International Journal of Arabic Language Teaching
P-ISSN: 2684-690X E-ISSN: 2686-214X
https://e-journal.metrouniv.ac.id/index.php/IJALT

Beyond Grammar-Translation: CLIL as a Communicative Strategy for Arabic Speaking Development at Madrasah Tsanawiyah Level

Nazwa Qomariyah[1]*, Ro'fat Hizmatul Himmah[2]
[1, 2]Universitas KH. Mukhtar Syafaat Banyuwangi, Indonesia

Article History:
Received : 14 February 2026
Revised : 11 May 2026
Published : 11 May 2026

Keywords:
CLIL; Contextual Fluency; Arabic Speaking Skills.

***Correspondence Address:**
nazwaqomariyah18@gmail.com

DOI: 10.32332/ijalt.v8i01.13057

***Abstract*:** The development of Arabic speaking skills in madrasahs is often hindered by a theoretical teaching pattern with minimal communicative practice. This study aims to analyze the effectiveness of the Content and Language Integrated Learning (CLIL) approach on the speaking skills of eighth-grade students at MTsN 9 Banyuwangi. Using a quasi-experimental quantitative method thru a pretest-posttest control group design, this study involved 50 students as the sample. Data were collected via oral tests using an analytic rubric, then analyzed with an independent t-test and N-Gain test. The research results showed a significant difference in posttest scores ($p < 0.001$), where the experimental group achieved a high category increase (N-Gain 0.7105), surpassing the control group which was in the moderate category (0.3703). The novelty of this research lies in the modification of content-based oral production tasks through the 4C framework to stimulate contextual fluency. This research suggests the implementation of CLIL as an innovative strategy in Arabic language learning at madrasahs, noting the need for further studies on its long-term impact.

INTRODUCTION

Although Arabic has been studied for years in Islamic educational institutions,[1] The majority of madrasah graduates still experience "communicative paralysis" where they are able to master complex grammar rules but fail in simple everyday conversations.[2] This real-world problem is rooted in the teaching patterns at madrasahs that are still trapped in the dominance of rigid grammatical theory (*qawa'id*), thereby marginalizing the space for oral practice for students. Efforts to reform grammar teaching have actually begun to be explored thru the application of the Content and Language Integrated Learning (CLIL) approach in

[1] Khalid Arar, Rania Sawalhi, and Munube Yilmaz, "The Research on Islamic-Based Educational Leadership since 1990: An International Review of Empirical Evidence and a Future Research Agenda," *Religions* 13, no. 1 (2022): 42. https://doi.org/10.3390/rel13010042

[2] Jimmy Van Rijt, Arina Banga, and Martijn Goudbeek, "Getting a Load of Linguistic Reasoning: How L1 Student Teachers Process Rules of Thumb and Linguistic Manipulations in Discussions about Grammar," *Applied Linguistics* 45, no. 1 (2024): 163–88. https://doi.org/10.1093/applin/amad011

the teaching of *qawa'id* based on the yellow book in *pesantren*.[3] However, the fact is that the focus on strengthening those principles often does not correspond with the improvement in students' oral production skills at the madrasah level.

Previous research has indeed attempted to apply various communicative methods, but often still separates the mastery of subject matter from language skills themselves.[4] The focus of this research is to evaluate the impact of the Content and Language Integrated Learning (CLIL) approach on the speaking proficiency of eighth-grade students at MTsN 9 Banyuwangi. The research problem raised is to what extent the integration of content material with the target language can bridge the speaking barriers of students. The scope of the study is limited to the theme of *A'māl al-Yaumiyyah* (daily activities) involving 50 students in a quasi-experimental design. Although this study holds significant value as a pedagogical innovation, there are limitations regarding the sample size, which is specific to one institution, and the relatively short duration of the intervention, which consists of eight meetings.

A review of previous works shows that CLIL has been highly successful in English language learning to enhance cognitive and motivational aspects, but its implementation in the Arabic language curriculum,[5] Several studies have begun to explore the implementation and challenges of this approach in *pesantren* environments to improve language proficiency. However, the specific implementation for producing oral content in the *Madrasah Tsanawiyah* curriculum remains a very rarely touched area.[6] Here lies the position and contribution of this research; the researcher identifies a gap in the literature regarding the lack of systematic content-based oral production task models for madrasah students. This research offers novelty thru the modification of oral tasks based on the 4C framework (Content, Communication, Cognition, Culture) to stimulate contextual fluency.

Conceptually, this research is based on the 4C theory framework by Do Coyle (1999), which views language not merely as an object of study but as an

[3] Salma Salsabila et al., "Content and Language Integrated Learning Approach in Qawāʿid Learning through Yellow Books in Islamic Boarding Schools," *International Journal of Arabic Language Teaching* 7, no. 01 (2025): 193–206. https://doi.org/10.32332/ijalt.v7i01.10991

[4] Jeremy Lamri and Todd Lubart, "Reconciling Hard Skills and Soft Skills in a Common Framework: The Generic Skills Component Approach," *Journal of Intelligence* 11, no. 6 (2023): 107. https://doi.org/10.3390/jintelligence11060107

[5] Hengzhi Hu, Nur Ehsan Mohd Said, and Harwati Hashim, "Sustaining Content and Language Integrated Learning in China: A Systematic Review," *Sustainability* 15, no. 5 (2023): 3894. https://doi.org/10.3390/su15053894

[6] Muhammad Iqbal, Mohammad Ahsanuddin, and Muhammad Alfan, "Enhancing Arabic Proficiency through Content and Language Integrated Learning: Implementation and Challenges in Pesantren," *International Journal of Arabic Language Teaching* 7, no. 02 (2025): 207–22. https://doi.org/10.32332/ijalt.v7i02.10769

active instrument for understanding content.[7] The main argument put forward is that the direct integration of subject content into language practice provides a more meaningful and authentic context for students compared to conventional methods that tend to be theoretical.[8] The findings of this study clearly indicate that the approach is capable of significantly accelerating oral proficiency, where students show a much more optimal increase in fluency and confidence after engaging in this content and language integrated learning.[9]

The structure of this paper is arranged in a logical order to facilitate the reader's understanding. After this introduction section, the second part will review the quantitative research methods, purposive sampling procedures, and the oral test instruments used. The third section will present the research data thru statistical tests (normality, homogeneity, and t-test) accompanied by an in-depth discussion on the effectiveness of CLIL. Finally, the fourth section will conclude the research findings, acknowledge the study's limitations, and provide practical suggestions for future Arabic language teaching.

METHOD

This research uses a quantitative approach with a quasi-experimental design of the pretest-posttest control group type, aiming to determine the effect of implementing Content and Language Integrated Learning (CLIL) on improving students' speaking skills. The quasi-experimental design was chosen because the researcher was unable to fully randomize the subjects, but could still objectively compare the effects of the treatment through the experimental group and the control group.[10] The research was conducted at MTsN 9 Banyuwangi, focusing on the material of daily activities (*a'mālul yaumiyyah*) because this topic is close to the students' daily lives and relevant for developing contextual speaking skills.

The study population for Class VIII of MTsN 9 Banyuwangi consists of 7 classes. However, the sampling technique used was purposive sampling, so only two classes were selected as the research sample. The selection of the two classes was based on considerations of initial ability equivalence, relatively homogeneous

[7] Jieting Jerry Xin, Daniel Fung, and Nicole Judith Tavares, "The 4Cs in Action: Evaluating Language-Driven CLIL Materials for Holistic Learning Experiences," *The Language Learning Journal*, 2025, 1–20, https://doi.org/10.1080/09571736.2025.2494017

[8] Mohammad H Al-Khresheh, "The Role of Presentation-Based Activities in Enhancing Speaking Proficiency among Saudi EFL Students: A Quasi-Experimental Study," *Acta Psychologica* 243 (2024): 104159. https://doi.org/10.1016/j.actpsy.2024.104159

[9] Ting-Yu Fan and Hsiu-Ling Chen, "Developing Cooperative Learning in a Content and Language Integrated Learning Context to Enhance Elementary School Students' Digital Storytelling Performance, English Speaking Proficiency, and Financial Knowledge," *Journal of Computer Assisted Learning* 39, no. 4 (2023): 1354–67. https://doi.org/10.1111/jcal.12804

[10] Muhamad Bisri Ihwan, Sumari Mawardi, and Ulin Ni'mah, "Pengaruh Penguasaan Ilmu Nahwu Dan Sharaf Terhadap Kemampuan Membaca Kitab Fathul Qarib," *TADRIS AL-ARABIYAT: Jurnal Kajian Ilmu Pendidikan Bahasa Arab* 2, no. 1 (2022): 61–77. https://doi.org/10.30739/arabiyat.v2i1.1422

student characteristics, and the ease of controlling research variables.[11] From those two classes, one class was designated as the experimental group and one class as the control group, with 25 students in each, resulting in a total research sample of 50 students.

The experimental group received treatment in the form of Arabic language learning using the CLIL method, which integrates the content of daily activities with the use of Arabic as an active communication tool through discussions, oral presentations, and context-based tasks. Meanwhile, the control group used conventional learning focused on explaining the material and limited exercises without communicative content integration. The data collection instruments in this study include oral tests (pretest and posttest of kalam skills) to measure the improvement of speaking ability, observation sheets to determine the implementation of CLIL learning and students' speaking activities, and a kalam skills assessment rubric covering aspects of fluency, vocabulary accuracy, sentence structure, and speaking courage. Observation is used to ensure that the treatment is applied consistently and according to procedure, as well as to strengthen the validity of quantitative data.[12]

In order to clarify the data analysis stages used in this study, the data processing steps are systematically arranged through several statistical tests performed sequentially. The data was obtained through several stages. The analysis stages include a normality test to determine if the data is normally distributed, a homogeneity test to ensure equal variances between groups, and an independent t-test to examine the difference in posttest mean scores between the experimental and control groups.[13] Additionally, the improvement in speaking skills was analyzed using the N-Gain test, which serves to determine the effectiveness of CLIL implementation in enhancing students' speaking abilities.[14]

The N-Gain calculation results were then interpreted using the following effectiveness categories:

Table 1. N-Gain Interpretation Criteria

N-Gain Range	Category
≥ 0,70	High
0,30 – 0,69	Medium
< 0,30	Low

[11] Konstantin Büchel et al., "The Relative Effectiveness of Teachers and Learning Software: Evidence from a Field Experiment in El Salvador," *Journal of Labor Economics* 40, no. 3 (2022): 737–77. https://doi.org/10.1086/717727

[12] Abdul Basith and Nur Imamatun Nisa, "Pengaruh Interaksi Edukatif Guru serta Disiplin Belajar Siswa terhadap Hasil Belajar Bahasa Arab," *BASA Journal of Language & Literature* 4, no. 2 (2024): 66–72. https://doi.org/10.33474/basa.v4i2.22678

[13] Meli Fiandini et al., "How to Calculate Statistics for Significant Difference Test Using SPSS: Understanding Students Comprehension on the Concept of Steam Engines as Power Plant," *Indonesian Journal of Science and Technology* 9, no. 1 (2024): 45–108. https://doi.org/10.17509/ijost.v9i1.64035

[14] Damar Isti Pratiwi and Budi Waluyo, "Autonomous Learning and the Use of Digital Technologies in Online English Classrooms in Higher Education.," *Contemporary Educational Technology* 15, no. 2 (2023). https://doi.org/10.30935/cedtech/13094

The purpose of using these statistical tests is to ensure that the differences in learning outcomes that emerge are truly caused by the implementation of the CLIL method, so that the research results have empirical strength and can be scientifically accounted for.[15]

RESULT AND DISCUSSION

Descriptive Analysis Test

Descriptive statistical analysis is an important initial step in quantitative research because it serves to provide a general overview of the characteristics of the research data before further statistical testing is conducted. Through descriptive analysis, researchers can determine data trends, value distribution, and the initial and final differences in students' learning outcomes. In the context of this study, descriptive statistical analysis is used to describe students' speaking skills before and after the implementation of the Content and Language Integrated Learning (CLIL) approach, as well as to compare the learning outcomes of the experimental and control groups.

In general, descriptive statistical analysis includes several key measures, namely the minimum value, maximum value, mean, and standard deviation. The minimum and maximum values serve to indicate the range of students' abilities, the mean is used to describe the average level of learning outcomes achievement, while the standard deviation shows the level of variation or diversity of data among students. Thus, descriptive analysis serves as the initial basis for interpreting the effectiveness of the treatment given in the study. As for the difference in the average learning outcomes between the experimental and control classes, it can be seen in Table below.

Table 2. Descriptive Statistics of Pretest and Posttest Results

	N	Minimum	Maximum	Mean	Std. Deviation
Class	50	1	2	1.50	.505
Pretest	50	40	80	60.00	10.690
Posttest	50	55	100	80,36	11.797
Valid N (listwise)	50				

Based on the table, the number of respondents in this study is 50 students, divided equally into the experimental and control groups, with 25 students per class. The average value of the class variable is 1.50 with a standard deviation of 0.505, indicating a relatively balanced distribution of subjects. The pretest results showed that students' initial speaking skills were at an intermediate level with an average score of 60.00. The fairly wide range of scores, from 40 to 80, and a

[15] Enola K Proctor et al., "Ten Years of Implementation Outcomes Research: A Scoping Review," *Implementation Science* 18, no. 1 (2023): 31. https://doi.org/10.1186/s13012-023-01286-z

standard deviation of 10.690, indicate that there was variation in students' speaking abilities before the treatment was administered.

Meanwhile, the posttest results showed an improvement in speaking skills with an average score of 80.36. The minimum and maximum scores of 55 and 100 respectively indicate a shift in learning achievement toward a higher level. A standard deviation of 11.797 indicates that improvement occurred in most students, although variations in ability are still evident. The comparison of pretest and posttest scores shows an average increase of 20.36 points. Descriptively, these findings indicate that implementing the CLIL approach contributes positively to improving students' speaking skills.

Thus, both conceptually and empirically, the results of the descriptive analysis in this study support the theory and previous findings that the implementation of CLIL is effective in improving students' speaking skills. This finding serves as a strong foundation for continuing the analysis at the inferential stage to statistically determine the significance of CLIL's influence on students' learning outcomes.

The results of the descriptive statistical analysis show a clear improvement in students' speaking skills after the implementation of the Content and Language Integrated Learning (CLIL) approach. The increase in the average score from pretest to posttest indicates that the integration of content and language was able to create a more meaningful and contextual learning environment. This finding aligns with the view that CLIL serves not only as a language teaching approach but also as a means of developing communicative competence thru the use of language as a tool for understanding content.[16]

The descriptive improvement in learning outcomes also indicates that the use of CLIL encourages students to more actively use the target language in authentic situations.[17] In learning speaking skills, contextual language exposure plays an important role in improving students' fluency, confidence, and speaking accuracy. CLIL provides broader opportunities for students to produce spoken language meaningfully,[18] thus positively impacting speaking skills. This is relevant to the findings of this study, where the increase in posttest scores reflects the development of students' speaking abilities after participating in CLIL-based learning.

---

[16] A D Owens and R L Hite, "Enhancing Student Communication Competencies in STEM Using Virtual Global Collaboration Project Based Learning," *Research in Science & Technological Education* 40, no. 1 (2022): 76–102. https://doi.org/10.1080/02635143.2020.1778663

[17] Karina Rose Mahan, "The Comprehending Teacher: Scaffolding in Content and Language Integrated Learning (CLIL)," *The Language Learning Journal* 50, no. 1 (2022): 74–88. https://doi.org/10.1080/09571736.2019.1705879

[18] Elizabeth Burke Hadley, Erica M Barnes, and HyeJin Hwang, "Purposes, Places, and Participants: A Systematic Review of Teacher Language Practices and Child Oral Language Outcomes in Early Childhood Classrooms," *Early Education and Development* 34, no. 4 (2023): 862–84. https://doi.org/10.1080/10409289.2022.2074203

Additionally, the variation in scores still evident in the posttest results indicates that although CLIL is generally effective, students' success rates are still influenced by differences in initial abilities and readiness to learn. However, the descriptively significant average increase confirms that this approach is capable of facilitating more inclusive and communicative language learning.[19] These results reinforce previous research findings stating that CLIL positively contributes to the mastery of productive language skills, particularly speaking, compared to conventional language learning.[20]

Shapiro-Wilk Normality

Test After descriptive statistical analysis was conducted to obtain an overview of the characteristics of students' speaking skills learning outcome data, the next step in this study was to perform a prerequisite analysis test, namely the normality test. The normality test aims to determine whether the pretest and posttest data in the experimental and control groups are normally distributed, so that the appropriate type of inferential statistical test can be determined for use in the next analysis stage.

In this study, the normality test was conducted using the Shapiro–Wilk test. The choice of the Shapiro–Wilk test was based on the consideration of the relatively small sample size, which was less than 50 subjects in each group. The Shapiro–Wilk test is recommended as a normality test that has better sensitivity and statistical power compared to other normality tests for small to medium sample sizes. Data is considered normally distributed if the significance value (Sig.) is greater than 0.05.

Normality testing is important in this study because parametric statistical tests, such as the t-test, require that the data being analyzed be normally distributed. Therefore, the results of the normality test serve as the basis for making decisions about the suitability of using further statistical tests. The results of the normality test for pretest and posttest scores in the experimental and control groups are presented in Table below.

Table 3. Shapiro-Wilk Normality Test Results

	Class	Statistic	df	Sig.	Information
Pretest	Experiment	.966	25	.539	Normal
	Control	.966	25	.539	Normal
Posttest	Experiment	.945	25	.197	Normal
	Control	.972	25	.701	Normal

[19] Hanaa ZainEldin et al., "Silent No More: A Comprehensive Review of Artificial Intelligence, Deep Learning, and Machine Learning in Facilitating Deaf and Mute Communication," *Artificial Intelligence Review* 57, no. 7 (2024): 188. https://doi.org/10.1007/s10462-024-10816-0

[20] Adriana Soto-Corominas, Helena Roquet, and Marta Segura, "The Effects of CLIL and Sources of Individual Differences on Receptive and Productive EFL Skills at the Onset of Primary School," *Applied Linguistics* 45, no. 2 (2024): 364–82. https://doi.org/10.1093/applin/amad031

Based on this, the results of the Shapiro-Wilk test show that the significance value for pretest scores in the experimental class and the control class are 0.539 each. This value is greater than 0.05, so it can be concluded that the pretest data in both groups are normally distributed. This indicates that the initial speaking skills of students in the experimental and control classes have a normal and relatively comparable data distribution.

Next, the results of the normality test on the posttest data showed significance values of 0.197 for the experimental class and 0.701 for the control class. Both of these values are also greater than 0.05, so the posttest data in both groups were declared to be normally distributed. The difference in significance values between the experimental and control classes indicates a variation in data distribution, but statistically, both still meet the assumption of normality.

Thus, it can be concluded overall that the pretest and posttest data in both the experimental and control groups are normally distributed. These results indicate that the research data meets one of the main prerequisites for parametric statistical analysis, allowing hypothesis testing using inferential statistical tests to proceed to the next stage.

Normality testing showed that both pretest and posttest data in the experimental and control groups were normally distributed. This is indicated by the Shapiro-Wilk test significance values for all groups being above 0.05, indicating that the data distribution did not significantly deviate from a normal distribution. This finding indicates that the initial and final kalam skills of students in both groups have a comparable distribution pattern.

The use of the Shapiro-Wilk test was considered appropriate because the sample size in each group was less than 50 subjects. This test has a higher sensitivity and statistical power for small to medium-sized samples compared to other normality tests, making the test results a reliable basis for further analysis[21].

Meeting the normality assumption is an important prerequisite for applying parametric statistical tests, such as the independent t-test, which is commonly used in experimental language education research. Fulfilling this assumption allows for a more accurate and valid interpretation of the hypothesis testing results.[22] Additionally, the similarity in data distribution at the pretest stage indicates that the experimental and control groups were in relatively equal initial conditions. Thus, the difference in learning outcomes on the posttest can be more convincingly attributed to the implementation of the CLIL approach, which,

[21] Andrea J Bingham, "From Data Management to Actionable Findings: A Five-Phase Process of Qualitative Data Analysis," *International Journal of Qualitative Methods* 22 (2023): 16094069231183620. https://doi.org/10.1177/16094069231183620

[22] Grigorios Kotronoulas et al., "An Overview of the Fundamentals of Data Management, Analysis, and Interpretation in Quantitative Research," in *Seminars in Oncology Nursing*, vol. 39 (Elsevier, 2023), 151398. https://doi.org/10.1016/j.soncn.2023.151398

according to recent research, contributes to a uniform improvement in speaking ability through communicative and meaningful learning.[23]

Homogeneity Test

After the research data is confirmed to be normally distributed through the normality test, the next prerequisite test step is to conduct a variance homogeneity test. The homogeneity test aims to determine whether the data variances between the experimental and control groups are the same or equivalent. Equal variance is one of the important assumptions in the use of parametric statistical tests, particularly the independent t-test, so that the results of hypothesis testing can be interpreted validly and objectively.

In this study, the homogeneity test was conducted using Levene's Test. This test was chosen because it can effectively test the equality of variances between two or more groups and is relatively insensitive to deviations from data normality. Data is considered homogeneous when the significance value (Sig.) is greater than 0.05. Conversely, if the significance value is less than or equal to 0.05, then the variance between groups is considered non-homogeneous. The results of the homogeneity test for pretest and posttest scores in the experimental and control groups are presented in Table below.

Table 4. Results of Variance Homogeneity Test

Class	Sig	Information
Pretest (Experiment+Control)	1.000	Homogeneous
Posttest (Experiment+Control)	.831	Homogeneous

Based on the table, the Levene's test results for the pretest data show a significance value of 1.000 for all testing bases. This value is greater than 0.05, so it can be concluded that the variance of pretest scores between the experimental and control groups is homogeneous. This finding indicates that both groups had relatively similar levels of initial speaking skills diversity before the treatment was administered.

The homogeneity test on the post-test data yielded a significance value of 0.831. This value exceeds the critical threshold of 0.05, indicating that the data variance is homogeneous. This condition establishes the suitability of the data for analysis using parametric statistical tests. Thus, it can be concluded that the pretest and posttest data meet the assumption of variance homogeneity. Fulfilling these prerequisites confirms that the data is suitable for analysis using parametric statistical tests, allowing for valid and scientifically justifiable testing of the impact of CLIL implementation on improving students' speaking skills.

[23] Masoud Rahimi and Jalil Fathi, "Employing E-Tandem Language Learning Method to Enhance Speaking Skills and Willingness to Communicate: The Case of EFL Learners," *Computer Assisted Language Learning* 37, no. 4 (2024): 924–60. https://doi.org/10.1080/09588221.2022.2064512

Meeting the assumption of variance homogeneity is an important prerequisite for using parametric statistical tests, particularly the independent t-test, as it allows for a more accurate and objective comparison of learning outcomes.[24] Homogeneity at the pretest stage also indicates that the experimental and control groups were in comparable initial conditions, so differences in posttest learning outcomes can be attributed to the implementation of the CLIL approach, not to differences in students' initial abilities.[25] Additionally, the consistent variance in the posttest data indicates that implementing CLIL did not lead to extreme disparities in learning achievement, but rather promoted a relatively even improvement in speaking skills. This finding aligns with recent research results stating that CLIL supports the development of speaking skills through communicative and contextual learning.[26]

Independent Samples t-test

After the pretest and posttest data were found to be normally distributed and have homogeneous variances, the next analysis step is to perform an Independent Samples t-test. This test aims to determine whether there is a significant difference in students' speaking skills learning outcomes between the experimental and control groups, both before and after the implementation of the Content and Language Integrated Learning (CLIL) approach.

The independent t-test was used because this study involved two independent groups (the experimental class and the control class) and aimed to compare the average learning outcomes of these two groups. The decision-making criteria in this test are as follows: if the significance value (Sig. (2-tailed)) is greater than 0.05, there is no significant difference. Conversely, if the significance value is less than 0.05, there is a significant difference between the two groups. The results of the Independent Samples t-Test are presented in Table below.

Table 5. Independent Samples t-Test Results

Class	N	Mean	Sig. (2-tailed)	Information
Pretest (Experiment+Control)	25	.000	1.000	There is no difference.
Posttest (Experiment+Control)	25	12.640	.000	There is a difference

[24] Christina Wekerle, Martin Daumiller, and Ingo Kollar, "Using Digital Technology to Promote Higher Education Learning: The Importance of Different Learning Activities and Their Relations to Learning Outcomes," *Journal of Research on Technology in Education* 54, no. 1 (2022): 1–17. https://doi.org/10.1080/15391523.2020.1799455

[25] Alexandra Vraciu and Anna Marsol, "Content-Specific Vocabulary in CLIL: Exploring L2 Learning Outcomes in a Primary Education Programme in Catalonia," *Language Teaching Research*, 2023, 13621688231170072. https://doi.org/10.1177/13621688231170073

[26] Christiane Dalton-Puffer, Julia Hüttner, and Ana Llinares, "CLIL in the 21st Century: Retrospective and Prospective Challenges and Opportunities," *Journal of Immersion and Content-Based Language Education* 10, no. 2 (2022): 182–206. https://doi.org/10.1075/jicb.21021.dal

The results of the Independent Sample t-Test at the pretest stage showed a significance value of 1.000 with a mean difference of 0.000. Given that this value is above 0.05, there is no significant difference between the experimental and control groups, confirming that the initial abilities of students in both groups were at an equivalent level before the treatment was given.

On the contrary, at the post-test stage, a significance value of 0.000 (< 0.05) was obtained with a mean difference of 12.640. This significant difference, where the experimental class showed higher scores than the control class, proves that the application of the CLIL approach positively contributes to the improvement of students' speaking skills. Thus, the observed improvement in learning outcomes can be directly attributed to the applied learning intervention.

The results of the independent sample t-test show that there is no significant difference in pretest scores between the experimental and control groups because the significance value is > 0.05. This indicates that the initial speaking abilities of students in both groups were at an equivalent level before the treatment was given. However, the posttest results showed a significant difference with a p-value < 0.001, indicating that the implementation of Content and Language Integrated Learning (CLIL) has a proven effect on improving students' speaking skills. Additionally, the average score of the experimental group was higher than that of the control group.

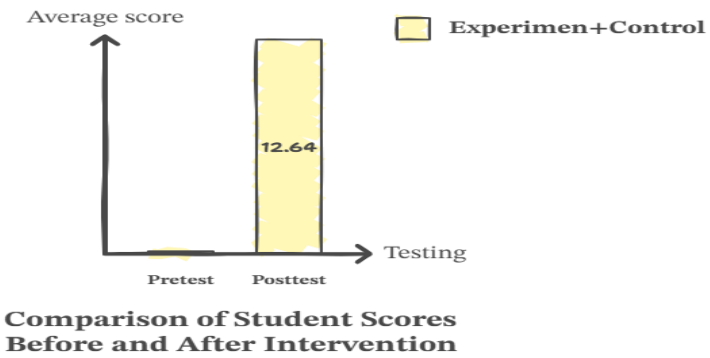

Figure 1. Comparison of Students' Speaking Skills Scores at the Pretest and Posttest Stages

The effectiveness of CLIL in this study is influenced by the application of the 4C principles, namely Content, Communication, Cognition, and Culture. The content aspect is evident from the use of Arabic in understanding learning materials. The communication aspect is apparent in speaking and discussion activities using Arabic. The cognition aspect is visible when students understand and respond to the material using the target language, while the culture aspect emerges thru the use of Arabic in authentic communication contexts.

The results of this study reinforce Do Coyle's 4C theory, which emphasizes the integration of content, communication, thinking processes, and culture in learning. In this study, Arabic is not only learned as a subject but is directly used as a means of communication.[27] Those conditions make students more active, confident, and brave to speak without fear of making mistakes, so their speaking skills develop thru real communication practice.

The findings of this study are also in line with the research by Xin et al. (2025), which states that CLIL is capable of creating more holistic and contextual learning compared to conventional methods.[28] The integration of language and content enhances student engagement and aids in the development of productive language skills. The difference is that this research is applied to Arabic speaking learning at the madrasah level, but it still shows that the 4C principles in CLIL effectively enhance students' communicative competence.

Overall, the research results prove that the implementation of CLIL based on the 4C theory significantly affects the improvement of students' speaking skills. The absence of differences at the pretest stage and the emergence of significant differences at the posttest stage indicate that the improvement in speaking ability is a direct impact of the CLIL implementation. Therefore, CLIL based on the 4C theory is worthy of being applied as a strategy for teaching Arabic in formal education.

N-Gain Test

Next is the N-Gain test. The N-Gain test is used to determine the level of improvement in students' speaking skills after the learning process has taken place, both in the experimental and control classes. This test serves to illustrate the effectiveness of learning by comparing standardized pretest and posttest scores, allowing for a more proportional interpretation of learning outcome improvements. The N-Gain value ranges from 0–1 or can be converted into a percentage, which is then classified into low, medium, or high categories. The results of the N-Gain calculation for the experimental and control classes are presented in Table below.

Table 6. Descriptive N-Gain of Speaking Skills

Descriptives			
N-Gain percentage	**Class**	**Mean**	**Persentase**
	Experiment	0.7105	71.05%
	Control	0.3703	37.03%

[27] Eiman Mustafawi et al., "Perceptions and Attitudes of Qatar University Students Regarding the Utility of Arabic and English in Communication and Education in Qatar," *Language Policy* 21, no. 1 (2022): 75–119, https://doi.org/https://doi.org/10.1007/s10993-021-09590-4.

[28] Xin, Fung, and Tavares, "The 4Cs in Action: Evaluating Language-Driven CLIL Materials for Holistic Learning Experiences." https://doi.org/10.1080/09571736.2025.2494017

Based on the table, the experimental class achieved an average N-Gain score of 0.7105, which is equivalent to 71.05%. This score falls into the high category, indicating that the improvement in students' speaking skills in the experimental class occurred optimally. This indicates that implementing the Content and Language Integrated Learning (CLIL) approach can significantly and sustainably improve students' speaking abilities. Meanwhile, the control class achieved an average N-Gain score of 0.3703 or 37.03%, which falls into the moderate category. This increase indicates that the learning process in the control class still has an impact on improving students' speaking skills, but the level of improvement is relatively lower compared to the experimental class.

The comparison of N-Gain values between the two classes shows a significant difference in improvement, which is 0.3402 or 34.02%. This difference confirms that the experimental class experienced a higher improvement in speaking skills compared to the control class. Thus, descriptively, it can be concluded that the implementation of the CLIL approach is more effective in improving students' speaking ability compared to learning that does not use this approach. The results of this N-Gain test reinforce the findings of the previous statistical tests and provide a more comprehensive picture of the effectiveness level of CLIL learning in improving students' speaking skills.

The N-Gain test results show a difference in the improvement of speaking skills between the experimental group and the control group. If the t-test proves the presence of an effect, the N-Gain analysis is used to measure the quality of that effect. The experimental group obtained an N-Gain score of 0.7105 or 71.05%, which falls into the high category, while the control group obtained a score of 0.3703 or 37.03%, which falls into the moderate category. These results indicate that the Content and Language Integrated Learning (CLIL) approach is more effective than conventional learning in improving students' speaking skills.

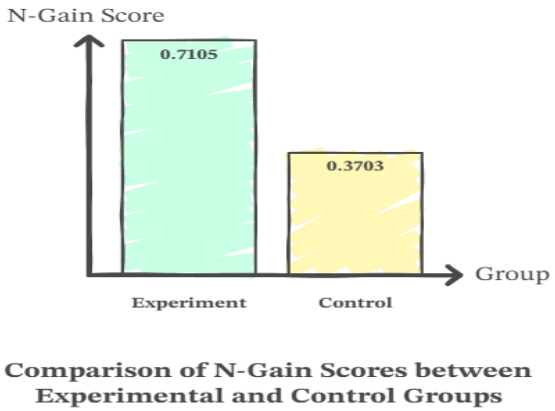

Figure 2. Hypothesis Test Results Using the Normal Distribution Curve *t*

The high N-Gain scores in the experimental group indicate that the integration of language and content in CLIL can create more effective, communicative, and meaningful learning.[29] In this study, students directly use Arabic to understand the *A'māl al-Yaumiyyah* material, thereby developing their speaking skills thru contextual interactions and real communication practices. The leap to the "high" category also indicates that CLIL successfully improved students' speaking fluency and confidence during the learning process.[30] On the contrary, the control group that only reached the moderate category shows that learning without content and language integration is still less optimal in developing students' speaking skills.[31]

The effectiveness can be explained thru the implementation of the 4C framework, which includes Content, Communication, Cognition, and Culture. The content aspect is evident from the use of *A'māl al-Yaumiyyah* materials as a learning context that is close to students' lives. The communication aspect is seen in speaking practice and active discussions using Arabic. The cognition aspect appears when students understand and respond to the material using the target language, while the culture aspect is visible using Arabic in social contexts and authentic communication. The integration of these four aspects makes students more actively use Arabic naturally during the learning process.

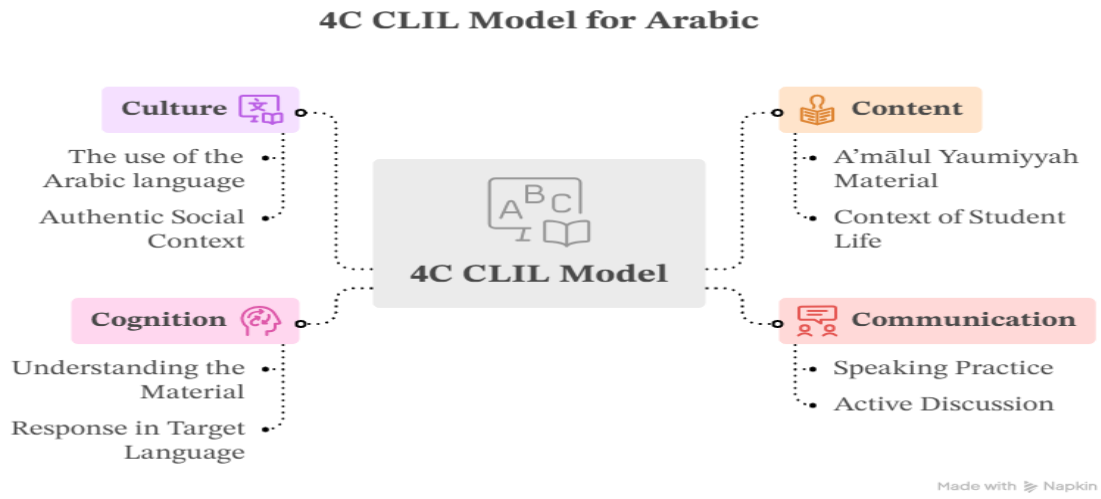

Figure 3. Integration Scheme of the 4C Framework (Content, Communication, Cognition, Culture) in the *A'māl al-Yaumiyyah* Material

[29] Nerea Villabona and Jasone Cenoz, "The Integration of Content and Language in CLIL: A Challenge for Content-Driven and Language-Driven Teachers," *Language, Culture and Curriculum* 35, no. 1 (2022): 36–50. https://doi.org/10.1080/07908318.2021.1910703

[30] Jenny Denman, Erik van Schooten, and Rick de Graaff, "Inclusive CLIL: Pre-Vocational Pupils' Target Language Oral Proficiency, Fluency, and Willingness to Communicate," *AILA Review* 35, no. 2 (2022): 321–50. https://doi.org/10.1075/aila.22020.den

[31] Rahimi and Fathi, "Employing E-Tandem Language Learning Method to Enhance Speaking Skills and Willingness to Communicate: The Case of EFL Learners." https://doi.org/10.1080/09588221.2022.2064512

This study's findings indicate that the application of Content and Language Integrated Learning (CLIL) utilizing Do Coyle's 4C framework -Content, Communication, Cognition, and Culture- substantially enhances students' Arabic speaking abilities at the madrasah level. The lack of statistically significant differences at the pretest stage, succeeded by distinct differences at the posttest stage, indicates a causal relationship between CLIL application and the development of speaking skills, rather than pre-existing disparities across groups. The N-Gain results further substantiate that not only were outcomes enhanced, but the rate and extent of that enhancement were significantly greater in the CLIL group, indicating that the technique expedites communicative development in ways that traditional training does not.

The implications of these results can be comprehended through the four principles themselves. The utilization of *A'māl al-Yaumiyyah* (Daily Activities) as teaching material fostered a personal and experiential connection for pupils to the subject matter, hence reducing the conceptual barrier between language and meaning. Communication activities centered on genuine Arabic usage, as opposed to grammatical exercises, necessitated that students generate language for authentic interactional reasons rather than merely exhibiting formal accuracy. The cognitive dimension emerged as students showed the ability to not only comprehend content delivered in Arabic but also to articulate responses in the target language - a distinctly higher demand than mere passive understanding. The cultural dimension was ultimately engaged through the contextual nature of communication: Arabic functioned not merely as a subject of analysis but as a dynamic medium for meaning construction, which research consistently correlates with enhanced vocabulary retention and greater communicative confidence.

These findings directly support the theoretical assertions of Do Coyle's 4C framework, which posits that language acquisition is enhanced when learners interact with content, participate in communication, employ higher-order thinking, and experience the target language in culturally authentic contexts. In this study, Arabic served concurrently as both the subject of learning and the medium of instruction (a dual role that reflects the circumstances in which first-language ability often evolves). This dual-function orientation seems to be exactly what differentiates CLIL outcomes from traditional methods, in which language and content are procedurally segregated.

The findings align with Xin, Fung, and Tavares (2025), who assessed CLIL materials developed within the 4C framework and discovered that language-driven CLIL facilitated more comprehensive and contextually integrated learning experiences compared to traditional methods, particularly enhancing productive

language skills.[32] This study expands that conclusion to the context of Arabic language acquisition at the madrasah level, a situation that has garnered relatively less attention in CLIL research compared to European bilingual education programs. This extension is significant as it indicates that the efficacy of the 4C framework is not confined to the Western European educational contexts where CLIL has predominantly been examined, but is also applicable to formal Islamic education environments in Indonesia, where Arabic holds a unique sociolinguistic and cultural significance.

The findings correspond with Vraciu and Marsol (2023), who illustrated that content-specific vocabulary instruction within CLIL yields enhanced learning outcomes relative to traditional methods, especially when learners can relate target language content to their personal experiential knowledge.[33] The thematic relevance of *A'māl al-Yaumiyyah* to students' daily experiences likely fulfilled a similar role in the current study, offering a semantic framework that rendered vocabulary acquisition more intuitive and contextually anchored. Both studies converge on a coherent principle: the combination of communication and content in CLIL is not only additive but generative, creating conditions for language development that neither content instruction nor language instruction alone can reproduce. This study's contribution is its demonstration that this principle applies within an Arabic as religious and academic language context, where learners possess cognitive and affective orientations towards the language that significantly differ from those of learners in typical foreign language environments.

The systematic review by Ruiz Hidalgo and Ortega Sánchez (2023) on CLIL in multilingual school environments offers additional background for these findings. Their review highlighted student engagement and effective language use as the most consistent results of CLIL across various educational contexts, a pattern that our study confirms in the domain of Arabic speaking skills.[34] The uniformity observed in these studies, encompassing various languages, educational tiers, and national contexts, bolsters the assertion that the efficacy of CLIL pertains to intrinsic aspects of communicative competence development, rather than being a mere byproduct of specific instructional environments.

This study theoretically enhances the existing evidence that communicative competence in Arabic as a foreign language is most effectively cultivated through

[32] Xin, Fung, and Tavares, "The 4Cs in Action: Evaluating Language-Driven CLIL Materials for Holistic Learning Experiences." https://doi.org/10.1080/09571736.2025.2494017

[33] Vraciu and Marsol, "Content-Specific Vocabulary in CLIL: Exploring L2 Learning Outcomes in a Primary Education Programme in Catalonia." https://doi.org/10.1177/13621688231170073

[34] David Ruiz Hidalgo and Delfín Ortega Sánchez, "CLIL (Content and Language Integrated Learning) Methodological Approach in the Bilingual Classroom: A Systematic Review," *International Journal of Instruction* 16, no. 3 (2023): 915–34, https://doi.org/https://doi.org/10.29333/iji.2023.16349a. https://doi.org/10.29333/iji.2023.16349a

instructional approaches that regard language as a medium rather than an object. The research provides an experimentally substantiated Arabic-language example to Coyle's 4C paradigm, which has been more thoroughly confirmed in European bilingual settings. The framework's four principles operate synergistically in the madrasah classroom: when content is personally relevant, communication is authentic, cognition is engaged, and culture is integrated into language use, the conditions for developing speaking skills align in ways that individualized instruction cannot replicate.

The findings have immediate consequences for the instructional design of Arabic language teachers at the madrasah level. A transition from grammar-translation or rote memorization methods to CLIL-based units -where students utilize Arabic to interact with relevant content from their daily lives or religious and academic studies- is both feasible within current curricular frameworks and substantiated by evidence. Educators do not inherently need specialized training in bilingual education to apply the fundamental 4C principles; rather, a reorientation of the classroom's communicative function is essential, ensuring that Arabic is consistently employed as the medium of instruction and interaction, rather than merely as a subject of study.

These findings advocate for the integration of CLIL-informed Arabic language teaching into national madrasah curriculum guidelines and teacher professional development program at the policy level. If communicative competence in Arabic is an articulated educational objective -which it undoubtedly is for madrasah students involved with Quranic, hadith, and classical textual traditions- then the instructional conditions necessary for cultivating such competence should get institutional endorsement. This encompasses supplying educators with resources centered on genuine communicative content, assessment structures that priorities oral proficiency in conjunction with grammatical precision, and institutional acknowledgement that language instruction and content delivery are interconnected duties.

This study has limitations that influence the interpretation of its conclusions. The study was performed in a singular madrasah environment, utilizing a sample from one educational context in Indonesia, hence constraining the generalizability of the results across other institutional settings, learner demographics, and regional educational cultures. The implementation period was brief; although the pretest to posttest methodology records instantaneous advancements, it fails to determine if CLIL-induced speaking enhancements are maintained over extended durations or applicable in contexts beyond the classroom. The study concentrated solely on speaking skills; it remains uncertain whether CLIL, grounded in the 4C framework, yields similar improvements in reading, writing, or listening within the Arabic environment.

Subsequent research ought to rectify these shortcomings by duplicating the study in various schools and areas, prolonging the implementation time to assess retention effects, and expanding the outcome measures to encompass more dimensions of language skills. Longitudinal approaches that monitor the same learners over an academic year or longer would be particularly beneficial for determining whether initial CLIL-based advancements accumulate with time or stabilize. Investigating teacher variables, such as previous CLIL training, proficiency in Arabic as a medium of instruction, and pedagogical flexibility, would elucidate the conditions under which the 4C framework is most likely to yield favorable outcomes and identify areas where further professional support would be advantageous.

CONCLUSION

This study concludes that the implementation of the Content and Language Integrated Learning (CLIL) approach has proven effective in improving students' Arabic speaking skills compared to conventional teaching methods. The research results show that the experimental group using the CLIL approach achieved a high category increase in speaking skills scores, reaching 71.05% based on the N-Gain test, while the control group using conventional methods only reached a moderate category of 37.03%. This success is based on the CLIL approach's ability to integrate content and language contextually, encouraging students to actively use the target language in meaningful communication.

However, there are several methodological limitations that need to be considered in the interpretation of the results of this study. This study includes a limited and localized sample, involving only 50 students at MTsN 9 Banyuwangi, so generalizing the findings to a broader population needs to be done with caution. The quasi-experimental design of the study means that the researcher cannot randomly assign subjects, which can affect internal validity if there are initial characteristic differences that are not fully detected. Although the researchers used the same teacher for both groups to maintain treatment fidelity, this has the potential to introduce a teacher effect on learning outcomes. Additionally, the use of oral test instruments, even when assessed with an analytic rubric, still carries the risk of subjectivity or evaluator bias in evaluating speaking skills aspects such as fluency and pronunciation. Lastly, the limited duration of the intervention, which was only eight sessions, indicates that these results reflect a short-term impact, thus the long-term effectiveness of this CLIL method still requires further research.

References

Al-Khresheh, Mohammad H. "The Role of Presentation-Based Activities in Enhancing Speaking Proficiency among Saudi EFL Students: A Quasi-Experimental Study." *Acta Psychologica* 243 (2024): 104159. https://doi.org/10.1016/j.actpsy.2024.104159

Arar, Khalid, Rania Sawalhi, and Munube Yilmaz. "The Research on Islamic-Based Educational Leadership since 1990: An International Review of Empirical Evidence and a Future Research Agenda." *Religions* 13, no. 1 (2022): 42. https://doi.org/10.3390/rel13010042

Basith, Abdul, and Nur Imamatun Nisa. "pengaruh interaksi edukatif guru serta disiplin belajar siswa terhadap hasil belajar bahasa arab." *BASA Journal of Language & Literature* 4, no. 2 (2024): 66–72. https://doi.org/10.33474/basa.v4i2.22678

Bingham, Andrea J. "From Data Management to Actionable Findings: A Five-Phase Process of Qualitative Data Analysis." *International Journal of Qualitative Methods* 22 (2023): 16094069231183620. https://doi.org/10.1177/16094069231183620

Büchel, Konstantin, Martina Jakob, Christoph Kühnhanss, Daniel Steffen, and Aymo Brunetti. "The Relative Effectiveness of Teachers and Learning Software: Evidence from a Field Experiment in El Salvador." *Journal of Labor Economics* 40, no. 3 (2022): 737–77. https://doi.org/10.1086/717727

Dalton-Puffer, Christiane, Julia Hüttner, and Ana Llinares. "CLIL in the 21st Century: Retrospective and Prospective Challenges and Opportunities." *Journal of Immersion and Content-Based Language Education* 10, no. 2 (2022): 182–206. https://doi.org/10.1075/jicb.21021.dal

Denman, Jenny, Erik van Schooten, and Rick de Graaff. "Inclusive CLIL: Pre-Vocational Pupils' Target Language Oral Proficiency, Fluency, and Willingness to Communicate." *AILA Review* 35, no. 2 (2022): 321–50. https://doi.org/10.1075/aila.22020.den

Fan, Ting-Yu, and Hsiu-Ling Chen. "Developing Cooperative Learning in a Content and Language Integrated Learning Context to Enhance Elementary School Students' Digital Storytelling Performance, English Speaking Proficiency, and Financial Knowledge." *Journal of Computer Assisted Learning* 39, no. 4 (2023): 1354–67. https://doi.org/10.1111/jcal.12804

Fiandini, Meli, Asep Bayu Dani Nandiyanto, Dwi Fitria Al Husaeni, Dwi Novia Al Husaeni, and M Mushiban. "How to Calculate Statistics for Significant Difference Test Using SPSS: Understanding Students Comprehension on the Concept of Steam Engines as Power Plant." *Indonesian Journal of Science and Technology* 9, no. 1 (2024): 45–108. https://doi.org/10.17509/ijost.v9i1.64035

Hadley, Elizabeth Burke, Erica M Barnes, and HyeJin Hwang. "Purposes, Places, and Participants: A Systematic Review of Teacher Language Practices and Child Oral Language Outcomes in Early Childhood Classrooms." *Early Education and Development* 34, no. 4 (2023): 862–84. https://doi.org/10.1080/10409289.2022.2074203

Hu, Hengzhi, Nur Ehsan Mohd Said, and Harwati Hashim. "Sustaining Content and Language Integrated Learning in China: A Systematic Review." *Sustainability* 15, no. 5 (2023): 3894. https://doi.org/10.3390/su15053894

Ihwan, Muhamad Bisri, Sumari Mawardi, and Ulin Ni'mah. "Pengaruh Penguasaan Ilmu

Nahwu Dan Sharaf Terhadap Kemampuan Membaca Kitab Fathul Qarib." *TADRIS AL-ARABIYAT: Jurnal Kajian Ilmu Pendidikan Bahasa Arab* 2, no. 1 (2022): 61–77. https://doi.org/10.30739/arabiyat.v2i1.1422

Iqbal, Muhammad, Mohammad Ahsanuddin, and Muhammad Alfan. "Enhancing Arabic Proficiency through Content and Language Integrated Learning: Implementation and Challenges in Pesantren." *International Journal of Arabic Language Teaching* 7, no. 02 (2025): 207–22. https://doi.org/10.32332/ijalt.v7i02.10769

Kotronoulas, Grigorios, Susana Miguel, Maura Dowling, Paz Fernández-Ortega, Sara Colomer-Lahiguera, Gülcan Bağçivan, Eva Pape, Amanda Drury, Cherith Semple, and Karin B Dieperink. "An Overview of the Fundamentals of Data Management, Analysis, and Interpretation in Quantitative Research." In *Seminars in Oncology Nursing*, 39:151398. Elsevier, 2023. https://doi.org/10.1016/j.soncn.2023.151398

Lamri, Jeremy, and Todd Lubart. "Reconciling Hard Skills and Soft Skills in a Common Framework: The Generic Skills Component Approach." *Journal of Intelligence* 11, no. 6 (2023): 107. https://doi.org/10.3390/jintelligence11060107

Mahan, Karina Rose. "The Comprehending Teacher: Scaffolding in Content and Language Integrated Learning (CLIL)." *The Language Learning Journal* 50, no. 1 (2022): 74–88. https://doi.org/10.1080/09571736.2019.1705879

Mustafawi, Eiman, Kassim Shaaban, Tariq Khwaileh, and Katsiaryna Ata. "Perceptions and Attitudes of Qatar University Students Regarding the Utility of Arabic and English in Communication and Education in Qatar." *Language Policy* 21, no. 1 (2022): 75–119. https://doi.org/https://doi.org/10.1007/s10993-021-09590-4. https://doi.org/10.1007/s10993-021-09590-4

Owens, A D, and R L Hite. "Enhancing Student Communication Competencies in STEM Using Virtual Global Collaboration Project Based Learning." *Research in Science & Technological Education* 40, no. 1 (2022): 76–102. https://doi.org/10.1080/02635143.2020.1778663

Pratiwi, Damar Isti, and Budi Waluyo. "Autonomous Learning and the Use of Digital Technologies in Online English Classrooms in Higher Education." *Contemporary Educational Technology* 15, no. 2 (2023). https://doi.org/10.30935/cedtech/13094

Proctor, Enola K, Alicia C Bunger, Rebecca Lengnick-Hall, Donald R Gerke, Jared K Martin, Rebecca J Phillips, and Julia C Swanson. "Ten Years of Implementation Outcomes Research: A Scoping Review." *Implementation Science* 18, no. 1 (2023): 31. https://doi.org/10.1186/s13012-023-01286-z

Rahimi, Masoud, and Jalil Fathi. "Employing E-Tandem Language Learning Method to Enhance Speaking Skills and Willingness to Communicate: The Case of EFL Learners." *Computer Assisted Language Learning* 37, no. 4 (2024): 924–60. https://doi.org/10.1080/09588221.2022.2064512

Rijt, Jimmy Van, Arina Banga, and Martijn Goudbeek. "Getting a Load of Linguistic Reasoning: How L1 Student Teachers Process Rules of Thumb and Linguistic Manipulations in Discussions about Grammar." *Applied Linguistics* 45, no. 1 (2024): 163–88. https://doi.org/10.1093/applin/amad011

Ruiz Hidalgo, David, and Delfín Ortega Sánchez. "CLIL (Content and Language Integrated Learning) Methodological Approach in the Bilingual Classroom: A Systematic Review." *International Journal of Instruction* 16, no. 3 (2023): 915–34. https://doi.org/10.29333/iji.2023.16349a.

Salsabila, Salma, Maman Abdurahman, Shofa Musthofa Khalid, and Muhamad Al Bani. "Content and Language Integrated Learning Approach in Qawāʿid Learning through Yellow Books in Islamic Boarding Schools." *International Journal of Arabic Language Teaching* 7, no. 01 (2025): 193–206. https://doi.org/10.32332/ijalt.v7i01.10991

Soto-Corominas, Adriana, Helena Roquet, and Marta Segura. "The Effects of CLIL and Sources of Individual Differences on Receptive and Productive EFL Skills at the Onset of Primary School." *Applied Linguistics* 45, no. 2 (2024): 364–82. https://doi.org/10.1093/applin/amad031

Villabona, Nerea, and Jasone Cenoz. "The Integration of Content and Language in CLIL: A Challenge for Content-Driven and Language-Driven Teachers." *Language, Culture and Curriculum* 35, no. 1 (2022): 36–50. https://doi.org/10.1080/07908318.2021.1910703

Vraciu, Alexandra, and Anna Marsol. "Content-Specific Vocabulary in CLIL: Exploring L2 Learning Outcomes in a Primary Education Programme in Catalonia." *Language Teaching Research*, 2023, 13621688231170072. https://doi.org/10.1177/13621688231170073

Wekerle, Christina, Martin Daumiller, and Ingo Kollar. "Using Digital Technology to Promote Higher Education Learning: The Importance of Different Learning Activities and Their Relations to Learning Outcomes." *Journal of Research on Technology in Education* 54, no. 1 (2022): 1–17. https://doi.org/10.1080/15391523.2020.1799455

Xin, Jieting Jerry, Daniel Fung, and Nicole Judith Tavares. "The 4Cs in Action: Evaluating Language-Driven CLIL Materials for Holistic Learning Experiences." *The Language Learning Journal*, 2025, 1–20. https://doi.org/10.1080/09571736.2025.2494017

ZainEldin, Hanaa, Samah A Gamel, Fatma M Talaat, Mansourah Aljohani, Nadiah A Baghdadi, Amer Malki, Mahmoud Badawy, and Mostafa A Elhosseini. "Silent No More: A Comprehensive Review of Artificial Intelligence, Deep Learning, and Machine Learning in Facilitating Deaf and Mute Communication." *Artificial Intelligence Review* 57, no. 7 (2024): 188. https://doi.org/10.1007/s10462-024-10816-0