrective feedback, and linguistic scaffolding to better support balanced speaking development.

The study has limitations that should be acknowledged. The findings were based on a single educational context and group of participants, which may limit the generalizability of the results to other CLIL settings or learner populations. Additionally, the intervention focused primarily on speaking proficiency, potentially overlooking other aspects of language learning that might benefit from CLIL. Future research should explore CLIL implementations across diverse contexts and include broader linguistic outcomes, such as writing skills, to provide a more comprehensive picture of its effects. Moreover, studies investigating how explicit language instruction can be seamlessly integrated into CLIL without detracting from its content-driven focus would offer valuable insights for improving its efficacy.

Funding: This paper is funded by the Faculty of Education, Universiti Kebangsaan Malaysia (grant code: GG-2024-012).

Acknowledgements: We would like to thank the participants of the study.

Conflict of interests: The authors declare no conflict of interest.

Authenticity: This manuscript is an original research paper.

Artificial Intelligence Statement: AI and AI-assisted technologies have been used to refine language and grammar. The authors take the responsibility for the accuracy and integrity of the content.

About the authors:
Hengzhi Hu is a research fellow and language teacher with a PhD in Teaching English as a Second Language. His research interests include bilingual/multilingual education, innovations in language teaching and learning, language assessment, and curriculums and pedagogies. ORCID: https://orcid.org/0000-0001-5232-913X.

Harwati Hashim is the Deputy Dean of the Faculty of Education, Universiti Kebangsaan Malaysia, with a PhD in Teaching English as a Second Language. Her research interests include innovations in language teaching and learning, teaching English as a second language, and technology-assisted language teaching. ORCID: https://orcid.org/0000-0002-8817-427X.

References

Alonso, R. A. (Ed.). (2018). *Speaking in a second language*. John Benjamins.

Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. Sage.

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: complexity, accuracy and fluency in SLA* (pp. 23–46). John Benjamins.

Cao, Y. (2021). Exploring content and language integrated learning (CLIL) teaching materials in Chinese universities: Teachers' and students' perceptions and reflections. *International Journal of English and Literature*, *12*(3), 23–32. https://doi.org/10.5897/IJEL2021.1444

Chen, Z. (2024). A study of Chinese undergraduate students' English language speaking anxiety, expectancy-value beliefs and spoken English proficiency. *Sage Open*, *14*(1), 1–17. https://doi.org/10.1177/21582440231219312

College Foreign Language Teaching Steering Committee of the Ministry of Education. (2020). *大学英语教学指南 [Guidelines for College English Teaching]*. Higher Education Press.

Costa, F. (2025). *Alternation between L1 (Italian) and L2 (English) in three CLIL and EMI contexts* (2nd ed.). Cambridge Scholars Publishing.

Coyle, D., & Meyer, O. (2021). *Beyond CLIL: Pluriliteracies teaching for deeper learning*. Cambridge University Press. https://doi.org/10.1017/9781108914505

Ding, Y. (2022). Translanguaging in content and language integrated learning (CLIL): Practices in the classroom of a Chinese university. *OLBI Journal*, *12*(1), 311–331. https://doi.org/10.18192/olbij.v12i1.5992

Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford University Press.

Fan, Y., & Zhang, S. (2024). Exploration on the construction of bilingual teacher development system for foreign students in university of traditional Chinese medicine based on the CLIL theory. *Chinese Medicine Modern Distance Education of China*, *22*(7), 27-29. https://doi.org/10.3969/j.issn.1672-2779.2024.07.009

Hidalgo, D. R., & Ortega-Sánchez, D. (2023). CLIL (Content and Language Integrated Learning) methodological approach in the bilingual classroom: A systematic review. *International Journal of Instruction*, *16*(3), 915–934. https://doi.org/10.29333/iji.2023.16349a

Hu, H. (2022). Factors pertinent to first language use in foreign language classroom: A case of content and language integrated learning *Arab World English Journal*, *13*(2), 177–191. https://doi.org/10.24093/awej/vol13no2.12

Hu, H., Said, N. E. M., & Hashim, H. (2022). Killing two birds with one stone? A study on achievement levels and affective factors in content and language integrated learning (CLIL). *International Journal of Learning, Teaching and Educational Research*, *21*(4), 150-167. https://doi.org/10.26803/ijlter.21.4.9

Hu, H., Said, N. E. M., & Hashim, H. (2023). Sustaining content and language integrated learning in China: A systematic review. *Sustainability*, *15*, 5, Article 3894. https://doi.org/10.3390/su15053894

Hussain, S. S. (2022). Content and language integrated learning (CLIL) in ELT as a link between language learning and content development. *Arab World English Journal*, *13*(2), 386–400. https://doi.org/10.24093/awej/vol13no2.26

Ikeda, M., Izumi, S., Watanabe, Y., Pinner, R., & Davis, M. (2022). *Soft CLIL and English language teaching: Understanding Japanese policy, practice and implications*. Routledge.

Joo, M. (2022). Effects of pre-task and on-line planning on complexity, fluency, and accuracy in computer-based English speaking and writing tests. *Korean Journal of English Language and Linguistics*, *22*, 938–956. https://doi.org/10.15738/kjell.22..202210.938

Klewitz, B. (2021). *Content and language integrated learning (CLIL): A methodology of bilingual teaching*. Ibidem Verlag.

Krashen, S. D. (1987). *Principles and practice in second language acquisition*. Pergamon.

Li, S. (2021). The CLIL model of teaching in the College English curriculum in China: Policies, perceptions, and practices. *Onomázein*, 1–9. https://doi.org/10.7764/onomazein.ne9.01

Li, Y. (2020). *Educational change amongst English language college teachers in China: Transitioning from teaching for general to academic purposes*. Springer.

Liao, Y. (2024). An Exploration of Teaching Reform of Integrating Moral Education into English Speaking Course Based on POA. *Journal of Qingdao Ocean Shipping Mariners College*, *45*(1), 73-77. https://doi.org/10.3969/j.issn.2095-3747.2024.01.016

Luoma, S. (2004). *Assessing speaking*. Cambridge University Press.

Mahan, K. R. (2023). Instructional scaffolding in CLIL: An overview of theory and research. In D. L. Banegas & S. Zappa-Hollman (Eds.), *The Routledge handbook of content and language integrated learning* (pp. 284–298). Routledge.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, *30*(4), 555–578. https://doi.org/10.1093/applin/amp044

Peérez, A. C., & Unamuno, V. (2024). The challenges of integrating linguistic and disciplinary knowledge in public secondary schools in the province of Córdoba, Argentina. In E. Codó (Ed.), *Global CLIL: Critical, ethnographic and language policy perspectives* (pp. 96–122). Routledge.

Siskova, Z. (2012). Lexical richness in EFL students' narratives. *Language Studies Working Papers*, *4*(2016), 26–36. https://www.reading.ac.uk/elal/-/media/project/uor-main/schools-departments/elal/lswp/lswp-4/elal_lswp_vol_4_siskova.pdf?la=en&hash=E13FADFF59A93DD925B67C972D8F972B

Skehan, P. (2009). Modelling second language performance: integrating complexity, accuracy, fluency and lexis. *Applied Linguistics*, *30*(4), 510–532. https://doi.org/10.1093/applin/amp047

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Newbury House.

Sylvén, L. K., & Tsuchiya, K. (2023). CLIL in various forms around the world. In D. L. Banegas & S. Zappa-Hollman (Eds.), *The Routledge handbook of content and language integrated learning* (pp. 373–386). Routledge.

Sysoyev, P. V. (2021). Teacher training for content and language integrated learning at the university. *Vysshee Obrazovanie v Rossii, 30*(5), 21–31. https://doi.org/10.31992/0869-3617-2021-30-5-21-31

Teaching Advisory Board in Tertiary Education of the Ministry of Education. (2018). *外国语言文学类教学质量国家标准 [National Standards for Teaching Quality in Foreign Languages and Literature]*. Higher Education Press.

Thomas, C. G. (2021). *Research methodology and scientific writing*. Springer.

Uhl, P., Steinlen, A., & Piske, T. (2023). Elementary school first graders' acquisition of productive L2 French grammar in regular and CLIL programs. *Languages*, *8*(2), Article 138. https://doi.org/10.3390/languages8020138

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

Zanoni, F. (2021). Students' view on CLIL: Perceived benefits and limitation. *EL.LE*, *10*(2), 261–278. https://doi.org/10.30687/ELLE/2280-6792/2021/02/005

Zhang, F. (2024). Contrast between Chinese and Western cultural values and its effects on English learning in China. *Trans/Form/Ação*, *47*(4), Article e0240062. https://doi.org/10.1590/0101-3173.2024.v47.n4.e0240062

Zhang, J., Ismail, H., & Sulaiman, N. A. (2024). Teachers' classroom roles under CLIL: A ten-year systematic review. *World Journal of English Language*, *15*(3), 35–49. https://doi.org/10.5430/wjel.v15n3p35

Zhang, Y. (2022). The application of CLIL approach in EFL writing course integrated with ideological and political elements. *Journal of Advances in Education and Philosophy*, *6*(12), 568–572. https://doi.org/10.36348/jaep.2022.v06i12.002

Zhang, Y. (2023). A Corpus-based NLP Study on Code-switching in English Majors' CLIL Classroom. *Journal of Intelligence and Knowledge Engineering*, *1*(2), 35–42. https://doi.org/10.62517/jike.202304206

Zheng, J. (2023). A comparison of CLIL teaching in China and the United Kingdom—Taking the English language teaching materials of the first grade 6 of SEPE as an example. *Transactions on Comparative Education*, *5*(8), 40–44. https://doi.org/10.23977/trance.2023.050807

Appendices
Appendix A
Mann-Whitney Test of CAFP

Test	Variable	Group	Mean	Standard Deviation	Asymp. Sig. (2-tailed)
Pre-Test	MLAS	EG	9.321	.268	.301
		CG	8.732	.281	
	MLC	EG	5.495	.481	.431
		CG	5.424	.432	
	RS	EG	1.485	.251	.271
		CG	1.472	.322	
	D-score	EG	54.312	1.874	.641
		CG	54.744	2.451	
	PEC	EG	.623	.024	.845
		CG	.615	.022	
	PEA	EG	.556	.012	.745
		CG	.551	.022	
	NP	EG	.365	.123	.810
		CG	.370	.111	
	NR	EG	.056	.003	.401
		CG	.054	.003	
	Intelligibility	EG	5.322	.783	.551
		CG	5.401	.693	
Post-Test	MLAS	EG	9.431	.431	.000
		CG	12.583	.593	
	MLC	EG	6.241	.412	.001
		CG	6.904	.235	
	RS	EG	1.713	.310	.000
		CG	1.945	.365	
	D-score	EG	58.312	2.314	.000
		CG	62.301	2.684	
	PEC	EG	.731	.043	.000
		CG	.824	.033	
	PEA	EG	.712	.022	.000
		CG	.835	.032	

	NP	EG	.214	.124	.000
		CG	.369	.133	
	NR	EG	.038	.004	.000
		CG	.051	.007	
	Intelligibility	EG	7.421	.245	.000
		CG	5.824	.684	

Appendix B: Wilcoxon Signed-Rank Test of CAFP

Group	Variable	Mean Difference (Pre-Post)	N	Asymp. Sig. (2-tailed)
EG	MLAS	-.109	12[a]	.412
			19[b]	
			9[c]	
	MLC	-.746	6[a]	.607
			21[b]	
			13[c]	
	RS	-.228	7[a]	.358
			16[b]	
			17[c]	
	D-score	-4.000	7[a]	.132
			13[b]	
			20[c]	
	PEC	-.108	7[a]	.573
			16[b]	
			17[c]	
	PEA	-.156	5[a]	.422
			14[b]	
			21[c]	
	NP	.051	4[a]	.000
			8[b]	
			28[c]	
	NR	.018	6[a]	.000
			10[b]	
			24[c]	

	Intelligibility	-2.099	5[a]	.000
			11[b]	
			24[c]	
CG	MLAS	-3.851	7[a]	.000
			11[b]	
			24[c]	
	MLC	-1.480	5[a]	.001
			11[b]	
			26[c]	
	RS	-.473	7[a]	.000
			9[b]	
			26[c]	
	D-score	-7.557	6[a]	.000
			10[b]	
			26[c]	
	PEC	-.209	5[a]	.001
			12[b]	
			25[c]	
	PEA	-.284	5[a]	.000
			13[b]	
			24[c]	
	NP	.001	10[a]	.348
			19[b]	
			13[c]	
	NR	.003	8[a]	.631
			26[b]	
			8[c]	
	Intelligibility	-.432	6[a]	.481
			22[b]	
			16[c]	

[a]: pre-test > post-test; [b]: pre-test = post-test; [c]: pre-test < post-test