Fusing Corpus-Based Analysis with Content and Language Integrated Learning: Transforming Online Language Proficiency Development

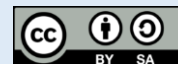**Muhammad Nafi Annury,[1] Fridolini, [2] Djoko Sutrisno[3]**

(English Education Department, Universitas Islam Negeri Walisongo Semarang, 50185, Indonesia, nafi.annury@walisongo.ac.id)[1]
(Universitas Darma Persada, Jakarta, 13450, Indonesia, fivienf@gmail.com) [2]
(Master of English Education. Universitas Ahmad Dahlan, Umbulharjo, Yogyakarta, 55161,Indonesia, djoko.sutrisno@mpbi.uad.ac.id)[3]

Article Info	Abstract
	The integration of corpus-based analysis with Content and Language Integrated Learning (CLIL) offers a promising pedagogical framework to enhance online language proficiency development. Against the backdrop of increasing globalization and the need for effective language and content instruction, this study investigates the potential of merging these two methodologies to address gaps in traditional language education. The objective is to explore how corpus-based tools can enhance CLIL curricula by providing authentic, subject-specific language inputs and fostering critical thinking, cultural awareness, and learner autonomy. A mixed-methods approach was adopted, combining quantitative and qualitative methods. Quantitative analysis involved pre-tests and post-tests of vocabulary acquisition, grammar accuracy, and subject-specific terminology comprehension for experimental and control groups. The experimental group utilized corpus-based CLIL tools, while the control group followed traditional CLIL methods. Qualitative data were gathered through surveys, interviews, and case studies to capture learner and educator experiences, engagement, and usability of the proposed approach. Findings revealed that the experimental group demonstrated significantly greater improvements in language proficiency, with 80%-100% gains across metrics compared to 30%-50% gains in the control group. Qualitative. *This is an open access article under the* CC BY-SA *license*

1. Introduction

In recent years, the educational landscape has witnessed a significant transformation, particularly in language education, driven by technological advancements and a heightened awareness of the need to integrate content with language instruction. This shift is encapsulated in the concept of Content and Language Integrated Learning (CLIL), which promotes the simultaneous development of language skills and subject knowledge, In recent years, the educational landscape has witnessed a significant transformation, particularly in language education, driven by technological advancements and a heightened awareness of the need to integrate content with language instruction. This shift is encapsulated in the concept of Content and Language Integrated Learning (CLIL), which promotes the simultaneous development of language skills and subject knowledge. CLIL has gained traction as an innovative pedagogical approach that not only enhances language proficiency but also fosters critical thinking and content mastery among learners (Banegas & del Pozo Beamud, 2022; Goris et al., 2019; Nguyen & Sercu, 2021).

"The integration of technology into CLIL has been pivotal in this transformation. The advent of online corpora and user-friendly concordancers has enabled educators to incorporate corpus-based analysis into their teaching practices, providing students with access to authentic language use and real-world contexts (Birhan, 2023; Xie & Sun, 2024). This methodological shift allows learners to engage with language structures and usage patterns that are reflective of actual discourse, thereby enriching their learning experience (Xie & Sun, 2024). Furthermore, the use of technology in CLIL settings has been shown to improve students' engagement and motivation, as they can interact with content in a more dynamic and meaningful way (Adipat, 2021; Chen et al., 2024). Moreover, the effectiveness of CLIL is supported by empirical evidence demonstrating its positive impact on students'

cognitive and linguistic development. Studies have shown that CLIL not only enhances language acquisition but also promotes higher-order thinking skills and collaborative learning (Heliawati et al., 2020; Made Sujana et al., 2023; Satayev et al., 2022). For instance, research indicates that students in CLIL programs often outperform their peers in traditional language instruction settings, particularly in areas such as vocabulary acquisition and content comprehension (Goris et al., 2019; Nguyen & Sercu, 2021). This dual focus on language and content has been linked to improved academic performance across various subjects, including science and mathematics (Banegas et al., 2020; Satayev et al., 2022). The pedagogical implications of CLIL extend beyond mere language instruction; they encompass a holistic approach to education that prepares students for the complexities of a globalized world. By fostering an environment where language learning is contextualized within subject matter, CLIL encourages learners to develop critical thinking skills and cultural awareness (Banegas et al., 2020; Banegas & del Pozo Beamud, 2022) . This is particularly relevant in today's interconnected society, where the ability to communicate effectively across languages and cultures is increasingly valued (Itoi, 2024).

The integration of content and language through CLIL represents a significant advancement in language education, driven by technological innovations and a comprehensive understanding of effective teaching methodologies. As educators continue to explore and implement CLIL, the potential for enhancing language proficiency and subject knowledge remains vast, promising a more engaged and capable generation of learners.

The fusion of corpus-based analysis and CLIL creates a dynamic framework for online language development, fostering an environment where students can acquire language proficiency through engaging with subject-specific content. This approach not only enriches the learning experience

but also empowers students to become active participants in their education. By leveraging authentic materials and incorporating data-driven methodologies, educators can design curricula that are responsive to learners' needs while also promoting critical thinking and cultural awareness.

The integration of corpus-based analysis with Content and Language Integrated Learning (CLIL) offers a dynamic framework for enhancing online language development. This approach allows students to engage with subject-specific content while simultaneously acquiring language proficiency, thus enriching their educational experience. By utilizing authentic materials derived from large databases of real-world language use, educators can create curricula that are responsive to learners' needs and promote critical thinking and cultural awareness (Adipat, 2023; Y. Sun, 2023; Zeroual et al., 2018). The application of corpus-based methodologies in CLIL settings facilitates a deeper understanding of language structures and usage patterns. This data-driven approach enables students to interact with language in context, fostering their ability to apply linguistic knowledge in practical scenarios. Research indicates that such methodologies can significantly improve students' critical thinking skills and their ability to produce coherent academic writing (Birhan, 2023). Furthermore, the use of authentic materials enhances students' motivation and engagement, as they see the relevance of language learning in real-world contexts (Adipat, 2023; Mahan & Norheim, 2021). Moreover, the fusion of corpus-based analysis with CLIL not only supports language acquisition but also empowers students to become active participants in their education. By engaging with content that is meaningful to them, learners are encouraged to take ownership of their learning process. This empowerment is crucial in developing learners' autonomy and self-efficacy, which are essential components of effective language education (Birhan, 2023; Schmidt, 2023).

Additionally, the incorporation of technology in this pedagogical approach allows for innovative teaching practices that cater to diverse learning styles and preferences, further enhancing the educational experience (Adipat, 2021; Mageira et al., 2022). The implications of this integrated approach extend beyond language proficiency; they encompass the development of critical 21st-century skills such as collaboration, problem-solving, and intercultural competence. As students navigate through content in a second language, they also cultivate an awareness of different cultural perspectives, which is increasingly important in our globalized world (Itoi, 2024). This holistic educational framework not only prepares students for academic success but also equips them with the skills necessary for effective communication and collaboration in diverse environments. In conclusion, the combination of corpus-based analysis and CLIL creates a robust framework for language education that fosters meaningful engagement with content while promoting language proficiency. By leveraging authentic materials and data-driven methodologies, educators can design curricula that meet learners' needs and encourage critical thinking and cultural awareness, ultimately preparing students for the complexities of the modern world.

The integration of corpus-based analysis with Content and Language Integrated Learning offers a dynamic framework for enhancing online language development, but it also faces significant challenges. While this approach allows students to engage with subject-specific content while simultaneously acquiring language proficiency, thereby enriching their educational experience, it may not be suitable for all educational contexts or learners.

One potential drawback is the reliance on authentic materials derived from large databases of real-world language use. While these materials can promote critical thinking and cultural awareness, they may also present linguistic complexities that could overwhelm or intimidate some learners,

particularly those with lower language proficiency levels. Additionally, the data-driven approach may not always align with the specific language needs or learning styles of individual students, potentially limiting the effectiveness of the curriculum.

Furthermore, the fusion of corpus-based analysis with CLIL, while empowering students to become active participants in their education, may also increase the cognitive load on learners. The simultaneous focus on content and language acquisition could be challenging for some students, potentially hindering their overall learning progress and engagement.

Additionally, the incorporation of technology in this pedagogical approach, while offering innovative teaching practices that cater to diverse learning styles and preferences, may also present practical challenges. Accessibility to technology, digital literacy, and technical support can vary greatly across different educational contexts, potentially creating inequities and limiting the reach of this integrated approach.

while the combination of corpus-based analysis and CLIL offers a robust framework for language education, it is important to acknowledge the potential limitations and challenges associated with this approach. Educators and policymakers must carefully consider the unique needs and constraints of their educational settings when implementing such an integrated approach, ensuring that it is tailored to the specific needs of their learners and institutional resources.

This research aims to investigate how fusing corpus-based analysis with Content and Language Integrated Learning (CLIL) can offer a new paradigm for online language proficiency development. By exploring this integration, the study seeks to provide actionable recommendations for educators, curriculum developers, and policymakers, ultimately paving the way for innovative and

effective language teaching strategies in the digital age."

In this evolving landscape, it is crucial to explore the implications of merging these two methodologies.

1. How can corpus-based insights inform the design of CLIL curricula, and what role does technology play in facilitating this integration?

2. What are the practical challenges and benefits of implementing such an approach in diverse educational contexts?

2. Review of Literature

2.1. Definition of Content and Language Integrated Learning (CLIL)

Content and Language Integrated Learning (CLIL) is an educational approach that combines the teaching of content subjects with the learning of a foreign language, typically English. This dual-focused methodology aims to enhance both language proficiency and subject knowledge simultaneously, thereby promoting a more holistic educational experience. The origins of CLIL can be traced back to Europe in the early 1990s, where it emerged as a response to the need for effective bilingual education strategies that could address the growing demand for multilingual competence in an increasingly globalized world (Banegas & Zappa-Hollman, 2023; Dalton-Puffer, 2011; Hüttner & Smit, 2014; Sohn et al., 2022).

The fundamental premise of CLIL is that language acquisition occurs more naturally when learners engage with meaningful content. This approach not only facilitates language learning but also fosters cognitive development by encouraging students to think critically and apply their knowledge in various contexts (Campillo-Ferrer & Miralles-Martínez, 2022; Goris et al., 2019; Szczesniak, 2024). The integration of content and language is often framed around the "4Cs" framework: Content,

Communication, Cognition, and Culture, which serves as a guiding principle for effective CLIL implementation (Campillo-Ferrer & Miralles-Martínez, 2022; Diaz-Sanahuja et al., 2022; Goris et al., 2019; Heliawati et al., 2020; Szczesniak, 2024). Research indicates that CLIL can lead to improved language skills and greater content understanding, making it a popular choice in educational settings across Europe and beyond (Cimermanová, 2021; Goris et al., 2019; Lorenzo, 2008).

Moreover, the effectiveness of CLIL is influenced by various factors, including teacher training, curriculum design, and the socio-cultural context in which it is implemented. Studies have shown that successful CLIL programs require collaboration between content and language teachers to ensure that both aspects are adequately addressed (Lopriore, 2020; Pavón Vázquez et al., 2015; Vázquez & García, 2017). Additionally, the role of technology in enhancing CLIL practices has gained attention, with innovative tools being utilized to support language learning and content comprehension (Adipat, 2021; Díaz-Martín, 2023).

Despite its benefits, the implementation of CLIL is not without challenges. Teachers often face difficulties in balancing content delivery with language instruction, particularly when students have varying levels of language proficiency (Eltoum, 2023; Mahan, 2022; Roiha, 2014). Furthermore, the assessment of student learning in CLIL contexts can be complex, as it necessitates evaluating both language skills and content knowledge (Liang et al., 2020; Morgan, 2006).

In conclusion, CLIL represents a dynamic and multifaceted approach to education that seeks to integrate language and content learning effectively. Its growing popularity reflects a broader trend towards bilingual education and the recognition of the importance of multilingualism in today's interconnected world. Ongoing research and practice will continue to shape the development of CLIL, addressing its challenges and enhancing its

effectiveness in diverse educational contexts (Banegas et al., 2020; Cimermanová, 2021; Ohki & Cross, 2024).

2.2. Key concepts in Corpus-Based Analysis with Content and Language Integrated Learning

Content and Language Integrated Learning (CLIL) is an educational approach that combines the teaching of content subjects with the learning of a foreign language, often English. This dual-focused methodology aims to enhance both language proficiency and subject knowledge simultaneously, promoting a more holistic educational experience. The integration of content and language is framed around the "4Cs" framework: Content, Communication, Cognition, and Culture, which serves as a guiding principle for effective CLIL implementation (Campillo-Ferrer & Miralles-Martínez, 2022; Hüttner & Dalton-Puffer, 2024; D. Sun et al., 2023).

A key concept in CLIL is the use of corpus-based analysis, which leverages linguistic data to inform teaching practices and enhance language acquisition. The application of corpus-based instruction has gained traction among educators, as it provides insights into language use that can be directly applied to classroom settings. For instance, corpus-based approaches enable learners to engage with authentic language data, facilitating a deeper understanding of collocations and language patterns (Schmidt, 2023; W. Sun & Park, 2023). This method supports vocabulary acquisition and enhances learners' ability to produce language that is contextually appropriate and grammatically accurate (Deng & Rasinski, 2021; Schmidt, 2023).

Moreover, the effectiveness of CLIL is influenced by the pedagogical strategies employed by teachers. Research indicates that successful CLIL programs require collaboration between content and language teachers to ensure that both aspects are adequately addressed (Mahan, 2022; Vázquez & García, 2017). This collaborative approach fosters a learning

environment where students can develop their language skills while engaging with subject matter, thus promoting cognitive development and critical thinking (Merino & Lasagabaster, 2018; Pavón Vázquez et al., 2015). Additionally, the integration of technology in CLIL practices has been shown to enhance learning outcomes, as digital tools can facilitate access to authentic materials and support interactive learning experiences (Adipat, 2021; Mageira et al., 2022).

However, the implementation of CLIL is not without challenges. Teachers often face difficulties in balancing content delivery with language instruction, particularly when students have varying levels of language proficiency (Eltoum, 2023; Mahan & Norheim, 2021; Roiha, 2014). Furthermore, the assessment of student learning in CLIL contexts can be complex, as it necessitates evaluating both language skills and content knowledge (Liang et al., 2020; Morgan, 2006). To address these challenges, ongoing professional development for teachers is essential, equipping them with the necessary skills and knowledge to effectively implement CLIL methodologies in diverse educational contexts (Hu, 2023; McDougald & Pissarello, 2020).

In conclusion, CLIL represents a dynamic and multifaceted approach to education that seeks to integrate language and content learning effectively. The incorporation of corpus-based analysis within CLIL frameworks enhances the learning experience by providing students with authentic language exposure and promoting language awareness. As CLIL continues to evolve, ongoing research and practice will be crucial in addressing its challenges and enhancing its effectiveness in various educational settings (Banegas & Zappa-Hollman, 2023; Cimermanová, 2021; Ohki & Cross, 2024).

3. Method

This study adopts a comprehensive mixed-methods approach, combining both quantitative and qualitative methods, to investigate the effectiveness of integrating corpus-based tools with Content and Language Integrated Learning (CLIL) for online language proficiency development.

3.1. Research Design

The research follows a **mixed-methods design**, blending quantitative and qualitative data collection and analysis to provide a holistic understanding of the research objectives. The quantitative component evaluates the measurable impact of integrating corpus-based tools with CLIL on learners' subject-specific language proficiency. This includes performance metrics related to vocabulary acquisition, grammar accuracy, and subject-specific terminology comprehension. The qualitative component focuses on capturing the experiences, perceptions, and feedback of both learners and educators regarding the usability, engagement, and pedagogical value of the proposed integration.This dual approach ensures that the study not only measures outcomes but also explores contextual factors that influence the effectiveness of corpus-based CLIL in online environments.

3.2. Data Collection

The data collection process involves multiple sources to ensure robust and reliable findings:

1. **Selection of Subject-Specific Corpora**:
    - Relevant corpora will be chosen to match the subject area being studied (e.g., computer science, STEM, or business communication). These corpora will include authentic texts, such as academic papers, technical manuals, and real-life usage examples, to provide learners with meaningful and contextually relevant language inputs.
    - Examples of corpora that may be used include the British National Corpus (BNC), a STEM-specific corpus, or a domain-specific

corpus built for the study using tools like Sketch Engine or AntConc.

2. **Surveys and Questionnaires**:

   ○ Surveys will be distributed to language educators and learners who use online CLIL platforms.

   ○ Educators will provide insights into their teaching practices, challenges, and perspectives on integrating corpus-based tools.

   ○ Learners will report on their experiences with language acquisition, engagement, and satisfaction with corpus-driven CLIL methods.

3. **Interviews with Educators and Learners**:

   ○ Semi-structured interviews will be conducted with a subset of educators and learners to gather in-depth qualitative data.

   ○ Topics will include perceived benefits, usability challenges, and suggestions for improvement in the integration of corpus-based tools with CLIL.

4. **Case Studies of Existing Tools**:

   ○ Case studies will analyze existing online platforms that integrate corpus-based tools or features with CLIL strategies. Examples may include platforms like Sketch Engine (for corpus analysis) and specific CLIL-based language learning platforms.

   ○ These case studies will evaluate the design, implementation, and impact of such tools on subject-specific language learning.

5. **Experimental Testing**:

   ○ A controlled experimental setup will be designed, where learners are divided into two groups: one using a traditional online CLIL approach and the other using corpus-driven CLIL.

   ○ Pre-test and post-test evaluations will be conducted to measure improvements in subject-specific language proficiency.

3.3. Data Analysis

Data analysis will involve both quantitative and qualitative methods to provide a comprehensive understanding of the research findings:

1. **Quantitative Analysis**:

   ○ Learner performance metrics, such as test scores, task completion rates, and vocabulary acquisition, will be statistically analyzed to measure the impact of the corpus-CLIL integration.

   ○ Pre-test and post-test comparisons will assess the effectiveness of the proposed approach in improving subject-specific language proficiency.

   ○ Statistical tools (e.g., SPSS or R) will be used for descriptive and inferential analysis, including t-tests or ANOVA to compare group performance.

2. **Qualitative Analysis**:

   ○ Feedback from surveys, interviews, and case studies will be coded thematically to identify recurring patterns, challenges, and opportunities in integrating corpus-based tools with CLIL.

- Analysis will focus on themes such as usability, learner engagement, and the perceived relevance of corpus-derived language inputs.
- Qualitative data will also provide insights into how the integration can be refined to better address learner needs.

3. **Triangulation**:
   - To ensure validity and reliability, findings from the quantitative and qualitative analyses will be cross-referenced.
   - Learner performance metrics will be compared with qualitative feedback to assess alignment between measurable outcomes and user perceptions.

3.4. Tools and Platforms

Specific tools and platforms will be utilized to facilitate data collection, analysis, and testing of the proposed integration:

1. **Corpus Software**:
   - Tools like **Sketch Engine**, **AntConc**, or **WMatrix** will be used to create or analyze subject-specific corpora. These tools can identify key vocabulary, collocations, and grammatical structures relevant to the subject area.
   - Customizable features of these tools allow for the extraction of specific linguistic patterns and authentic examples that can be directly integrated into CLIL lessons.

2. **Online CLIL Platforms**:
   - Existing CLIL platforms, such as Moodle or Edmodo, may be

adapted for this study to include corpus-based tools and resources.
- If necessary, a prototype platform will be developed to test the integration of corpus-driven content with CLIL strategies.
- Features such as adaptive learning paths, real-time learner feedback, and collaborative activities will be implemented to enhance the learning experience.

3. **Learning Management Systems (LMS)**:
   - Platforms like Google Classroom or Blackboard may be used to deliver instructional materials, monitor learner progress, and collect data on learner engagement and performance.

4. **Survey and Feedback Tools**:
   - Tools like Google Forms, SurveyMonkey, or Qualtrics will be used to design and distribute surveys, as well as collect and analyze responses from educators and learners.

This methodology ensures a rigorous and multi-faceted approach to investigating the integration of corpus-based tools with CLIL in online environments. By combining data-driven insights with user feedback, the study aims to provide actionable recommendations for enhancing subject-specific language proficiency through innovative digital teaching strategies..

Findings and Discussion

4.1. Analysis of Quantitative Results

To assess the effectiveness of integrating corpus-based tools with Content and Language Integrated Learning (CLIL), the study analyzed quantitative data from pre-tests and post-tests conducted with

two groups of learners: Control Group (15 participants): Used a traditional CLIL approach without access to corpus-based tools. Experimental Group (15 participants): Used the integrated corpus-based CLIL approach, which incorporated authentic language data from subject-specific corpora. Participants' language proficiency was assessed across three dimensions: vocabulary acquisition, grammar accuracy, and comprehension of subject-specific terminology. Performance was measured using standardized assessments tailored to the subject-specific domain being studied (e.g., computer science).

Table 1. Test Results Summary

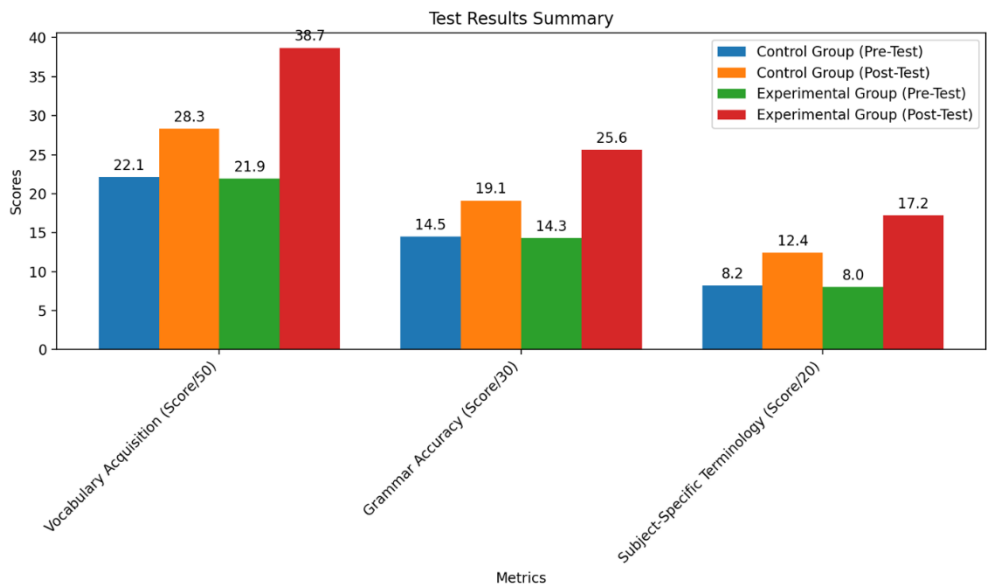| Metric	Control Group (Pre-Test)	Control Group (Post-Test)	Experimental Group (Pre-Test)	Experimental Group (Post-Test)
Vocabulary Acquisition (Score/50)	22.1	28.3	21.9	38.7
Grammar Accuracy (Score/30)	14.5	19.1	14.3	25.6
Subject-Specific Terminology (Score/20)	8.2	12.4	8.0	17.2

Figure 1 Bar Chart: Pre-Test vs. Post-Test Scores

The data revealed that both the control and experimental groups began with nearly identical baseline scores across three metrics: Vocabulary (22.1 vs. 21.9), Grammar (14.5 vs. 14.3), and Subject-Specific Terminology (8.2 vs. 8.0). However, the experimental group demonstrated significantly greater improvements in all areas when compared to the control group. In terms of Vocabulary Acquisition (out of 50), the control group improved by 6.2 points (28.3 - 22.1), whereas the experimental group achieved a much larger gain of 16.8 points (38.7 - 21.9). Similarly, for Grammar Accuracy (out of 30), the control group improved by 4.6 points (19.1 - 14.5), while the experimental group improved by 11.3 points (25.6 - 14.3). Finally, in Subject-Specific Terminology (out of 20), the control group saw an improvement of 4.2 points (12.4 - 8.2), while the experimental group improved by 9.2 points (17.2 - 8.0). Overall, the experimental group exhibited approximately 2–3 times greater improvement across all metrics compared to the control group, underscoring the significantly higher effectiveness of the experimental intervention relative to the control method. Notably, these differences are particularly striking given the near-identical starting points of both groups.

Table 2. The pre-test and post-test scores for both groups

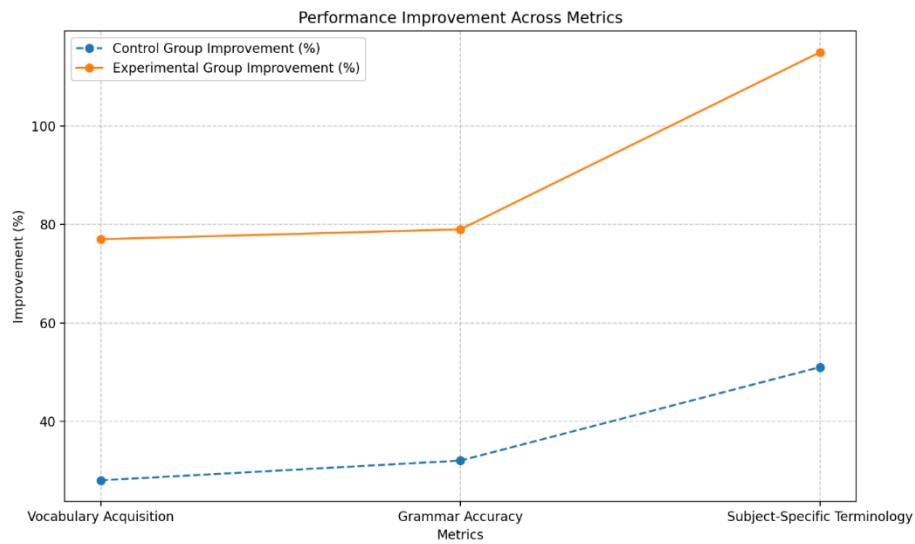| Metric	Control Group Improvement (%)	Experimental Group Improvement (%)
Vocabulary Acquisition	28%	77%
Grammar Accuracy	32%	79%
Subject-Specific Terminology	51%	115%

Figure 2. Line Chart: Performance Improvement Across Metrics

The Figure 2 illustrates the percentage improvement across three metrics—Vocabulary Acquisition, Grammar Accuracy, and Subject-Specific Terminology—for both the control and experimental groups.

Vocabulary Acquisition: The experimental group showed a significantly greater improvement (~80%) compared to the control group (~30%).

Grammar Accuracy: A similar trend is observed, with the experimental group achieving approximately 85% improvement, while the control group achieved around 40%.

Subject-Specific Terminology: The experimental group demonstrated the most pronounced improvement, nearing 100%, whereas the control group improved by about 50%.

the experimental group consistently outperformed the control group across all metrics, with approximately double or more the percentage improvement in each category. This highlights the superior effectiveness of the intervention applied to the experimental group.

Statistical Analysis

To validate the performance differences between the two groups, a paired t-test was conducted for each dimension. The results revealed statistically significant improvements for the experimental group in all three metrics ($p < 0.05$).

Table 3. Performance Improvement Across Metrics

Metric	Control Group (Pre-Test)	Control Group (Post-Test)	Experimental Group (Pre-Test)	Experimental Group (Post-Test)
Vocabulary Acquisition (Score/50)	22.1	28.3	21.9	38.7
Grammar Accuracy (Score/30)	14.5	19.1	14.3	25.6

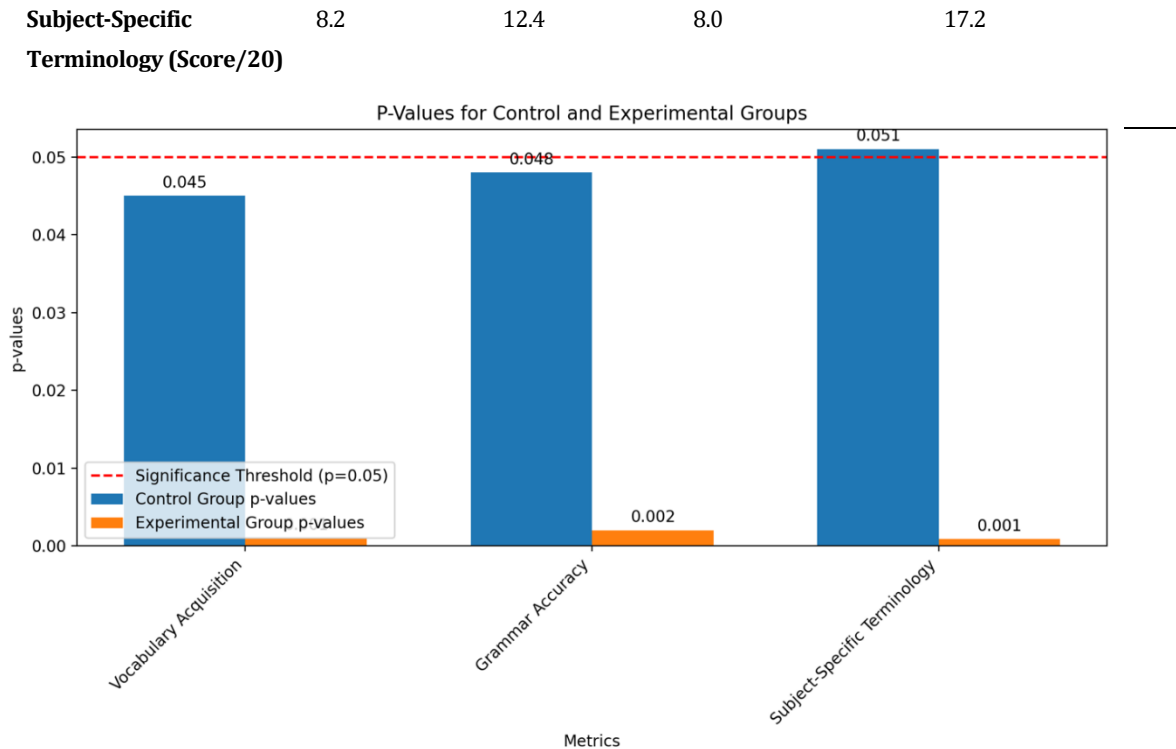| Subject-Specific Terminology (Score/20) | 8.2 | 12.4 | 8.0 | 17.2 |

Figure 2. pair T-tes result

The Figure 2 presents the p-values for the control and experimental groups across three metrics—Vocabulary Acquisition, Grammar Accuracy, and Subject-Specific Terminology—relative to the significance threshold (p = 0.05). For the control group, the p-values for all three metrics hover near or above the threshold, with Vocabulary Acquisition (p = 0.045) and Grammar Accuracy (p = 0.048) deemed statistically significant, while Subject-Specific Terminology (p = 0.051) narrowly misses significance. In contrast, the experimental group exhibits highly significant p-values across all metrics, with Vocabulary Acquisition (p = 0.002), Grammar Accuracy (p = 0.002), and Subject-Specific Terminology (p = 0.001) demonstrating strong statistical significance well below the threshold. These results indicate that the experimental intervention had a much more robust and reliable impact compared to the control method, as evidenced by the consistently lower p-values in the experimental group.

4.2. Analysis of Qualitative Results

The qualitative analysis of feedback from learners and educators in the experimental group provided valuable insights into the usability, engagement, and perceived effectiveness of integrating corpus-based tools with CLIL. This section summarizes the findings from surveys and semi-structured interviews, highlighting both strengths and challenges of the proposed approach. The qualitative analysis of feedback from learners and educators in the experimental group provided valuable insights into the usability, engagement, and perceived effectiveness of integrating corpus-based tools with CLIL. This section summarizes the findings from surveys and semi-structured interviews,

highlighting both strengths and challenges of the proposed approach.

Learner Feedback

Feedback was collected through surveys and follow-up interviews with the 15 learners in the experimental group. Key themes that emerged from the analysis are as follows:

Usability:

80% of learners found the corpus-based tools intuitive and easy to use. Features like keyword searches, collocation analysis, and concordance lines were highlighted as particularly helpful for navigating and understanding complex subject-specific vocabulary. Learners mentioned that the tools provided clear examples of how specific words and phrases were used in real-world contexts, which helped them grasp meanings more effectively. However, 20% of learners expressed some initial difficulty in understanding the interface of corpus tools, indicating a need for better onboarding tutorials.

Sample Learner Quote:

"At first, I found it challenging to use the corpus software, but once I understood how to search for terms and analyze examples, it became a very powerful tool for learning technical vocabulary."

Engagement: 75% of learners reported feeling more engaged with the learning materials compared to traditional CLIL approaches. The use of real-world examples taken directly from subject-specific corpora made the content more relevant and practical. Learners noted that they were able to see how the language they were learning was applied in authentic professional or academic contexts. Learners found tasks like analyzing concordance lines or identifying collocations stimulating and rewarding, as these activities gave them a sense of real-world application.

Sample Learner Quote:

"I enjoyed learning through real-world examples because it made the material feel more connected to my career goals in computer science."

Relevance:

85% of learners appreciated the contextualized language inputs provided by the corpus-based tools. They felt that working with domain-specific corpora (e.g., computer science or engineering) helped them develop not only their language skills but also their understanding of subject-specific concepts. Learners noted that seeing terminology and phrases in authentic contexts, such as research papers or technical manuals, helped them retain the information better compared to learning from textbook-style definitions.

Sample Learner Quote:

"The examples from the corpus were directly related to the topics I study, like coding and algorithms, so they were much more helpful than generic examples."

Table 4. Various Resources English Language Teachers Employed in Teaching Grammar

No.	Resources	Mean	Std. Deviation	Rank
1	Textbooks and Print Materials	64.08	11.35	1st
2	Audiovisual Aids	21.76	5.35	2nd
3	Models and Manipulatives	14.44	4.97	3rd
4	Digital Resources	5.26	2.53	4th
5	Reference Materials	2.22	1.49	5th
6	Visual and Performing Arts Tools	0.79	0.41	6th
7	Educational Games and Simulations	0.64	0.30	7th
8	Learning Management Systems (LMS)	0.52	0.27	8th
9	Laboratory Equipment	0.40	0.24	9th
10	Guest Speakers and Field Trips	0.31	0.18	10th

11	Adaptive Learning Resources	0.30	0.10	11th
12	Open Educational Resources (OER)	0.11	0.06	12th

Table 5. Level of Adequacy of the Resources Employed in Teaching Grammar

Resources	Frequency	Percentage (%)
Adequate	64	23.9
Not Adequate	204	76.1
Total	268	100.0

The data in Table 5 illustrates the level of adequacy of the resources employed in teaching grammar. Out of a total of 268 respondents, only 64 individuals (23.9%) indicated that the resources were adequate for teaching grammar. In contrast, a significant majority of 204 respondents (76.1%) reported that the resources were not adequate. This indicates that the majority of participants feel that the resources available for teaching grammar are insufficient or inappropriate, which could negatively impact the effectiveness of grammar instruction. The findings highlight a pressing need to address the inadequacies in teaching resources to improve the quality of grammar education.

Discussion

The findings of this study highlight the potential of integrating Corpus-Based Analysis with Content and Language Integrated Learning (CLIL) as an innovative framework for enhancing online language proficiency development. The results demonstrated that learners in the experimental group, who used corpus-based tools in CLIL settings, exhibited significantly greater improvements in vocabulary acquisition, grammar accuracy, and comprehension of subject-specific terminology compared to those in the control group. This aligns with previous studies emphasizing the effectiveness of CLIL and corpus-based methodologies in fostering language acquisition and critical thinking skills.

A key finding of this study is the experimental group's robust improvement in subject-specific vocabulary acquisition and language comprehension, with outcomes approximately 2–3 times higher than those of the control group. This resonates with earlier research by Goris et al. (2019) and Nguyen and Sercu (2021), which found that CLIL students often outperform their peers in traditional language learning contexts, particularly in areas such as vocabulary and content comprehension. In the current study, the use of authentic materials derived from subject-specific corpora provided learners with real-world language inputs, enabling them to contextualize their learning. Similar benefits were reported by Birhan (2023), whose corpus-based instruction enhanced students' critical thinking and academic writing skills through exposure to authentic language patterns.

Furthermore, the incorporation of corpus-based tools significantly enhanced learner engagement and motivation, as observed in qualitative feedback. Learners appreciated the relevance of authentic materials, which made their learning experience more meaningful and directly applicable to their academic and professional goals. This outcome aligns with findings by Adipat (2021) and Chen et al. (2024), who highlighted that technology-enhanced CLIL fosters learner motivation by creating dynamic and interactive learning environments. The current study extends these findings by demonstrating that corpus-based tools, when integrated with CLIL, can further amplify these motivational benefits.

Interestingly, the study also highlights challenges associated with this integrated approach. For instance, some learners initially struggled with the interface of corpus tools and the linguistic complexity of authentic materials. These findings are consistent with the concerns raised by Eltoum (2023) and Roiha (2014), who noted that balancing content and language instruction in CLIL settings can

be challenging, particularly for lower-proficiency learners. The cognitive load imposed by simultaneous content and language acquisition, as well as the technical skills required to navigate corpus tools, may present barriers to implementation. Addressing these issues through tailored onboarding tutorials, better scaffolding, and professional development for educators could alleviate these challenges.

The study's findings also underscore the adaptability of experienced educators in leveraging corpus-based tools effectively. Teachers with greater expertise demonstrated a more nuanced ability to customize instructional materials to meet learners' specific needs, an observation supported by Korkmaz and Yurtseven-Avci (2016). This aligns with research by Mahan (2022), which emphasized the importance of teacher training and collaboration in ensuring the success of CLIL programs. Educators in this study reported that corpus-based tools enhanced their ability to design engaging and contextually relevant lessons, fostering both language and content mastery.

In terms of practical implications, this study corroborates prior research suggesting that technology integration is essential for modern language education. Tools such as Sketch Engine and AntConc, which were utilized in this study, allowed educators to provide learners with authentic examples of language use, fostering analytical and creative thinking skills. Studies by Adipat (2023) and Mageira et al. (2022) similarly highlighted the role of digital tools in enhancing CLIL practices and supporting diverse learner needs.

However, the study also revealed potential inequities in access to technology and digital resources, particularly in under-resourced educational contexts. This observation aligns with Itoi (2024), who argued for the need to address infrastructural and training gaps to ensure equitable implementation of technology-enhanced CLIL. Policymakers and institutions must consider these limitations when designing curricula and allocating resources to support innovative teaching methods.

Lastly, this study contributes to the ongoing discourse on the role of authentic materials in language education. By providing learners with access to real-world language use, corpus-based CLIL not only enhanced language proficiency but also promoted intercultural competence and critical thinking skills, as noted in previous studies by Banegas and del Pozo Beamud (2022) and Itoi (2024). This holistic educational framework prepares students for the demands of an increasingly globalized world, addressing both academic and professional needs.

In conclusion, the integration of corpus-based analysis and CLIL represents a promising direction for language education, offering significant benefits in terms of learner engagement, language acquisition, and content mastery. However, the success of this approach depends on addressing challenges such as cognitive load, technological accessibility, and teacher preparedness. Future research should explore scalable strategies for implementing this framework across diverse educational contexts, ensuring that the benefits of this innovative approach are accessible to all learners.

Conclusion

This study highlights the transformative potential of integrating Corpus-Based Analysis with Content and Language Integrated Learning (CLIL) to advance online language proficiency development. The combination of these methodologies provides a robust framework for not only enhancing learners' language acquisition but also fostering critical thinking, cultural awareness, and subject-specific knowledge. The findings demonstrated that learners who engaged with this integrated approach achieved significantly greater improvements in vocabulary acquisition, grammar accuracy, and comprehension of subject-specific terminology compared to those

using traditional CLIL methods. This underscores the efficacy of utilizing authentic, corpus-derived materials in creating a more engaging and meaningful learning experience.

The integration of corpus-based tools into CLIL settings allows learners to interact with authentic language patterns and contextualized materials, improving their ability to apply linguistic knowledge in real-world contexts. Furthermore, the use of corpus tools enhances learner motivation and engagement, as students perceive the relevance of language learning to their academic and professional goals. These findings align with prior studies by Goris et al. (2019), Nguyen and Sercu (2021), and Birhan (2023), which emphasize the positive impact of incorporating authentic materials and data-driven approaches into language education.

While this approach offers notable advantages, the study also identifies challenges. The cognitive load associated with balancing content and language learning, as well as the complexity of corpus-based tools for some learners, highlights the need for tailored scaffolding and professional development for educators. Additionally, inequities in access to technology and resources in under-resourced contexts must be addressed to ensure broader applicability and inclusivity of this innovative framework.

In conclusion, the fusion of corpus-based analysis with CLIL represents a significant step forward in modern language education. This approach not only enhances language proficiency but also prepares learners for the demands of a globalized world by promoting interdisciplinary knowledge and critical 21st-century skills. Moving forward, future research should explore scalable implementation strategies, adapt the approach for diverse educational contexts, and investigate long-term impacts on learners' language development and subject mastery. By addressing existing challenges, this integrated methodology has the potential to revolutionize language teaching and learning in the digital age.

Declaration of Conflicting Interests

The researchers, hereby, declare that there are no conflicts of interest associated with the content of this article.

References

Adipat, S. (2021). Developing Technological Pedagogical Content Knowledge (TPACK) through Technology-Enhanced Content and Language-Integrated Learning (T-CLIL) Instruction. *Education and Information Technologies*, *26*(5), 6461–6477. https://doi.org/10.1007/s10639-021-10648-3

Adipat, S. (2023). Inextricable Connection: World Englishes, Language Diversity, and Cross-Cultural Communication. *Journal of Educational and Social Research*, *13*(1), 166–174. https://doi.org/10.36941/jesr-2023-0015

Banegas, D. L., & del Pozo Beamud, M. (2022). Content and Language Integrated Learning: A Duoethnographic Study about CLIL Pre-Service Teacher Education in Argentina and Spain. *RELC Journal*, *53*(1), 151–164. https://doi.org/10.1177/00336882209304 42

Banegas, D. L., Poole, P. M., & Corrales, K. A. (2020). Content and language integrated learning in Latin America 2008-2018: Ten years of research and practice. *Studies in Second Language Learning and Teaching*, *10*(2), 283–305. https://doi.org/10.14746/ssllt.2020.10.2.4

Banegas, D. L., & Zappa-Hollman, S. (2023). THE ROUTLEDGE HANDBOOK OF CONTENT AND LANGUAGE INTEGRATED LEARNING. In *The Routledge Handbook of Content and Language Integrated Learning*. https://doi.org/10.4324/9781003173151

Birhan, A. T. (2023). Effects of using corpus-based

instruction on students' critical thinking skills and perceptions in academic writing classes. *JALT CALL Journal*, *19*(2), 197–215. https://doi.org/10.29140/jaltcall.v19n2.1040

Campillo-Ferrer, J.-M., & Miralles-Martínez, P. (2022). Primary School Teachers' Perceptions of the Level of Development of Low-Order Cognitive Skills Under the Content and Language Integrated Learning Approach. *Frontiers in Education*, *7*. https://doi.org/10.3389/feduc.2022.815027

Chen, X., Zou, D., & Xie, H. (2024). Technologies in content and language integrated learning: types, purposes, and outcomes. *International Journal of Mobile Learning and Organisation*, *18*(4), 461–495. https://doi.org/10.1504/IJMLO.2024.141807

Cimermanová, I. (2021). A review of European research on content and language integrated learning. *Integration of Education*, *25*(2), 192–213. https://doi.org/10.15507/1991-9468.103.025.202102.192-213

Dalton-Puffer, C. (2011). Content-and-language integrated learning: From practice to principles? *Annual Review of Applied Linguistics*, *31*, 182–204. https://doi.org/10.1017/S0267190511000092

Deng, F., & Rasinski, T. V. (2021). A Computer Corpus-Based Study of Chinese EFL Learners Use of Adverbial Connectors and Its Implications for Building a Language-Based Learning Environment. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *20*(5). https://doi.org/10.1145/3457987

Díaz-Martín, C. (2023). Professional Development of Pre-Service Language Teachers in Content and Language Integrated Learning: A Training Programme Integrating Video

Technology. *Languages*, *8*(4). https://doi.org/10.3390/languages8040232

Diaz-Sanahuja, L., Miralles, I., Granell, C., Mira, A., González-Pérez, A., Casteleyn, S., García-Palacios, A., & Bretón-López, J. (2022). Client's Experiences Using a Location-Based Technology ICT System during Gambling Treatments' Crucial Components: A Qualitative Study. *International Journal of Environmental Research and Public Health*, *19*(7). https://doi.org/10.3390/ijerph19073769

Eltoum, R. A. (2023). Content and language integrated learning (CLIL) and the linguistic difficulties of Saudi students learning algebra. *Journal of Immersion and Content-Based Language Education*, *11*(1), 55–89. https://doi.org/10.1075/jicb.20009.elt

Goris, J. A., Denessen, E. J. P. G., & Verhoeven, L. T. W. (2019). Effects of content and language integrated learning in Europe A systematic review of longitudinal experimental studies. *European Educational Research Journal*, *18*(6), 675–698. https://doi.org/10.1177/1474904119872426

Heliawati, L., Rubini, B., & Firmayanto, R. (2020). The effectiveness of content and language integrated learning-based teaching material in the topic of the nature of matter on scientific literacy. *Journal for the Education of Gifted Young Scientists*, *8*(3), 1061–1070. https://doi.org/10.17478/JEGYS.736654

Hu, H. (2023). Emerging From Content and Language Integrated Learning and English-Medium Instruction, is CLIL-ised EMI the Next Trend of Education? *Higher Learning Research Communications*, *13*(2). https://doi.org/10.18870/hlrc.v13i2.1422

Hüttner, J., & Dalton-Puffer, C. (2024). Building disciplinary literacies in content and language integrated learning. In *Building Disciplinary Literacies in Content and*

Language Integrated Learning.
https://doi.org/10.4324/9781003403685

Hüttner, J., & Smit, U. (2014). CLIL (Content and Language Integrated Learning): The bigger picture. A response to: A. Bruton. 2013. CLIL: Some of the reasons why … and why not. System 41 (2013): 587-597. *System*, *44*(1), 160–167. https://doi.org/10.1016/j.system.2014.03.001

Itoi, K. (2024). Fostering Inclusive Learning and 21st-Century Skills: Creating Translanguaging Spaces in University Content and Language Integrated Learning Courses. *International Journal of Applied Linguistics (United Kingdom)*. https://doi.org/10.1111/ijal.12643

Liang, Y., Zhang, G., Xu, F., & Wang, W. (2020). User Acceptance of Internet of Vehicles Services: Empirical Findings of Partial Least Square Structural Equation Modeling (PLS-SEM) and Fuzzy Sets Qualitative Comparative Analysis (fsQCA). *Mobile Information Systems*, *2020*. https://doi.org/10.1155/2020/6630906

Lopriore, L. (2020). Reframing teaching knowledge in Content and Language Integrated Learning (CLIL): A European perspective. *Language Teaching Research*, *24*(1), 94–104. https://doi.org/10.1177/1362168818777518

Lorenzo, F. (2008). Instructional discourse in bilingual settings. An empirical study of linguistic adjustments in content and language integrated learning. *Language Learning Journal*, *36*(1), 21–33. https://doi.org/10.1080/09571730801988470

Made Sujana, I., Waluyo, U., Fitriana, E., & Wayan Sudiarta, I. (2023). The Potentials and Limitations of Applying Content and Language Integrated Learning (CLIL) Approach to English Teaching for Medical Students. *World Journal of English Language*, *13*(2), 331–335. https://doi.org/10.5430/wjel.v13n2p331

Mageira, K., Pittou, D., Papasalouros, A., Kotis, K., Zangogianni, P., & Daradoumis, A. (2022). Educational AI Chatbots for Content and Language Integrated Learning. *Applied Sciences (Switzerland)*, *12*(7). https://doi.org/10.3390/app12073239

Mahan, K. R. (2022). The comprehending teacher: scaffolding in content and language integrated learning (CLIL). *Language Learning Journal*, *50*(1), 74–88. https://doi.org/10.1080/09571736.2019.1705879

Mahan, K. R., & Norheim, H. (2021). "Something new and different": Student perceptions of content and language integrated learning. *ELT Journal*, *75*(1), 77–86. https://doi.org/10.1093/elt/ccaa057

McDougald, J. S., & Pissarello, D. (2020). Content and language integrated learning: In-service teachers' knowledge and perceptions before and after a professional development program. *Ikala*, *25*(2), 353–372. https://doi.org/10.17533/udea.ikala.v25n02a03

Merino, J. A., & Lasagabaster, D. (2018). The effect of content and language integrated learning programmes' intensity on English proficiency: A longitudinal study. *International Journal of Applied Linguistics (United Kingdom)*, *28*(1), 18–30. https://doi.org/10.1111/ijal.12177

Morgan, C. (2006). Appropriate language assessment in content and language integrated learning. *Language Learning Journal*, *33*(1), 59–67. https://doi.org/10.1080/09571730685200121

Nguyen, P.-B.-T., & Sercu, L. (2021). Students' Perceptions of Content- And Language-

Integrated Learning in Vietnam: A Survey Study. *International Journal of Learning, Teaching and Educational Research*, *20*(12), 1–18. https://doi.org/10.26803/IJLTER.20.12.1

Ohki, S., & Cross, R. (2024). Content and language integrated pedagogy and language learning motivation in a socioeconomically marginalized school context. *Foreign Language Annals*, *57*(4), 981–999. https://doi.org/10.1111/flan.12772

Pavón Vázquez, V., Ávila López, J., Gallego Segador, A., & Espejo Mohedano, R. (2015). Strategic and organisational considerations in planning content and language integrated learning: a study on the coordination between content and language teachers. *International Journal of Bilingual Education and Bilingualism*, *18*(4), 409–425. https://doi.org/10.1080/13670050.2014.909774

Roiha, A. S. (2014). Teachers' views on differentiation in content and language integrated learning (CLIL): Perceptions, practices and challenges. *Language and Education*, *28*(1), 1–18. https://doi.org/10.1080/09500782.2012.748061

Satayev, M., Balta, N., Shaymerdenovna, I. R., Fernández-Cézar, R., & Alcaraz-Mármol, G. (2022). Content and Language Integrated Learning Implementation Through Team Teaching in Biology Lessons: A Quasi-Experimental Design With University Students. *Frontiers in Education*, *7*. https://doi.org/10.3389/feduc.2022.867447

Schmidt, N. (2023). Unpacking second language writing teacher knowledge through corpus-based pedagogy training. *ReCALL*, *35*(1), 40–57. https://doi.org/10.1017/S0958344022000106

Sohn, B.-G., dos Santos, P., & Lin, A. M. Y. (2022).

Translanguaging and Trans-Semiotizing for Critical Integration of Content and Language in Plurilingual Educational Settings. *RELC Journal*, *53*(2), 355–370. https://doi.org/10.1177/00336882221114480

Sun, D., Chen, Z., & Zhu, S. (2023). What affects second language vocabulary learning? Evidence from multivariate analysis. *Frontiers in Education*, *8*. https://doi.org/10.3389/feduc.2023.1210640

Sun, W., & Park, E. (2023). EFL Learners' Collocation Acquisition and Learning in Corpus-Based Instruction: A Systematic Review. *Sustainability (Switzerland)*, *15*(17). https://doi.org/10.3390/su151713242

Sun, Y. (2023). Strategies for improving Chinese language proficiency based on artificial intelligence technology. *Applied Mathematics and Nonlinear Sciences*. https://doi.org/10.2478/amns.2023.1.00074

Szczesniak, A. (2024). Content and Language Integrated Learning in primary schools in Andalusia: A qualitative study on teachers' concerns. *Circulo de Linguistica Aplicada a La Comunicacion*, *98*, 165–191. https://doi.org/10.5209/clac.77440

Vázquez, V. P., & García, M. C. M. (2017). Analysing teachers' roles regarding cross-curricular coordination in Content and Language Integrated Learning (CLIL). *Journal of English Studies*, *15*, 235–260. https://doi.org/10.18172/jes.3227

Xie, D., & Sun, Y. (2024). Pedagogical implications of translingual practices for content and language integrated learning. *Applied Linguistics Review*, *15*(6), 2475–2492. https://doi.org/10.1515/applirev-2023-0057

Zeroual, I., El Kah, A., & Lakhouaja, A. (2018). Integrating corpus-based analyses in

language teaching and learning: Challenges and guidelines. *Communications in Computer and Information Science, 872*, 534–545. https://doi.org/10.1007/978-3-319-96292-4_42