BILINGUAL PUBLISHING GROUP

Forum for Linguistic Studies

https://journals.bilpubgroup.com/index.php/fls

ARTICLE

The Effects of CLIL on Enhancing EFL Writing Accuracy and Language Learning Motivation in an Indonesian Vocational High School Setting

Faisal Idris [1] , ***Nur Ehsan Mohd Said*** [1,2*] , ***Nur Ainil Sulaiman*** [1]

[1] *Research Centre for Innovation in Teaching and Learning, Faculty of Education, Universiti Kebangsaan Malaysia (UKM), Bangi 43000, Malaysia*

[2] *Centre for Shaping Advanced and Professional Education, Universiti Kebangsaan Malaysia (UKM), Bangi 43000, Malaysia*

ABSTRACT

Writing accuracy remains a persistent challenge for EFL students in Indonesian vocational high schools, where grammar instruction is often decontextualized. Additionally, low language learning motivation further impedes proficiency, as conventional non-CLIL methods fail to provide meaningful engagement. Content and Language Integrated Learning (CLIL) presents a pedagogically sound alternative, merging subject content with language acquisition to create an interactive and cognitively enriching learning environment. This study explores the efficacy of CLIL in enhancing writing accuracy and language learning motivation among third-grade vocational high school students in Aceh, Indonesia. Employing a mixed-methods, quasi-experimental design, the 12-week intervention integrated CLIL-based lesson plans and e-modules adapted from the national curriculum. The instructional approach emphasized linguistic structures such as cause-and-effect clauses, active and passive voice, modal auxiliaries, and selected tenses (present, past perfect, and future tenses), with structured activities strategically designed to foster motivation. A total of 44 students participated. Writing accuracy was assessed using the Barkovska EFL CLIL assessment grid, while language learning motivation levels were measured through a structured questionnaire. Statistical analyses—including paired t-tests, 95% confidence intervals, and effect size

*CORRESPONDING AUTHOR:

Nur Ehsan Mohd Said, Research Centre for Innovation in Teaching and Learning, Faculty of Education, Universiti Kebangsaan Malaysia (UKM), Bangi 43000, Malaysia; Centre for Shaping Advanced and Professional Education, Universiti Kebangsaan Malaysia (UKM), Bangi 43000, Malaysia; Email: nurehsan@ukm.edu.my

ARTICLE INFO

Received: 29 March 2025 | Revised: 30 April 2025 | Accepted: 17 May 2025 | Published Online: 2 June 2025
DOI: https://doi.org/10.30564/fls.v7i6.9281

CITATION

Idris, F., Mohd Said, N.E., Sulaiman, N.A., 2025. The Effects of CLIL on Enhancing EFL Writing Accuracy and Language Learning Motivation in an Indonesian Vocational High School Setting. Forum for Linguistic Studies. 7(6): 122–148. DOI: https://doi.org/10.30564/fls.v7i6.9281

computations—revealed minimal, non-significant differences between the experimental and control groups. These findings suggest that while CLIL enhances learner engagement, further pedagogical refinement and targeted teacher training are essential to optimize its impact on measurable language proficiency outcomes in vocational EFL settings.
Keywords: CLIL; EFL; Writing Accuracy; Language Learning Motivation; Vocational High School (VHS)

1. Introduction

Content and Language Integrated Learning (CLIL) has emerged as a prominent instructional approach, particularly effective in contexts where language acquisition must complement specific subject knowledge. By integrating content and language learning, CLIL fosters both linguistic and cognitive development, embedding language skills within subject-matter learning [1, 2]. Within the English as a Foreign Language (EFL) educational framework, CLIL is particularly valuable for enhancing students' communicative competence and engagement through context-rich learning experiences [3]. This integration is crucial in Indonesia, where Bahasa Indonesia is the national language and English is primarily taught as a foreign language within formal education [4].

Despite its potential, implementing CLIL within the accuracy of writing and language learning motivation in Indonesian vocational high schools (VHS) faces significant challenges. Vocational education in Indonesia aims to equip students with specific professional skills, yet the level of English proficiency among VHS students remains insufficient, particularly in written communication [5, 6]. Studies indicate that English writing accuracy and language learning motivation in Indonesian EFL classrooms are low, largely due to limited exposure to English in real-world contexts and a curriculum focused heavily on grammatical accuracy rather than applied language use [7–9]. Consequently, Indonesian students often struggle with constructing grammatically correct sentences, understanding complex sentence structures, and accurately using vocabulary in vocational contexts, impeding their professional readiness [10, 11]. Potentially, these factors will indicate a decrease in students' language learning motivation to enhance their existing English language knowledge [12].

Research in various EFL contexts worldwide underscores CLIL's effectiveness in improving writing skills and enhancing language learning motivation by providing students with a purpose for language use beyond rote memorization and grammar drills [13–16]. However, in Indonesia, studies on CLIL are limited and have produced mixed results, with some reporting improvements in language skills and others highlighting challenges due to varying English proficiency levels and resource limitations [17, 18]. For instance, Mukadimah & Sahayu found that while CLIL improved students' vocabulary and comprehension, many struggled with the cognitive load of learning complex content in English, leading to a higher dropout rate from CLIL-based classes [18].

This study specifically targets writing accuracy—focusing on linguistic features such as cause-and-effect clauses, modal auxiliaries, voice (active/passive), and selected tenses (present, past perfect, and future)—as these elements are not only fundamental to academic writing but are also crucial in the workplace communication scenarios expected of vocational graduates. The emphasis on these grammatical features is informed by the demands of vocational tasks such as report writing, instruction giving, and procedural explanations, where precise and effective language use is essential.

Simultaneously, the study addresses language learning motivation (both internal and external) as a core affective factor influencing students' engagement and persistence in English language learning. Given the often utilitarian and goal-oriented nature of vocational education, understanding and fostering motivation is critical for aligning language instruction with students' career-oriented aspirations.

Although CLIL is traditionally associated with content integration, this study applies it as a dual-focused pedagogical approach in which language learning and vocational content mutually support each other. Integrating subject-relevant English instruction—delivered through CLIL-informed lesson plans—enables learners to develop language skills in a meaningful, contextualized environment, which supports both linguistic accuracy and language learning motivation.

The dual nature of the inquiry justifies the mixed-methods design: (1) quantitatively measuring changes in writing accuracy and language learning motivation before and after the CLIL intervention (RO1–RO4), and (2) qualita-

tively exploring the extent of CLIL's influence on these domains (RO5–RO6). This design allows for triangulation and a more nuanced understanding of the intervention's impact. The central hypothesis—suggesting that CLIL enhances both writing accuracy and language learning motivation—is supported by the two null hypotheses (H_{01} and H_{02}), which are used to test the statistical significance of the intervention's effects.

In addition to the overarching research aim, this study outlines six specific research objectives, as detailed below:

RO1: To determine the vocational school students' writing accuracy scores before the intervention.

RO2: To determine the vocational school students' learning motivation scores before the intervention.

RO3: To measure the gain scores of vocational school students' writing accuracy after the intervention.

RO4: To measure the gain scores of vocational school students' language learning motivation after the intervention.

RO5: To explore the extent of CLIL effects on vocational school students' writing accuracy

RO6: To explore the extent of CLIL effects on vocational school students' language learning motivation.

The central hypothesis posits that CLIL instruction in an Indonesian vocational high school enhances third-grade students' writing accuracy and language learning motivation. To determine the statistical significance of the quantitative data, employing null hypotheses remains fundamental [19, 20], as described below:

H_{O_1}. *There is no significant score gain in the learners' writing accuracy after the intervention among the treatment group participants.*

H_{O_2}. *There is no significant score gain in the learners' language learning motivation after the intervention among the treatment group participants.*

2. Literature Review

2.1. CLIL and Writing Accuracy in the EFL Setting

Recent research increasingly supports using Content and Language Integrated Learning (CLIL) to enhance writing accuracy among EFL learners by promoting the simultaneous development of linguistic competence and content knowledge. Studies consistently show that CLIL environments improve lexicogrammatical accuracy, helping students reduce syntactic and lexical errors and produce more structurally complex writing than their non-CLIL peers [21, 22]. These gains are attributed to CLIL's emphasis on cognitively engaging, content-driven tasks that require learners to apply grammar in authentic contexts [23].

While traditional EFL classrooms often reveal gender-based performance gaps—typically favoring female students—emerging evidence suggests that CLIL may help equalize writing outcomes across genders by shifting focus away from rote drills to meaningful, task-based learning [24]. This equalizing effect reinforces the pedagogical value of CLIL in diverse classrooms.

Several studies also demonstrate that embedding grammar instruction within subject content—such as in history or science modules—leads to better internalization and application of grammatical forms [25, 26]. This finding aligns with Swain's Output Hypothesis, which emphasizes the role of meaningful language production in developing accuracy.

CLIL has shown promise across various educational contexts, including vocational high schools in EFL settings. Learners in these settings, such as in Indonesia, have reported improvements in formal writing tasks when exposed to CLIL instruction [27]. This suggests that CLIL's dual focus enhances grammatical control and supports broader academic and career readiness. Therefore, the literature broadly agrees on CLIL's benefits for writing accuracy, with evidence of consistent gains in grammatical precision, learner engagement, and task performance across contexts. However, variation in outcomes points to the need for further research on contextual factors and instructional design. This study contributes to that conversation by evaluating CLIL's effectiveness in a vocational EFL setting, addressing a gap in applied CLIL research in underrepresented educational environments.

2.2. CLIL and Language Learning Motivation in the EFL Setting

Growing evidence suggests that the CLIL approach is pivotal in enhancing language learning motivation, particularly within EFL settings. Research indicates improvements in students' intrinsic engagement, positive emotional orientation, and interest in academic content when language is

taught through authentic, content-driven instruction [28–30]. These outcomes are linked to CLIL's integrated model, which combines language development with meaningful subject learning, thereby promoting relevance, autonomy, and self-efficacy—factors essential for maintaining long-term motivation [1, 31].

Empirical studies across primary, secondary, and tertiary levels generally report positive motivational outcomes, though the extent and durability of these effects vary. For instance, some studies have found statistically significant increases in learner motivation [32, 33], while others have noted only modest improvements, often influenced by factors such as program duration, implementation quality, or learner proficiency levels [34]. The authenticity and cognitive challenge of the CLIL environment are often cited as enhancing motivation, but they can also lead to frustration or demotivation if not adequately scaffolded [35, 36].

CLIL's impact on teacher motivation and classroom dynamics is also crucial. Studies indicate that CLIL fosters a collaborative learning culture, enhancing teacher-student interactions and promoting spontaneous language use and confidence [37–39]. Coyle identifies three dimensions—learner environment, engagement, and identity—as central to understanding motivation in CLIL, a framework supported by recent findings linking classroom practices to affective learner outcomes [15, 40, 41].

In Asian contexts, particularly in Indonesia, CLIL has shown promise in enhancing not only language skills but also values, activeness, and character development [6, 42]. Incorporating genre-based pedagogy or ICT tools into CLIL has been shown to reduce anxiety and boost writing motivation, particularly in vocational high schools. However, challenges such as limited teacher proficiency, rigid curricula, and insufficient resources—especially in rural areas—continue to impede effective implementation. Addressing these structural barriers through targeted teacher training, curricular flexibility, and policy support is essential to fully realizing CLIL's motivational potential in Indonesian EFL classrooms. Thus, while the literature generally agrees that CLIL supports motivation and language achievement, its success depends on contextual and pedagogical factors. This study contributes to the ongoing discussion by examining motivation in a vocational EFL setting—where CLIL remains underexplored—highlighting how instructional design and learner support influence motivational outcomes.

Furthermore, Content and Language Integrated Learning (CLIL) fosters writing accuracy and language learning motivation through the integrated principles of Coyle's 4Cs framework—*Content, Communication, Cognition, and Culture*. Firstly, *Content* provides a meaningful context for writing by embedding language learning within subject-specific material, such as science or history. This connection gives purpose to writing tasks and promotes the use of accurate vocabulary and structures required to communicate discipline-specific ideas effectively. Secondly, *Communication* emphasizes the functional use of language, positioning writing as a tool for real expression and interaction rather than isolated language practice. When students write for genuine audiences and purposes, their motivation to improve and revise increases, naturally supporting greater linguistic accuracy. Thirdly, *Cognition* pushes learners to engage in higher-order thinking, such as analyzing, synthesizing, or evaluating content, which demands more precise and complex language use. This cognitive challenge encourages careful attention to grammatical and syntactic accuracy to convey nuanced ideas. Finally, *Culture* enriches writing by allowing students to explore and articulate diverse perspectives, including their own, about global and intercultural themes. Writing with cultural awareness often requires thoughtful language choices, further supporting accuracy, while also boosting motivation through personally and socially relevant expression. Collectively, the 4Cs framework supports a dynamic learning environment where motivation and writing accuracy reinforce one another in the pursuit of meaningful communication across languages and disciplines.

3. Material and Methods

This study investigates the impact of the CLIL approach on enhancing the writing accuracy and language learning motivation of third-grade vocational high school (VHS) students, within the context of a recent curriculum promoting pedagogical CLIL applications among Indonesian students. The intervention content is based on the first-semester curriculum of third-grade students. Specifically, a teacher engaged in collaborative classroom intervention research with a given CLIL English module over 12 weeks of lesson plans.

3.1. Research Materials

The research employed classroom research and a quasi-experimental design with non-equivalent pre-post-test groups to analyze teaching and learning in context, addressing pedagogical-related issues[43, 44]. To examine students' language learning motivation—both internal and external—and its impact on writing accuracy, a questionnaire was administered. The questionnaire explored both positive and negative perceptions of learning writing through the CLIL approach, adapt and adapt from[45, 46]. It included four-point Likert scale questions: 1 (strongly disagree), 2 (disagree), 3 (agree), and 4 (strongly agree), measuring students' attitudes toward English and the frequency of using English outside the classroom. The participants comprised both Bahasa Indonesia and Acehnese speakers, with varying levels of English language practice, as shown in **Table 1**. Additionally, a semi-structured interview was conducted to evaluate the post-intervention impacts of CLIL. To further analyze students' writing performance, six post-writing test document samples—representing high, middle, and low scores—were examined.

Table 1. Summary of Students' Demographic Information of Experiment and Control Group.

SEX					
MALE			**FEMALE**		
17			27		
L1					
Indonesia	Aceh	Java	Indonesia	Aceh	Java
14	4	0	24	2	2
A predisposition towards liking English (5 to 1)					
Strongly Like	Like	Neutral	Dislike	Strongly Dislike	
6	17	18	0	0	
Frequency of use of English outside the classroom (times per week)					
0	1–3	4–6	7–9	10 and more	
0	4	1	1	0	

3.2. Research Procedures

This mixed-method study was conducted among third-grade students in the Accounting Department at a vocational high school (VHS) to examine their writing accuracy and language learning motivation. The study employed a multifaceted approach to ensure comprehensive data collection, incorporating intervention strategies, pre- and post-writing assessments, questionnaires, semi-structured interviews, and document analysis.

3.3. Participants

3.3.1. Criteria for Participant Selection

The selection of respondents in this study was guided by purposeful and practical considerations to ensure data validity. For the teacher-participant, a convenience sampling method was applied. The criteria included substantial teaching experience at the vocational high school level, current involvement in teaching third-grade students, and a demonstrated willingness to collaborate in implementing the 12-week CLIL intervention. For student participants, the study employed intact classroom groups, selected based on academic comparability and alignment with the quasi-experimental research design. This approach ensured a natural classroom environment without randomization while maintaining equivalency between groups. Additional considerations included linguistic diversity among students and their relevance to the target population. Furthermore, two independent raters were selected based on their expertise in EFL writing instruction and assessment, each with more than a decade of experience. They were chosen from different schools to maintain objectivity and ensure scoring reliability.

3.3.2. Number of Participants

In total, 45 participants were involved in this study. This included one English teacher, who served as the teacher-participant responsible for implementing the CLIL-based instruction during the 12-week intervention. The student participants comprised 44 third-grade vocational high school students. The experimental group (EG) consisted of 23 students (9 males and 14 females) from the highest academi-

cally performing class, although their English proficiency was moderate. The control group (CG) included 21 students (8 males and 13 females) from another class with similar academic characteristics, who continued with non-CLIL instruction throughout the study.

3.4. Research Methods

This study employed a mixed-methods research approach within a classroom research framework, utilizing a quasi-experimental design to examine the impact of the Content and Language Integrated Learning (CLIL) approach on students' writing accuracy and language learning motivation. Two intact third-grade classes from a vocational high school were selected—one as the experimental group, receiving CLIL-based instruction, and the other as the control group, following non-CLIL teaching methods. The quasi-experimental design allowed for natural classroom settings without randomization while still maintaining comparability between groups.

To ensure the validity and reliability of the research instruments, a pilot study was conducted before the main research. This pilot involved 26 third-grade students (aged 17–18) from the Multimedia Department of the same vocational school, who were not included in the main study. The primary objective was to identify and resolve potential issues related to the design, administration, and scoring of both the writing test and the language learning motivation questionnaire. For the writing task, students—many of whom had limited proficiency in English—were asked to produce a 150–200-word essay in response to the prompt: *"Describe the most memorable experience while you are studying English at your school."* The pilot revealed issues concerning prompt clarity and content relevance, as some students misunderstood the task. For example,

> *Hi*
>
> *Whoever read this, I hope you are happy I do not give up easily remember all children have their own fortune, not all of them can run smoothly, tired? Of course everyone feels. It's okay, humane feel free to share your story with other, but choose people you trust, believe in something that ends beautifully? Yes I don't stop moving forward. I always pray for your happiness. hope we meet another time*
>
> *Oh yeah, I wanted to be Psychologist haha*
>
> *But I took a different path. Please pray that I can become a psychologist in own way, I want to help many people there.*
>
> *Umm … thank you*

In response, the prompt was revised to include more explicit language features, and clearer instructions along with sample responses were provided. Additionally, a more supportive testing environment was established through the integration of the CLIL approach, which helped activate students' prior knowledge and provided meaningful content context.

Student writing performance was evaluated using the Barkovska EFL CLIL assessment grid, which aligns with the objectives of CLIL instruction. The language learning motivation questionnaire was also assessed for clarity, item functionality, timing, and reliability. The instrument demonstrated acceptable internal consistency, with a Cronbach's alpha coefficient of 0.710, as in **Table 2** [47, 48], exceeding the commonly accepted threshold of 0.6 [46], thereby confirming its reliability in measuring both intrinsic and extrinsic motivation. Based on the pilot findings, minor revisions were made to enhance the clarity and appropriateness of the questionnaire items for the target student population. For instance, some students had difficulty understanding terms such as academic (akademis) and mother tongue (Bahasa ibu), which were subsequently rephrased or supplemented with simpler explanations.

Table 2. Reliability Statistics.

N Items	Cronbach's Alpha
26	0.710

In the main study, quantitative data were collected through two primary instruments: a writing accuracy test and a structured language learning motivation questionnaire. For the writing test, students were asked to compose a 150–200-

word essay in response to a prompt such as: *"Describe the most memorable experiences while studying English at your school. Please insert the elements of writing accuracy, such as cause-and-effect relationship clauses, modal auxiliaries, active and passive voice, and specific tenses (present tense, past perfect tense, and future tense) where necessary."* The students' written responses were evaluated by two independent raters—selected to ensure objectivity and avoid scoring bias—using an adapted Barkovska EFL CLIL assessment grid (see **Appendix A Table A1**)[49].

In addition, the structured motivation questionnaire consisted of 30 items designed to measure language learning motivation, with 20 items focusing on positive motivation and 10 items addressing negative motivation. The data obtained from both instruments were analyzed using various statistical techniques, including tests of normality (Kolmogorov-Smirnov and Shapiro-Wilk), outlier detection, paired-sample and independent-sample t-tests, Hedges' *g* for effect size calculation, 95% confidence intervals, and descriptive statistics such as mean, standard deviation, skewness, and kurtosis.

A semi-structured interview was conducted to examine the strengths and weaknesses of CLIL implementation after the intervention. Face-to-face interviews were conducted with seven students and a teacher-participant, supplemented by 15 and 13 questionnaires, respectively. The collected data were transcribed and thematically analyzed for qualitative analysis. Through a rigorous coding process, three key themes emerged, such as a comparative analysis of whole-class instruction versus CLIL instruction in students' writing acquisition, the impact of the classroom intervention program, and the potential of CLIL instruction to enhance writing skills in the EFL context of Aceh, Indonesia[46]. To ensure consistency in data representation, a teacher participant was coded as ***TP***, respondents as ***R*** (regardless of gender), and questions as ***Q***.

Furthermore, for document analysis, we reviewed six samples of students' written work (the highest, middle, and lowest scores) of both groups to explore the six elements of writing aspects, such as cause-and-effect relationship clauses, modal auxiliaries, active and passive voice, and specific tenses (present tense, past perfect tense, and future tense) within the students' writing. This approach aligns with the methodologies discussed by Mashuri et al. and Cardno[50, 51], who emphasize the effectiveness of semi-structured interviews in qualitative research. Additionally, the thematic analysis process is supported by Onwuegbuzie et al.[52], highlighting its relevance in educational studies. The document analysis methodology is consistent with the best practices outlined by Hamed and Liu and Zu[53, 54].

3.5. Intervention Procedures

The intervention was implemented in five sequential stages: Pre-research, Need Analysis, Intervention Planning, Intervention Implementation, and Post-intervention, each carefully documented to ensure transparency and procedural integrity.

In the Pre-research stage, the research team secured ethical clearance from the National University of Malaysia (UKM) and received approval from the university's confirmation review committee. Additionally, formal permission was obtained from the school principal to conduct the study within the participating institution.

The Need Analysis phase was conducted through a combination of formal and informal meetings with the teacher-participant. These sessions were used to review the annual teaching schedule and plan for subsequent school visits. A sample lesson plan and accompanying student activity sheets were designed and reviewed with guidance from academic supervisors (**Appendix A Table A2**). The lesson plan checklists were then validated by educational experts to ensure content appropriateness and alignment with pedagogical goals. Necessary revisions were made based on this expert feedback to improve clarity, structure, and instructional coherence.

During the Intervention Planning stage, a comprehensive 12-week instructional module was developed, featuring lesson plans aligned with the national curriculum and activity sheets tailored to the CLIL framework. To ensure methodological fidelity, the module was evaluated by two experts from a reputable university in Indonesia. Following this, the module was reviewed by the teacher-participant to confirm alignment with the official syllabus. Revisions were made as needed to ensure both pedagogical integrity and curricular coherence.

The Intervention Implementation began with the administration of the pre-tests—a writing accuracy inventory and a language learning motivation test—across both the exper-

imental (CLIL) and control (non-CLIL) groups. To ensure consistent delivery of the CLIL module, the researcher conducted the first lesson demonstration in the classroom. The remaining sessions were carried out by the trained teacher-participant, who adhered to the standardized lesson plans provided. The control group received instruction covering similar content topics but without CLIL methodology, thereby isolating the instructional approach as the independent variable.

In the post-intervention phase, all participants completed post-tests assessing writing accuracy and language learning motivation. The phase concluded with a school visit, during which the researcher formally expressed gratitude to the administrators, teachers, and students. Participants also received compensation, marking the ethical closure of the study. This documented process demonstrates that the intervention was not only systematically designed but also implemented with careful attention to fidelity, standardization, and ethical research conduct. The structured execution enhances the credibility of the findings while underscoring the need for more robust fidelity monitoring in future studies.

4. Result and Discussion

4.1. Quantitative Analysis

This study aims to investigate the impact of the CLIL approach on EFL writing accuracy and language learning motivation among Indonesian VHS students and examine how this pedagogical method can be adapted to better serve vocational education needs.

4.1.1. Analysis of Writing Test and Language Learning Motivation Scores Pre-and Post-CLIL Intervention

Figures 1 and **2** display individual scores in writing accuracy and language learning motivation across pre- and post-test phases for the Experimental Group (EG), which received CLIL instruction, and the Control Group (CG), which followed traditional instruction. These visualizations highlight learner-level responses to the intervention.

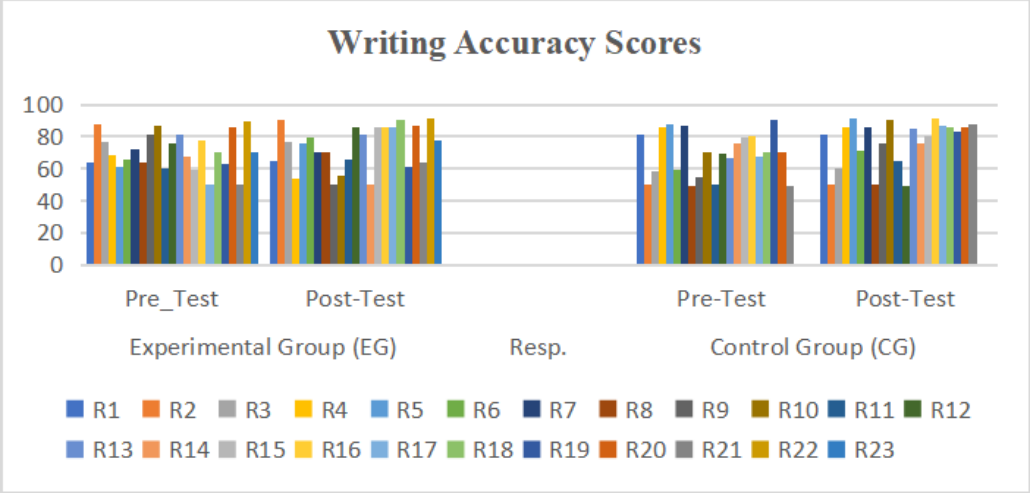

Figure 1. Summary of Scores of Writing Accuracy Test Differences Between CLIL and Non-CLIL Intervention in Both Groups.

Learners in the EG, such as R2, R5, and R10, demonstrated clear improvements in writing accuracy following the intervention, suggesting a beneficial effect of CLIL. However, gains were uneven; learners like R4 and R15 showed limited progress, possibly due to the cognitive demands of CLIL and individual differences in adaptability or prior language competence[34, 55]. In contrast, the CG showed more consistent improvements, particularly among R9, R11, and R17. This may reflect the advantages of traditional instruction's familiarity and structure, which can support more immediate language development through reduced cognitive load and clearer pedagogical framing[56, 57].

The EG began with relatively high motivation levels (e.g., R1, R4, R6), aligning with research suggesting that CLIL fosters initial engagement through content-language integration and novelty[58, 59]. However, post-test data re-

veal a decline in motivation for some (e.g., R4, R10, R14), potentially due to ongoing cognitive strain and a lack of affective support [60, 61]. Meanwhile, the CG maintained stable or slightly improved motivation (e.g., R7, R20), possibly due to the predictability and emotional safety offered by traditional instruction [62, 63].

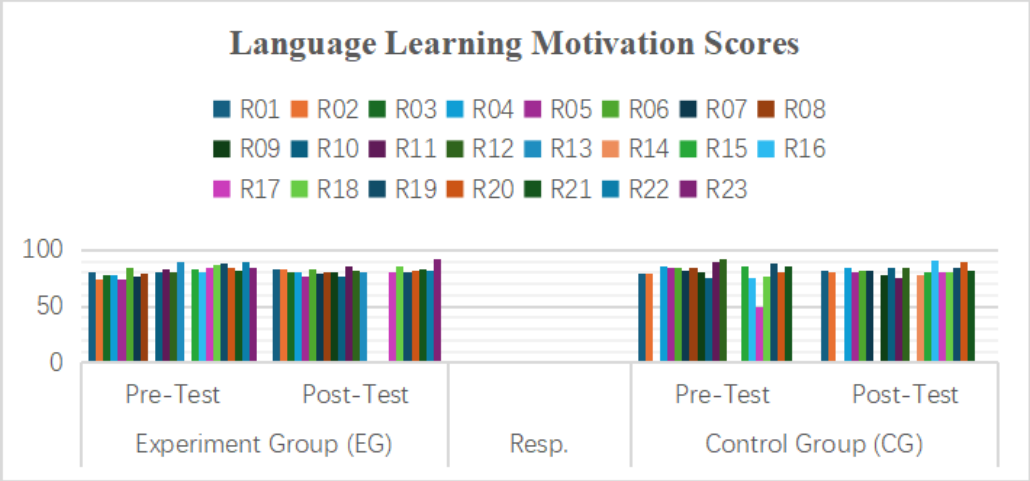

Figure 2. Summary of Scores of Language Learning Motivation Differences Between CLIL and Non-CLIL Intervention in Both Groups.

Overall, CLIL may enhance writing accuracy for certain learners, but its effectiveness appears contingent on learner readiness, scaffolding, and instructional design. Motivational outcomes are more nuanced: while initial interest may be high, sustaining motivation requires addressing cognitive and emotional demands [64]. Non-CLIL instruction, while less innovative, may promote more consistent short-term outcomes due to its lower complexity and greater familiarity.

These findings underscore the importance of maintaining intervention fidelity. Variations in teaching quality, learner support, or instructional context can significantly influence results. Future research should incorporate fidelity monitoring and control for confounding factors to ensure robust and generalizable conclusions [22, 65].

4.1.2. Analysis of Distribution of Normality

Before conducting a paired-sample t-test, the data were assessed for normality and outliers. Normality was tested using the Shapiro-Wilk test and by examining skewness and kurtosis, while outliers were identified using Tukey's Outlier Labelling Rule [66–68]. A nonparametric alternative was used if normality was violated, and outliers were excluded. The null hypothesis, stating that the data follow a normal distribution in writing accuracy and language learning motivation scores, was rejected if $p < 0.05$. Statistical analyses of pre-and post-test scores are presented in **Tables 3** and **4**.

Table 3. Tests of Normality for Overall Scores of Writing Accuracy Tests of Both Groups.

	Group	Kolmogorov-Smirnov [a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Writing Accuracy	Pre-Test Experiment (CLIL)	0.09	23	0.20	0.96	23	0.56
	Post-Test Experiment (CLIL)	0.16	23	0.16	0.92	23	0.07
	Pre-Test Control (non-CLIL)	0.11	21	0.20	0.93	21	0.14
	Post-Test Control (non-CLIL)	0.21	21	0.00	0.82	21	0.00

Note: This is a lower bound of the true significance.
[a] Lilliefors Significance Correction.

Table 3 presents the normality test results for the experimental (CLIL) and control (non-CLIL) groups, using the Kolmogorov-Smirnov (K-S) and Shapiro-Wilk (S-W) tests, both before and after the intervention. In the experimental group, normality was confirmed for both the pre-test (K-S = 0.085, $p = 0.20$; S-W = 0.96, $p = 0.56$) and post-test (K-S

= 0.155, p = 0.161; S-W = 0.921, p = 0.070), indicating no significant deviations. Similarly, the control group exhibited normality in the pre-test (K-S = 0.11, p = 0.200; S-W = 0.930, p = 0.135); however, the post-test results (K-S = 0.21, p = 0.002; S-W = 0.816, p = 0.001) demonstrated a significant deviation from normality. These findings suggest that while most of the dataset adhered to normal distribution assumptions, the control group's post-test scores deviated significantly. Furthermore, concerning language learning motivation, the S-W test yielded p-values of 0.000 for both the pre-test and post-test, supporting the assumption that the data approximates a normal distribution.

Table 4. Tests of Normality for Overall Scores of Language Learning Motivation of Both Groups.

	Group	Kolmogorov-Smirnov [a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Motivation Scores	Pre-Test Experiment (CLIL)	0.40	23	0.00	0.50	23	0.00
	Post-Test Experiment (CLIL)	0.45	23	0.00	0.50	23	0.00
	Pre-Test Control (non-CLIL)	0.38	21	0.00	0.62	21	0.00
	Post-Test Control (non-CLIL)	0.42	21	0.00	0.54	21	0.00

[a] Lilliefors Significance Correction.

The analysis of writing accuracy and language learning motivation across CLIL and non-CLIL groups reveals distinct trends in learner performance (**Table 5**). Writing accuracy improved in both groups, with the CLIL group showing a moderate increase (M = 70.74 to 73.87) and reduced variability, while the non-CLIL group exhibited a larger gain (M = 69.14 to 77.23) but greater score dispersion. Skewness and kurtosis shifts suggest that CLIL instruction fostered more stable linguistic outcomes, whereas non-CLIL instruction led to greater individual variability. In contrast, language learning motivation declined in the CLIL group (M = 74.83 to 71.35) and remained relatively stable in the non-CLIL group (M = 69.43 to 70.52). The extreme pre-test kurtosis in the CLIL group (7.851) suggests initially high motivation levels, which became more dispersed post-intervention, likely due to the cognitive demands of CLIL instruction [69, 70]. Meanwhile, the non-CLIL group exhibited more stable motivation patterns, indicating that non-CLIL instruction may provide a less demanding but more consistent motivational environment. These findings align with research suggesting that while CLIL enhances linguistic competence, its impact on motivation is complex and context-dependent, highlighting the need for scaffolding strategies to sustain engagement while maintaining linguistic gains [71].

Table 5. Descriptives of Skewness and Kurtosis Levels of Writing Accuracy and Language Learning Motivation.

		Writing Accuracy		Language Learning Motivation	
		Statistics	Std. Error	Statistics	Std. Error
Pre-Test EG Scores (CLIL)	Mean	70.74	2.36	74.83	5.00
	Skewness	−0.43	0.48	−2.96	0.48
	Kurtosis	−0.74	0.94	7.85	0.94
Post-Test EG Scores (CLIL)	Mean	73.87	2.80	71.35	5.93
	Skewness	−0.43	0.48	−2.29	0.48
	Kurtosis	−0.09	0.94	3.70	0.94
Pre-Test CG Scores (non-CLIL)	Mean	69.14	2.98	69.43	6.61
	Skewness	−0.10	0.50	−1.90	0.50
	Kurtosis	−1.25	0.97	2.18	0.97
Post-Test CG Scores (non-CLIL)	Mean	77.23	3.05	70.52	6.49
	Skewness	−1.14	0.50	−2.13	0.50
	Kurtosis	−0.04	0.97	2.97	0.97

4.1.3. Homogeneity of Variance

In addition to normality and descriptive analyses, Levene's Test of Homogeneity of Variance was conducted to evaluate whether the assumption of equal variances across groups was satisfied for both writing accuracy and language learning motivation. The results revealed no significant differences in variance among the four groups.

For writing accuracy, all Levene statistics based on various estimation methods (mean, median, trimmed mean, and adjusted degrees of freedom) produced p-values well above the 0.05 significance level (e.g., based on mean: Levene = 0.492, $p = 0.690$), indicating homogeneity of variances. Similarly, for language learning motivation, Levene's test confirmed equal variances across groups (e.g., based on mean: Levene = 0.728, $p = 0.540$).

These findings affirm the suitability of parametric tests for analyzing writing accuracy, as both the normality and homogeneity of variance assumptions were met. In contrast, although variance was equally distributed for motivation scores, the significant departures from normality warrant the use of non-parametric tests for motivation-related analyses. Overall, these results, as presented in **Table 6** and **Table 7**, enhance the validity of subsequent statistical comparisons and suggest that observed differences in learner performance are unlikely to be attributed to unequal variances across groups.

Table 6. Test of Homogeneity of Variance of Writing Accuracy.

		Levene Statistic	**df1**	**df2**	**Sig.**
Scores	Based on Mean	0.49	3	84	0.69
	Based on Median	0.28	3	84	0.84
	Based on the Median and with adjusted df	0.28	3	68.2	0.84
	Based on trimmed mean	0.44	3	84	0.72

Table 7. Test of Homogeneity of Variance of Language Learning Motivation.

		Levene Statistic	**df1**	**df2**	**Sig.**
Scores	Based on Mean	0.728	3	84	0.54
	Based on Median	0.214	3	84	0.89
	Based on the Median and with adjusted df	0.214	3	81.7	0.89
	Based on trimmed mean	0.647	3	84	0.59

Moreover, the results of the normality tests and descriptive statistics provide critical insights into the differential effects of CLIL (Content and Language Integrated Learning) and non-CLIL instruction on learners' writing accuracy and language learning motivation. Normality was assessed using both the Shapiro-Wilk and Kolmogorov-Smirnov tests [66, 72]. For writing accuracy, the CLIL group's pre-test (S-W = 0.96, $p = 0560$) and post-test (S-W = 0.921, $p = 0.070$) scores showed no significant deviations from normality, indicating the data were suitable for parametric testing. Similarly, the control group's pre-test met the normality assumption (S-W = 0.930, $p = 0.135$); however, its post-test scores significantly violated normality (S-W = 0.816, $p = 0.001$). In line with standard statistical protocols [73], non-parametric methods were therefore considered appropriate for analyzing post-test data in the control group where normality was not upheld.

In contrast, language learning motivation scores across both groups violated the normality assumption at all time points, as consistently indicated by highly significant Shapiro-Wilk results ($p = 0.000$). This consistent deviation necessitates the use of non-parametric approaches to ensure the validity and reliability of any statistical inferences drawn from the motivational data [74]. Additionally, Tukey's outlier labeling rule was applied to identify influential data points, and any identified outliers were excluded to preserve the integrity of the analysis [68].

Descriptive statistics revealed meaningful patterns in learner outcomes. In terms of writing accuracy, the CLIL group experienced a moderate improvement in mean scores, rising from 70.74 to 73.87. This was accompanied by a

decrease in kurtosis and stable skewness (−0.426), suggesting more consistent and symmetric performance following CLIL instruction. Conversely, the non-CLIL group showed a larger mean gain (from 69.14 to 77.23), but this was accompanied by increased negative skewness (−1.136) and a near-zero kurtosis (−0.037), indicating a broader spread of scores and more variability in learner outcomes. These differences suggest that while non-CLIL instruction may lead to substantial gains for some learners, the effects are less evenly distributed[1, 16, 75].

Language learning motivation exhibited a different trend. The CLIL group started with high motivation levels (M = 74.83) and extreme kurtosis (7.851), indicating a tightly clustered distribution among highly motivated learners. However, post-test scores declined (M = 71.35) and became more dispersed (kurtosis = 3.70), suggesting a drop in motivation and a widening range of affective responses. The consistently high negative skewness (−2.96 to −2.29) indicates a subset of learners remained highly motivated, but the overall decline points to the cognitive load of CLIL as a potential demotivating factor[76–78]. In contrast, the non-CLIL group maintained relatively stable motivation scores (M = 69.43 to 70.52), with moderate skewness and kurtosis, indicating a more uniform affective response to instruction.

Furthermore, the assumption of homogeneity of variance was tested using Levene's Test across all groups and conditions. Results showed no significant differences in variances for either writing accuracy or language learning motivation scores. All p-values exceeded the .05 threshold across multiple methods—based on the mean (F = 0.728, p = 0.538), median (F = 0.214, p = 0.887), adjusted median (F = 0.214, p = 0.887), and trimmed mean (F = 0.647, p = 0.587)—indicating that the assumption of equal variances was met. This further justifies the use of parametric statistical tests for writing accuracy where normality was also confirmed. Although motivation scores did not meet the normality criterion, the equality of variances implies that any group differences are likely attributable to instructional effects rather than variance instability. This supports the interpretation that CLIL's cognitively demanding nature, while effective for enhancing linguistic competence, may present challenges for maintaining learner motivation over time. Thus, these findings highlight the dual impact of CLIL instruction. While it appears to facilitate improvements in writing accuracy through cognitively engaging tasks, it may also lead to reduced motivation in some learners, particularly those with initially high enthusiasm. This underscores the importance of integrating scaffolding and affective support strategies within CLIL frameworks to sustain learner engagement while ensuring academic and linguistic advancement[79–81].

4.1.4. Detecting Outliers Between Writing Accuracy and Language Learning Motivation

The second step before conducting a parametric test is to eliminate identifiable outliers, which are observations that deviate significantly from the overall distribution pattern[82–84]. Using these formulae: Upper boundary: Q3 + 1.5 × (Q3−Q1) and Lower boundary: Q1−1.5 × (Q3−Q1), where Q3 represents the 75th percentile and Q1 represents the 25th percentile of the scores. The upper and lower boundary values for the pre-test and post-test were determined. The percentiles were obtained through SPSS and are displayed in **Tables 8** and **9**.

Table 8. Percentiles for Writing Accuracy in the Pre-and Post-Tests for Both Groups.

		Percentiles						
		5	**10**	**25**	**50**	**75**	**90**	**95**
Weighted Average (Definition 1)	Pre-Test Experiment	50.05	52.30	61.75	68.00	79.25	86.40	87.40
	Post-Test Experiment	50.50	51.10	62.75	75.50	85.50	89.40	90.45
	Pre-Test Control	49.05	49.60	52.75	70.00	80.75	87.30	89.75
	Post-Test Control	48.20	50.10	68.00	83.50	86.50	90.80	91.45
Tukey's Hinges	Pre-Test Experiment			62.50	68.00	77.50		
	Post-Test Experiment			64.00	75.50	85.50		
	Pre-Test Control			55.00	70.00	80.00		
	Post-Test Control			71.00	83.50	86.00		

Table 9. Percentiles for Language Learning Motivation in the Pre-and Post-Tests for Both Groups.

		Percentiles						
		5	**10**	**25**	**50**	**75**	**90**	**95**
Weighted Average (Definition 1)	Pre-Test Experiment	0.00	14.80	77.00	81.00	84.50	87.80	88.90
	Post-Test Experiment	0.00	0.00	78.00	81.00	83.00	85.40	86.00
	Pre-Test Control	0.00	0.00	75.00	81.00	86.00	89.60	91.80
	Post-Test Control	0.00	0.00	78.00	81.00	84.00	89.00	90.90
Tukey's Hinges	Pre-Test Experiment			78.00	81.00	84.00		
	Post-Test Experiment			79.00	81.00	83.00		
	Pre-Test Control			75.00	81.00	86.00		
	Post-Test Control			78.00	81.00	84.00		

An outlier detection analysis was conducted using Tukey's Outlier Labeling Rule. In the context of writing accuracy, pre-test results showed no outliers, with EG scores ranging from 61.75 to 79.25 and CG scores from 52.75 to 61.75. Post-test results confirmed validity, with EG scores between 62.75 and 85.50 and CG scores between 52.75 and 86.50. Similarly, pre-test scores ranged from 75 to 91.80 for language learning motivation, and post-test scores from 75 to 90, with no extreme values detected. These findings confirm the absence of outliers in both datasets, ensuring their reliability for further statistical analysis [85].

4.1.5. Calculating the Significance of Score Differences

A paired-sample t-test was conducted to calculate the p-value. This test compares two population means from paired samples, such as "before-and-after" measurements [86, 87]. The p-value, ranging from 0 to 1, indicates the significance of the difference between pre-test and post-test scores. A small p-value (≤ 0.05) suggests strong evidence against the null hypothesis, while a larger p-value (> 0.05) indicates weak evidence. Means and standard deviations were calculated using SPSS Version 26, as shown in **Tables 6** and **7**.

In **Tables 10** and **11**, the statistical analysis comparing writing accuracy and language learning motivation between the Experimental Group (EG) and Control Group (CG) showed differing trends. In writing accuracy, the EG's mean scores increased slightly (70.73 to 73.87) with rising variability (SD: 11.30 to 13.41), while the CG showed greater improvement (68.21 to 77.19) with reduced variability (SD: 14.27 to 14.05). Paired t-tests confirmed that the EG's improvement was insignificant ($p = 0.348$), while the CG showed a significant gain ($p = 0.001$), suggesting that external factors influenced CG's performance. To support this, CLIL improves students' written performance, and writing accuracy in particular, except coherence and cohesion [88], but the CLIL environment across disciplines has a positive impact on the students' results in the post-task writing activity [89]. Similarly, the EG's mean scores in language learning motivation dropped slightly (74.83 to 71.35), while the CG saw a marginal increase (69.43 to 70.52). However, paired t-tests showed no significant changes in motivation for either group ($p > 0.05$). The cognitive load of integrating content and language may have hindered writing improvements, while the 12-week duration may have been too short for measurable gains. Although CLIL fosters engagement, it may not directly address key motivational drivers such as self-efficacy and goal-setting [29, 90]. To maximize its effectiveness, longer intervention periods, stronger scaffolding, and targeted motivational strategies may be necessary in EFL settings.

Table 10. Paired Samples Statistics of Writing Accuracy.

		Mean	**N**	**Std. Deviation**	**Std. Error Mean**
EG	Pre-Test Experiment	70.73	23	11.3	2.36
	Post-Test Experiment	73.87	23	13.41	2.8
CG	Pre-Test Control	68.21	21	14.27	3.11
	Post-Test Control	77.19	21	14.05	3.06

Table 10. *Cont.*

Paired Samples Test								
	Paired Differences					t	df	Sig. (2 tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Pre-Test Experiment Post-Test Experiment	−3.13	15.64	3.26	−9.89	3.63	−0.96	22	0.348
Pre-Test Control Post-Test Control	−8.98	10.8	2.36	−13.89	−4.06	−3.81	20	0

Table 11. Paired Samples Statistics of Language Learning Motivation.

		Mean	N	Std. Deviation	Std. Error Mean
EG	Pre-Test Experiment	74.83	23	24	5
	Post-Test Experiment	71.35	23	28.41	5.92
CG	Pre-Test Control	69.43	21	30.29	6.61
	Post-Test Control	70.52	21	29.72	6.49

Paired Samples Test								
Paired Differences								
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2 tailed)
				Lower	Upper			
Pre-Test Experiment Post-Test Experiment	3.48	30.16	6.29	−9.56	16.52	0.55	22	0.59
Pre-Test Control Post-Test Control	−1.10	27.5	6	−13.61	11.42	−0.18	20	0.86

4.1.6. Calculating Within 95% Confidence Intervals

The 95% confidence intervals (CI) were calculated for a within-subjects design. For writing accuracy, the experimental group's CI ranged from −9.89 to 3.63, showing high variability and no statistical significance. In contrast, the control group's CI (−13.89 to −4.06) indicated significant post-test improvement, highlighting the impact of non-CLIL instruction. Similarly, for language learning motivation, the experimental group's CI (−9.89 to 3.63) showed dispersion, whereas the control group's CI (−13.89 to −4.06) confirmed statistically robust gains. These results suggest the control group consistently improved, while the experimental group's variability may reflect individual differences or contextual influences and warrant further investigation. This study aligns with the findings of Camacho et al. and Liu et al. [91, 92], highlighting the impact of different instructional methods on writing accuracy and motivation.

Hedges' g is an effect size measure used to assess the difference between two groups, especially when sample sizes are unequal. It is similar to Cohen's d but includes a correction factor for small sample sizes, improving its accuracy in such situations [93]. The formula for Hedges' g is $\mathbf{g = M_1 - M_2/SD_{pooled}}$, where $\mathbf{M_1}$ and $\mathbf{M_2}$ are the means of the two groups, and $\mathbf{SD_{pooled}}$ is the pooled standard deviation. This measure is commonly used in meta-analyses and research studies to provide a standardized effect size, facilitating comparison across different studies [93].

4.1.7. Identifying the Effect Size

Hedges' g is valuable for correcting bias in small sample sizes, ensuring more reliable results [46, 94]. For the pre-test, Hedges' g was calculated as (68.21−70.73)/12.800519=0.2, and for the post-test, it was (77.19−73.87)/13.718486 = 0.2. Both groups had a Hedges' g of 0.2. Conversely, the Hedges' g values of language learning motivation calculated from the pre-test and post-test data of both groups were Hedges' g = (69.43 − 74.83) / 27.177407 = 0.2 and Hedges' g = (70.52 − 71.35) / 29.04118 = 0.03. indicating small effect sizes and minor differences.

The effect sizes of both writing accuracy and language

learning motivation revealed that the CLIL intervention did not produce significant improvements in either variable. For writing accuracy, the effect sizes measured using Hedges' g were small, with both pre-test and post-test values at g = 0.2. The total effect size of 0.44 indicated only minor improvements, and the control group slightly outperformed the experimental group, though not significantly enough to reject the null hypothesis. Similarly, for language learning motivation, the effect size decreased from g = 0.2 in the pre-test to g = 0.03 in the post-test, resulting in a total effect size of 0.23, suggesting that any observed differences were likely due to random variation rather than the impact of the CLIL intervention. The small effect size may indicate that both groups employ similar sentence processing strategies, especially for sentence types with low cognitive load [95].

The statistical analyses revealed nuanced patterns in the development of writing accuracy across the CLIL (experimental) and non-CLIL (control) groups. Tukey's Outlier Labelling Rule confirmed that both pre- and post-test writing accuracy scores were free from significant outliers, validating the reliability of the data for further parametric testing [73, 82]. Paired-sample t-tests demonstrated that while the CLIL group showed a moderate increase in writing accuracy (M = 70.73 to 73.87), this gain was not statistically significant (p = 0.348). In contrast, the control group exhibited a larger, statistically significant improvement (M = 68.21 to 77.19; p = 0.001), indicating more consistent gains.

Despite the statistical significance favoring the control group, the effect size analysis using Hedges' g revealed small effects for both groups (g = 0.2), suggesting that the differences in performance were modest. This aligns with previous research suggesting that short-term CLIL interventions may not yield significant gains in writing accuracy without sufficient cognitive and linguistic scaffolding [21, 61]. Furthermore, the wide 95% confidence interval for the CLIL group (−9.89 to 3.63) suggests high inter-individual variability, possibly due to differing levels of language proficiency, content familiarity, or learner autonomy.

While CLIL instruction has been found to support grammatical development and syntactic complexity in writing [70, 96], the current findings suggest that a 12-week intervention may be insufficient to produce statistically significant improvements, particularly in aspects such as cause-and-effect relationship clauses, modal auxiliaries, active and passive voice, and specific tenses (present tense, past perfect tense, and future tense) where necessary, or critical thinking. The marginal gains observed may reflect a shift in student awareness of formal accuracy rather than immediate improvement, as CLIL promotes deeper language processing through exposure to academic discourse [1].

Language learning motivation was also assessed across both groups, focusing on both intrinsic and extrinsic dimensions. The quantitative results indicated a slight decline in the CLIL group's mean motivation scores (M = 74.83 to 71.35) and a marginal increase in the control group (M = 69.43 to 70.52). However, paired-sample t-tests revealed no statistically significant changes in motivation for either group (p = 0.59 for CLIL, p = 0.86 for non-CLIL), and the confidence intervals confirmed substantial overlap. These results suggest that motivation was largely stable across the study period, though individual experiences varied.

Effect size calculations further reinforce this interpretation. Hedges' g decreased from 0.2 to 0.03 in the CLIL group, indicating a negligible effect. These findings align with earlier studies that emphasize the complexity of motivation in CLIL settings, especially in short-term implementations [69, 71, 97]. While CLIL may offer cognitive and linguistic benefits, it does not inherently guarantee motivational enhancement unless paired with explicit motivational strategies such as goal setting, feedback, and emotional support [28, 98].

Qualitative data highlighted that some learners found CLIL intrinsically motivating due to increased confidence and task autonomy, while others reported anxiety and confusion, likely stemming from the cognitive demands of dual-focused instruction. This supports the idea that motivation in CLIL is highly context-sensitive and mediated by factors such as instructional clarity, task design, and learners' self-efficacy [99, 100].

Taken together, the findings suggest that while CLIL can foster moderate improvements in writing accuracy, its short-term impact on motivation is less predictable. The small effect sizes and wide confidence intervals indicate that learners' responses to CLIL are highly individualized. To enhance its efficacy, longer intervention periods, increased scaffolding, and targeted motivational techniques should be incorporated. These could include formative feedback loops, peer collaboration, and content-language differentiation tailored to student proficiency levels.

This study supports existing research emphasizing the need for sustained and context-responsive CLIL integration in EFL classrooms [25, 101]. Future studies might adopt a mixed-methods approach over a longer duration to capture deeper motivational shifts and academic progress.

Findings support the hypothesis that CLIL enhances writing quality, syntactic and lexical complexity, and key writing criteria like task achievement and coherence. However, improvements vary, with areas like vocabulary and critical thinking needing additional strategies. Integrating these insights can refine CLIL's impact and inform its application in language education [102–104].

4.2. Qualitative Analysis

This section presents a qualitative analysis of semi-structured interviews conducted with seven student participants and one teacher-participant. The aim was to examine two key dimensions of EFL learning: writing accuracy and language learning motivation, with consideration given to both intrinsic and extrinsic motivational factors. The findings are organized around three principal themes: (a) a comparative analysis of whole-class instruction versus CLIL (Content and Language Integrated Learning) instruction in supporting students' writing development; (b) the impact of the CLIL-based classroom intervention on student engagement, writing performance, and motivation; and (c) the prospects and challenges of implementing CLIL to improve writing skills in the EFL context of Aceh, Indonesia.

4.2.1. Whole-Class Instruction vs. CLIL Instruction on Students' Writing Acquisition

The transition from conventional whole-class instruction to the CLIL approach marked a substantial shift in the classroom environment, particularly in terms of students' writing development. According to the teacher-participant, while the core materials such as grammar structures and writing components remained largely the same, the delivery method under CLIL fundamentally altered the classroom dynamic. She noted, *"There are some similarities with my previous lesson plan in terms of the theory and materials presented. But the difference is in the way how to transfer it in the classroom teaching-learning process"* ***(TP_Q6_271123)***. This reflects a movement away from teacher-centered transmission toward a more interactive and learner-centered instructional model.

Students reinforced this perspective. They indicated that CLIL made language learning more meaningful by embedding grammar instruction within relevant, real-world content. One student stated, *"Presenting a video about learning and then explaining the video as in CLIL"* ***(R09_Q14_281123)***, highlighting how visual and thematic integration enhanced comprehension and retention. Another commented, *"Interesting activities are activities that are done in groups. Because group work will increase the enthusiasm for learning by discussing with friends"* ***(R20_Q14_271123)***, underscoring the motivational value of collaborative learning in CLIL settings.

Nonetheless, while CLIL improved student engagement and contextual language use, challenges emerged in mastering grammatical accuracy. Some students found it difficult to manage the dual cognitive load of processing both content and language without sufficient instructional scaffolding. Weaker students, in particular, struggled to absorb complex language forms embedded in content-heavy lessons. In contrast, whole-class instruction—though more grammar-focused—lacked the engagement and contextual authenticity that CLIL offered. These observations suggest that while CLIL is effective in enhancing motivation and contextual writing ability, its success in improving linguistic precision relies heavily on the teacher's ability to scaffold language learning appropriately.

4.2.2. Impact of the Classroom Intervention Program

The CLIL intervention had a positive and measurable impact on students' participation, engagement, and writing outcomes. Over the 12-week program, students engaged in tasks such as interpreting vocationally relevant videos, composing reflective writing, and participating in group-based technical writing activities. According to the teacher-participant, this content-integrated approach helped students significantly improve their written communication. She observed, *"They are now able to write more effectively and comprehend the lessons better, for instance, they can now describe their daily and past activities in their writing with greater clarity"* ***(TP_Q5_221123)***. The strategic integration of vocational content—such as business procedures and job application skills—with language features

like modal verbs, passive voice, and temporal structures gave students a clear, practical rationale for developing their writing accuracy.

Students also expressed increased enjoyment and motivation. One noted, *"When watching videos and group work. Because watching videos is more fun and group work can share knowledge with friends"* ***(R15_Q14_281123),*** while another stated, *"Learning English is more exciting if we watch videos about learning English with subtitles and materials that are easy to understand"* ***(R13_Q14_281123)***nature of the intervention enhanced both comprehension and motivation.

Moreover, the intervention positively influenced student autonomy and classroom rapport. The teacher remarked, *"They are now more eager to present arguments and confidently support their points with strong examples"* ***(TP_Q9_221123)****,* indicating a growth in student confidence and ownership of learning. However, some limitations were noted, particularly regarding instructional time and lesson complexity: *"Some topics, particularly the lesson on Passive Sentences and Past Perfect, were quite challenging for students... made more difficult by the time constraints"* ***(TP_Q8_221123)***. This suggests that while the CLIL model is pedagogically effective, its implementation requires thoughtful pacing, lesson design, and professional development to be sustainable and impactful.

4.2.3. Prospects of CLIL on Writing Skills Within the EFL Context in Aceh, Indonesia

Looking forward, both teacher and student feedback points to a strong potential for CLIL to enhance writing instruction in Aceh's vocational high schools. The integration of academic language learning with career-relevant content was seen as particularly valuable. The teacher-participant voiced her intent to continue incorporating CLIL principles: *"I may employ and take some ideas in CLIL lessons or methods I will use... So the CLIL method will be one of my references in my teaching methodology"* ***(TP_Q11_221123)***. She further stated, *"The CLIL method is simpler and can enhance students' learning in terms of cognitive and learning motivation"* ***(TP_Q13_221123),*** indicating its accessibility and pedagogical merit.

Students similarly appreciated the real-world relevance of CLIL. One noted, *"Group presentation and when the teacher provides us with certain video animations"* ***(R18_Q14_271123)****,* while others emphasized that CLIL made learning English "less boring" and "more insightful." These responses suggest that CLIL not only aligns well with students' vocational goals but also fosters deeper engagement with writing tasks.

Despite its promise, the implementation of CLIL in Aceh is not without challenges. Key issues include limited teacher proficiency in English, insufficient training in CLIL pedagogy, and a lack of localized instructional resources. Additionally, students with lower proficiency levels expressed difficulty in handling the linguistic and cognitive demands of CLIL lessons. The teacher acknowledged these difficulties, noting that students initially struggled with the complexity of the integrated tasks.

Nevertheless, participants from both groups expressed optimism about CLIL's long-term value. With adequate teacher training, curriculum alignment, and resource development, CLIL can become a sustainable and effective strategy for improving writing outcomes in EFL vocational contexts. Hence, the qualitative findings suggest that CLIL creates a motivational learning environment by addressing both intrinsic factors—such as enjoyment, cognitive stimulation, and confidence—and extrinsic factors, including perceived academic gains and career applicability. However, for these motivational gains to translate into improved performance, appropriate pedagogical scaffolding is essential.

These results are consistent with Swain's Output Hypothesis [105], which posits that language production in meaningful contexts promotes linguistic development. Likewise, Meyer argues that CLIL fosters deeper language processing through content-driven tasks [106], which was evident in students' improved engagement and writing fluency. Supporting this, Siepmann et al. affirm that CLIL enhances learner motivation by integrating real-world themes [107], although Meyer cautions that excessive cognitive load may hinder outcomes unless tasks are carefully calibrated [25].

Taken together, the findings indicate that when appropriately supported, CLIL has the potential to significantly enhance both writing accuracy and language learning motivation among EFL students in vocational settings such as those in Aceh, Indonesia.

4.3. Document Analysis of Post-Writing Accuracy Tests from Six Selected Respondents in Both Groups

The analysis examines the frequency of four elements (cause-and-effect relationship clauses, modal auxiliaries, active and passive voice, and specific tenses (present tense, past perfect tense, and future tense) of students' writing accuracy by selecting six post-test samples—two high (R22 (EG) and R05 (CG)), two middle (R16 (EG) and R18 (CG)), and two low (R14 (EG) and R12 (CG)) scores—after the CLIL intervention to assess its impact on writing accuracy (see **Appendix A Table A3**).

The highest-scoring Experimental Group (EG) students exhibit greater grammatical complexity, frequently using cause-and-effect sentences, modal auxiliaries, passive constructions, and varied tense usage, including present, future, and occasional past perfect. This aligns with the findings of Dalton-Puffer[79], who found that CLIL learners develop more advanced syntactic structures and greater grammatical accuracy. In contrast, the Control Group (CG) predominantly employs present tense with minimal use of modal auxiliaries, passive voice, or past perfect, highlighting EG's superior grammatical accuracy through CLIL. Lorenzo et al. support this by noting that CLIL students show higher syntactic variety and accuracy compared to non-CLIL peers[108]. Moreover, middle-score samples reveal distinct grammatical patterns. EG students use cause-and-effect and passive constructions more frequently but lack modal auxiliaries, future, and past perfect tenses. CG students use modal auxiliaries and cause-and-effect structures but show limited passive voice usage. The EG group demonstrates slightly more varied grammatical structures. This supports findings by Ruiz de Zarobe[109], who observed that CLIL students attempt more complex grammatical structures, even when errors occur. Conversely, lowest-score comparisons show that EG students attempt complexity using cause-and-effect but rely mainly on past tense, lacking modal auxiliaries and passive voice. CG students predominantly use the present tense without complex structures. While both groups display limited grammatical variety, EG students attempt more sophisticated structures, whereas CG writing remains factual and simplistic. This is consistent with the work by Lahuerta[110], who found that CLIL students display greater syntactic complexity and risk-taking in writing, even at lower proficiency levels.

5. Conclusions

This study investigated the impact of CLIL (Content and Language Integrated Learning) instruction on writing accuracy and language learning motivation among vocational high school students in Aceh, Indonesia. The Experimental Group (EG), taught through CLIL, showed modest improvements in writing accuracy but with inconsistent results, while the Control Group (CG) achieved statistically significant and more stable gains under non-CLIL instruction. Language learning motivation changes in both groups were minimal and not statistically significant, with a slight decline observed in the EG, likely due to the cognitive demands of integrating content and language.

Qualitative data revealed that high-performing EG students displayed advanced grammatical structures and reasoning, supporting CLIL's potential to promote academic writing skills. However, many students—especially those with lower proficiency—struggled with cognitive overload, highlighting the need for targeted scaffolding and differentiated instruction. While CLIL's real-world relevance was motivating for some, others found the dual-focus approach overwhelming.

To improve CLIL implementation in vocational EFL settings, teachers should apply tiered scaffolding, use authentic and collaborative tasks, and receive training in balancing language and content. Curriculum designers should develop localized, modular CLIL materials aligned with vocational goals, while policymakers should promote hybrid models, teacher development programs, and investment in supportive resources.

This study was limited by its short intervention period and lack of triangulated data. Future research should examine long-term CLIL effects using mixed methods, explore the integration of adaptive technologies, and investigate motivational strategies that sustain engagement in content-based instruction.

Author Contributions

Conceptualization, F.I. and N.E.M.S.; methodology, N.A.S.; software, N.A.S.; validation, F.I., N.E.M.S., and N.A.S.; Formal analysis, N.E.M.S.; investigation, F.I.; re-

sources, N.A.S.; data curation, N.A.S.; writing—original draft preparation, F.I.; writing—review and editing, N.E.M.S. and N.A.S.; visualization, F.I.; supervision, N.E.M.S. and N.A.S.; project administration, N.A.S.; funding acquisition, N.E.M.S. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the Local Government of Aceh under the Acehnese Resources Empowerment Institution (BPSDM) and the National University of Malaysia grant number [TAP-K017971].

Institutional Review Board Statement

The study was carried out by the main author as a doctoral researcher. The research ethics review was carried out by the internal thesis examiner and the Postgraduate Committee prior to data collection, monitored by both supervisors closely.

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

Data Availability Statement

The data supporting this study's findings are available on request. To access the data, please contact Faisal Idris at p113827@siswa.ukm.edu.my.

Acknowledgments

We acknowledge the Local Government of Aceh under the Acehnese Resources Empowerment Institution (BPSDM), the National University of Malaysia, and related parties in this current study.

Conflicts of Interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

Table A1. Writing Rubric for CLIL Pre-Post Writing Tests.

	Criteria	86–100 Excellent	70–85 Good	51–69 Satisfactory	0–50 Unsatisfactory
CONTENT	**Use of basic subject concepts and knowledge**	Content completely relevant to the topic. Arguments are supplied with examples. Facts justified with appropriate examples. Appropriately used all/main basic concepts. Evidence of understanding principles of the topic.	The topic was discussed successfully. One argument can lack support.	An attempt to discuss the topic has been made. Does not give enough arguments or aspects of the problem.	The topic is not revealed or is misunderstood
	Organization	Clearly and logically arranged. Linkers are varied and used successfully. Proportions observed.	Clear layout. Link words used, though some may be repeated.	It is possible to follow the ideas. Paragraphing may be wrong or absent. Proportions are not observed. Too long or short. (The normal number of words is +/− 10%.)	The organization is hard to follow or trace. Too short.

Table A1. *Cont.*

	Criteria	86–100 Excellent	70–85 Good	51–69 Satisfactory	0–50 Unsatisfactory
LANGUAGE	**Vocabulary and spelling**	Vocabulary is relevant to the topic. A wide repertoire of vocabulary and terms. No spelling mistakes.	Appropriate use of topic vocabulary. Inappropriate use of synonyms may occur. A few spelling mistakes (1–3).	Basic vocabulary is used. Vocabulary repertoire is limited. There are mistakes in using synonyms.	Inappropriate use of basic vocabulary. Sometimes (3 or more cases) spelling inhibits understanding.
	Grammar	Grammar No grammar mistakes (especially in the use of cause and relationship effect clauses, modal auxiliaries, passive and active voice, and tenses (present tense, past perfect tense, and future tense), except for some (1–2) in articles.	Good grammar (especially in the use of cause and relationship effect clauses, modal auxiliaries, passive and active voice, and tenses (present tense, past perfect tense, and future tense). Some mistakes (1–3) may be present.	There are mistakes (5 – 10) in the use of cause and relationship effect clauses, modal auxiliaries, passive and active voice, and tenses (present tense, past perfect tense, and future tense). Some mistakes (1–2) are disturbing	Mistakes prevent understanding of content

Table A2. Sample of CLIL Lesson Plan.

CLIL Lesson Plan **Writing Accuracy class**	
Group: Third-Grade	Time: 4 × 45 minutes
Topic: Cause and effect relationship clauses	
Previous Knowledge/Skills: Dependent and independent clauses	
Aim(s): • To introduce cause and effect relationship clauses • To compare similarities and differences to the functions of cause-and-effect relationship clauses • To compose texts of written transactional interactions that involve the act of giving and asking for information related to causal relationships	
Teaching Objectives	Learning Outcomes
A. Content	**A. Content**
• The function of cause-and-effect relationship clauses in learning English. • The use of cause-and-effect relationship clauses in written communication	• Students acquire the vocabulary related to the unit • Students are encouraged to write simple sentences consisting of cause-and-effect relationship vocabulary such as because of, owing to, to identify the 'Why' question, etc.
B. Cognition (Functions)	**B. Cognition**
• Evaluate: determine the possible outcomes of processes • Analyse (write sentences based on the given topic) • Understand: arrange sentences from the words given and identify the differences among them	• Students can describe the use of cause-and-effect relationship clauses in certain discourse or utterances • Students understand the function cause-and-effect relationship clauses in contexts • Students identify and arrange properly the cause-and-effect relationship clauses in a short paragraph
C. Communication	
C.1 Language of Learning: Key Vocabulary Writing Accuracy: *because of…, due to …, in the reason of …, since…owing to…*	

Table A2. *Cont.*

C.2 Language for Learning (=language needed to operate in the learning environment or in a particular lesson)

- Language to combine the sentences
- Language to compose a short paragraph

For example, I have a terrible headache because of ...

- Language to explain the function of cause-and-effect relationship clauses
- Language to explain how to write cause and effect relationship sentences.

a. because of …

b. due to …

c. in the reason of …

d. since …

e. owing to …

C.3 Language Through Learning
Language for carrying on worksheets

D. Culture/Citizenship

- To associate the contents of the lesson with real-world situations
- To be aware and meticulous while writing cause-and-effect relationship sentences.
- To be respectful with partners when working in groups/individually

Materials & resources
Internet hotspot, laptop, English book Grade-Three, worksheets

Teaching plan (type, timing & sequence of activities)

Stages and time	Lesson procedure (describe the activities and instructional strategies)	Justification for the activity (content, language, cognition, culture)
Lead-in 10 minutes	Watch YouTube videos from native or EFL/ESL about cause-and-effect relationship clauses	To warm up To revise cause-and-effect relationship clauses
Presentation 20 minutes	Presentation of the concept of cause-and-effect relationship clauses	The purpose of this activity is to present the concept of cause-and-effect relationship clauses. The teacher picks up and shows 5 sample sentences from a YouTube video and writes them to the board and draws the similarities and the differences among them. The teacher explains that cause-and-effect relationship clauses have similarities and differences while writing sentences or paragraphs and asks students to connect a sentence as in 'most students this year failed an exam due to/because/because of …. The teacher asks a high-level thinking activity as in 'write several sentences that consist of cause-and-effect relationship clauses'
Practice 1 10 minutes	Worksheet Listen and point activity	The purpose of this activity is to listen and point out the words of cause-and-effect relationship clauses
Practice 2 25 minutes	English book Grade-Three: Reading certain texts that consist of cause-and-effect relationship clauses	The purpose of this activity is to listen to the teacher reading a book about cause-and-effect relationship clauses. The teacher asks questions while reading the book about the cause-and-effect relationship clauses that appear in the text. Where are the cause-and-effect relationship clauses located? Students also mention the number of cause-and-effect relationship clauses in the text that has been read.

Table A2. *Cont.*

Practice 3 20 minutes	Write less than two or three sentences of each cause-and-effect relationship clause.	The purpose of this activity is to encourage Ss to choose cause-and-effect relationship clauses and produce them in sentences. The teacher writes a model of the sentence on the board, for instance: • I have a terrible headache *because of* my sleeping time all day. • The dam burst *owing to* the heavy rain • Owing to the snow, the train was delayed, etc.
Conclusion 5 minutes	Post-Activities	• Recalling students' cognitive learning • Language of motivation
Assessment: The teacher checks on understanding and adjusts instructions to keep students on track. No grades or scores are given. Incidental observation and guiding students' understanding.		

Table A3. Document Analysis.

The Highest Score (EG)	The Highest Score (CG)
I ***will tell*** about the memorable experiences while studying english in my school. It started when I was 1st grade, the teacher who taught English in my class at that time was Mr. Rikki. Actually, he used teach Japanese at my school, but since the COVID-19 virus exists, he wasn't teach Japanese anymore at my school. And then at that time, he started to teach English. Studying English with Mr. Rikki was really fun and I really enjoyed that study. I ***was improved*** my English too, and at that time I got very good score. Continued when I ***was studied*** English in 2nd grade, the teacher who ***was taught*** English was not Mr. Rikki anymore, but ***has been replaced*** by Mrs. Ti Aminah. There weren't many interesting things when I was studied English in 2nd grade, it***'s*** ***because*** Mrs. Ti Aminah rarely teach English in my class. Besides being an English teacher, Mrs. Ti Aminah also ***serves*** ***as*** deputy principal for student affairs, therefore she rarely ***teaches*** in my class. Honestly, I was a little disappointed, ***because*** when I was in 2nd grade, my English scores dropped quite a bit. And last, when I ***was started*** learn English in the 3rd grade, the teacher who taught English ***has been replaced*** again. Currently, the teacher who ***teaches*** English in my class *is* Mrs. Zainatuddar or usually called Mrs. Ina. *So far*, studying with Mrs. Ina ***has been*** very enjoyable. But now, English lessons ***are sometimes taught*** by Mr. Faisal, *because* Mr. Faisal ***is*** doing his research to continue his study. Hopefully, in this 3rd grade, I ***can improve*** more about my English skills, and I ***can get*** good scores.	When I was I Junior high School. That was the first time I ***was interested and wanted to*** study it well and truly. Previously, I ***was not interested*** in learning english. ***because*** I ***feel*** studying english ***is*** boring. ***I don't know*** if it***'s*** ***because*** I ***don't understand*** it easily. but that all changed when entered junior high School. why ***I am interested***? And what ***make*** me serious about learning english well? ***Because*** the language used to be boring and I didn't understand. In fact that day and now it actually attracted my attention to learn it ***because*** it saw that my friends ***could speak*** english and I thought it was quite cool, really, that ***become*** a motto for me to want to study it seriously. My teacher's way of teaching ***is*** very unique and easy to understand. Sometimes we ***learn*** through the cartoon video ***are*** very funny and we *laugh* together. It *is* the most unforgettable moment and he also always asked questions if we didn't understand well. But it was different when ***he was asked*** to repeat the material he was very scary. the reason ***is*** ***because*** when we ***couldn't repeat*** it he was very angry which different when he taught. That why he often ***ask*** questions if there *is* something he ***doesn't understand***. And that***'s*** why sometimes I***'m*** really afraid to come to the lesson. But I ***don't give up*** easily, ***because I am interested*** in the subject. I ***am*** serious about studying it myself. ***So*** there ***is*** a good improvement in me from before, than I ***was given*** praise for my improvement learning English and now I *like* it.
The Middle Score (EG)	**The Middle Score (CG)**
I ***will share*** my experience of learning English through out my life while at school. English ***is*** one of the subjects that I ***am*** least interested in. But since middle school I have liked hearing English songs from older sister. And before there was no Google, now you ***can't find*** the lyrics and translation of the song. But I just sang along based on the pronunciation I heard. In junior high school, English lessons have become mandatory. But unfortunately my English teacher ***is not*** a student-centered Englis teacher so I ***don't like*** it. Because of the lesson ***is*** just doing questions in the book. Sometime it***'s*** not explained it***'s*** just corrections together exchanges with friends. The result ***is*** that nothing changes in my English. When entered vocational school in first grade, I studied with Rikki Sensei, the experience of studying with Rikki Sensei was very fun and exciting. And when I entered second grade I didn't take English lessons because of practicum. After going up to third grade I studied with Mrs. Zaina, I ***learn*** a lot with her from cause to effects and others.	I first learned English when I was in junior high school. now I ***am*** in vocational school and I thought middle school and vocational school english ***are*** different. Vocational english *is* more memorable and I ***am*** quite afraid of that lessons. But when my english teacher first ***come*** into the class I ***feel*** nervous ***because*** I was not an experts in this area of english. I thought english was difficult and boring. The first time my englsih teacher came in. we ***were all told*** to introduce ourselves using english. I saw my friend who was the first to come forward who wanted to introduce himself. She ***can learn*** english very well, ***because of*** that I was very nervous ***because*** I was afraid of speaking English. what I used was wrong, but my teacher said it was okay. Ita was wrong we were studying, that***'s*** when my nervousness started to subside. After the lesson ***was finished*** we all went home and when I got home I told my mother about what happened that day. this ***is*** my first experience or bring ***able to fix*** my nervousness about learning English.

Table A3. *Cont.*

The Lowest Score (EG)	The Lowest Score (CG)
Monday I went to school on a motorbike, I had an accident, ***because*** I was riding very fast. I suffered a very serious injury to my honda motor bike ***doesn't have*** **to** be fast, ***because*** there ***will be*** very high risk.	English ***is the most widely spoken*** language in the world, this language ***is*** the mother tongue of more than 400 million people throughout the world. Every day, millions of people ***use*** English at work and in social life when heads of State meet each other at world-level conferences.

References

[1] Coyle, D., Hood, P., Marsh, D., 2010. CLIL: Content and Language Integrated Learning. Cambridge University Press: Cambridge, UK. Available from: https://www.cambridge.org/core/product/identifier/9781009024549/type/book (cited 20 March 2025).

[2] de Larios, J.R., Coyle, Y., García, V., 2022. The effects of using cognitive discourse functions to instruct 4th-year children on report writing in a CLIL science class. Studies in Second Language Learning and Teaching. 12(4), 597–622. DOI: https://doi.org/10.14746/ssllt.2022.12.4.4

[3] Richards, J.C., Rodgers, T.S., 2014. Approaches and Methods in Language Teaching, 3rd ed. Cambridge University Press: Cambridge, UK.

[4] Puspitasari, E., Pratolo, B.W., Mahfiana, A.M.I., 2020. Teacher's Belief about the Implementation of Curriculum 2013 in English Language. Ethical Lingua: Journal of Language Teaching and Literature. 7(1), 101–111. DOI: https://doi.org/10.30605/25409190.171

[5] Yulistio, D., Fhitri, A., 2019. Improving Short Story Writing Skills Using the Genre Pedagogy, Scientific, and CLIL (Content and Language Integrated Learning) Learning Models in Grade XI Students of Senior High School No. 2 Bengkulu City [Peningkatan Kemampuan Menulis Cerpen Menggunakan Model Pembelajaran Pedagogi Genre, Saintifik, dan CLIL (Content and Language Integrated Learning) pada Siswa Kelas XI SMAN 2 Kota Bengkulu]. Jurnal Ilmiah KORPUS. 3(1), 9–20. DOI: https://doi.org/10.33369/jik.v3i1.7342

[6] Muhammad, K., Wahyu, S., Heru, S., 2020. The Influence of Genre Pedagogical and Content and Language Integrated Learning on Elementary Students' Writing Skills. Seloka: Jurnal Pendidikan Bahasa dan Sastra Indonesia. 9(2), 168–175. DOI: https://doi.org/10.15294/seloka.v9i2.40099

[7] Foster, P., Skehan, P., 1999. The influence of source of planning and focus of planning on task-based performance. Language Teaching Research. 3(3), 215–247. DOI: https://doi.org/10.1191/136216899672186140

[8] Sulistiyo, U., 2016. Learning English as a Foreign Language in an Indonesian University: A Study of Non-English of Department Students' Preferred Activities Inside and Outside the Classroom. IJET (Indonesian Journal of English Teaching). 5, 1–26.

[9] Budianto, S., Sulistyo, T., Widiastuti, O., et al., 2020. Written corrective feedback across different levels of efl students' academic writing proficiency: Outcomes and implications. Studies in English Language and Education. 7(2), 472–485.

[10] Widiati, U., Cahyono, B.Y., 2006. The teaching of EFL writing in the Indonesian context: The state of the art. Jurnal Ilmu Pendidikan. 13(3), 139–150.

[11] Fitria, V.N., Susilawati, 2019. Video Sharing in Content and Language Integrated Learning (CLIL) Context: Fostering Junior High School students' productive skills. Proceedings of the Second Conference on Language, Literature, Education, and Culture (ICOLLITE 2018); 24–25 October 2018; Bandung, Indonesia. pp. 282–286. DOI: https://doi.org/10.2991/icollite-18.2019.62

[12] Dörnyei, Z., Al-Hoorie, A., 2017. The Motivational Foundation of Learning Languages Other Than Global English: Theoretical Issues and Research Directions. Modern Language Journal. 101(3), 455–568. DOI: https://doi.org/10.1111/modl.12408

[13] Alrabah, S., Wu, S., 2017. Promoting Writing Competence and Positive Attitudes among College Students in a CLIL English Course. International Journal of English Linguistics. 7(1), 58–68. DOI: https://doi.org/10.5539/ijel.v7n1p58

[14] Chumbay, J., Ochoa, J.F.Q., 2020. Language-Driven CLIL: Developing Written Production at the Secondary School Level. English Language Teaching. 13(8), 74–90. DOI: https://doi.org/10.5539/elt.v13n8p74

[15] Coyle, D., 2006. Content and language integrated learning: Motivating learners and teachers. Scottish Languages Review. 13(5), 1–18.

[16] Faisal, I., Ehsan, M.S.N., Ainil, S.N., 2025. Content and language integrated learning (CLIL) on the acquisition of writing skills in secondary education: a systematic literature review. Journal of Education and Learning (EduLearn). 19(3), 1449–1458. DOI: https://doi.org/10.11591/edulearn.v19i3.21974

[17] Nugroho, A., 2020. Content and Language Integrated Learning Practice in English for Accounting Course.

International Journal of Emerging Technologies in Learning. 9(2), 172–181. DOI: https://doi.org/10.15642/IJET2.2020.9.2.172-181

[18] Mukadimah, H., Sahayu, W., 2021. Embedding Content and Language Integrated Learning (CLIL) into English for specific purposes curriculum for vocational high school. Jurnal Pendidikan Vokasi. 11(2), 204–212.

[19] Mackey, A., Gass, S.M., 2015. Second Language Research: Methodology and Design, 2nd ed. Routledge: New York, NY, USA.

[20] Creswell, J.W., Creswell, J.D., 2017. Research design: Qualitative, quantitative, and mixed methods approaches, 5th ed. Sage Publications: London, UK.

[21] Lo, Y.Y., Macaro, E., 2015. Getting used to content and language integrated learning: what can classroom interaction reveal? The Language Learning Journal. 43(3), 239–255. DOI: https://doi.org/10.1080/09571736.2015.1053281

[22] Pérez-Cañado, M.L., 2017. CLIL Teacher Education: Where do we Stand and Where do we Need to Go? In: Gómez, M.E., Johnstone, R. (eds.). Bilingual Education: Educational Trends and Key Concepts. Ministerio De Educacion, Cultura Y Deporte: Madrid, Spain. pp. 129–144.

[23] Park, H., 2023. Content-language integrated learning (CLIL) in EFL primary English writing based on coding theme. The Korea Association of Primary English Education. 29(4), 179–200. DOI: https://doi.org/10.25231/pee.2023.29.4.179

[24] Ebenberger, A., 2017. Developing Individual Language Competences via Task-Based Learning and Content and Language Integrated Learning (CLIL). Lublin Studies in Modern Languages and Literature. 41(1), 174–190. DOI: https://doi.org/10.17951/lsmll.2017.41.1.174

[25] Meyer, O., 2010. Towards quality CLIL: successful planning and teaching strategies. Pulso Revista de educación. 33, 11–29. DOI: https://doi.org/10.58265/pulso.5002

[26] Navarro-Pablo, M., López Gándara, Y., 2020. The effects of CLIL on L1 competence development in monolingual contexts. The Language Learning Journal. 48, 18–35. DOI: https://doi.org/10.1080/09571736.2019.1656764

[27] Yadi, K., Yayan, S., 2017. Improving students' factual report writing skill by using content and language integrated learning (CLIL) Method. Indonesian EFL Journal. 3(1), 21–30. DOI: https://doi.org/10.25134/ieflj.v3i1.650

[28] San Isidro, X., Lasagabaster, D., 2022. Students' and families' attitudes and motivations to language learning and CLIL: a longitudinal study. Language Learning Journal. 50(1), 119–134. DOI: https://doi.org/10.1080/09571736.2020.1724185

[29] Doiz, A., Lasagabaster, D., Sierra, J.M., 2014. CLIL and motivation: The effect of individual and contextual variables. Language Learning Journal. 42(2), 209–224. DOI: https://doi.org/10.57087/Verbeia.2021.4395

[30] Sylvén, L.K., 2017. Motivation, second language learning and CLIL. In: Llinares, A., Morton, T. (eds.). Applied Linguistics Perspectives on CLIL. John Benjamins: Amsterdam, Netherlands. pp. 51–56.

[31] Dornyei, Z., Ushioda, E., 2014. Teaching and Researching Motivation, 2nd ed. Routledge: London, UK.

[32] Pladevall-Ballester, E., 2019. A longitudinal study of primary school EFL learning motivation in CLIL and non-CLIL settings. Language Teaching Research. 23(6), 765–786. DOI: https://doi.org/10.1177/1362168818765877

[33] Hu, H., Ehsan, M.S.N., Harwati, H., 2023. Sustaining Content and Language Integrated Learning in China: A Systematic Review. Sustainability. 15(5), 3894. DOI: https://doi.org/10.3390/su15053894

[34] Navarro Pablo, M., 2018. Are CLIL Students More Motivated? An Analysis of Affective Factors and their Relation to Language Attainment. Porta Linguarum Revista Interuniversitaria de Didáctica de las Lenguas Extranjeras. 29, 71–90. DOI: https://doi.org/10.30827/Digibug.54023

[35] Kanoksilapatham, B., Khamkhien, A., 2022. CLIL implemented and scaffolded in an EFL higher education context: Engineering research methodology course. Indonesian Journal of Applied Linguistics. 11(3), 502–514. DOI: https://doi.org/10.17509/ijal.v11i3.36913

[36] Coelho, M., 2017. Scaffolding Strategies in CLIL Classes–supporting learners towards autonomy. In: Ribeiro, M.C.A., Gonçalves, A., Silva, M.M., (eds.). Languages and the Market: A ReCLes.pt Selection of International Perspectives and Approaches. ReCLes.pt: Lisbon, Portugal. pp. 106–114.

[37] Hu, H., Harwati, H., Li, X., 2025. Bridging language and learning: A CLIL-ized TMI model for international students in traditional Chinese medicine. Journal of Medical Education Development. 18(1), 140–146.

[38] Del Pozo Beamud, M., 2021. Are CLIL Teachers-to-be Motivated? A Qualitative Study. Babel – AFIAL: Aspectos de Filoloxía Inglesa e Alemá. 30, 47–67. DOI: https://doi.org/10.35869/afial.v0i30.3702

[39] Zaripova, R., Salekhova, L., Grigoriev, S., et al., 2019. Increasing academic motivation through Integrated Language and Content Learning (CLIL) and Information and Communication Technologies mediated by the constructivist approach. Dilemas Contemporáneos: Educación, Política y Valores. 6, 1–12.

[40] Mahan, K.R., 2020. Teaching Content and Language

Integrated Learning (CLIL): Classroom practices and student perspectives in three Norwegian classrooms [PhD Thesis]. University of South-Eastern Norway: Kongsberg, Norway.

[41] Hita, S., 2021. Implementing CLIL in the Foreign Language Class: Teaching English through History [MSc thesis]. Hellenic Open University: Patras, Greece.

[42] Evi, P., 2016. Classroom Activities in Content and Language Integrated Learning. Journal of Foreign Language Teaching and Learning. 1(2), 1–13. DOI: https://doi.org/10.18196/ftl.129

[43] Dörnyei, Z., 2007. Research methods in applied linguistics. Oxford University Press: Oxford, UK.

[44] Nunan, D., 1990. Second Language Classroom Research. ERIC: Washington, DC, USA.

[45] Yunisrina, Q.Y., Nurul, I., Mutiarani, 2020. Teachers' reinforcement: Building students' motivation to learn English. International Journal of Language Studies. 14(4), 105–128.

[46] Ehsan, M.S.N., 2019. The Effects of Differentiated Instruction on Students' Language Attitude and Critical Thinking in an ESL Context [PhD Thesis]. University Of Sheffield: Sheffield, UK.

[47] Edelsbrunner, P.A., Simonsmeier, B.A., Schneider, M., 2024. The Cronbach's Alpha of Domain-Specific Knowledge Tests Before and After Learning: A Meta-Analysis of Published Studies. Educational Psychology Review. 37(1), 1–43. DOI: https://doi.org/10.31234/osf.io/m8d7t

[48] Malapane, T.A., Ndlovu, N.K., 2024. Assessing the Reliability of Likert Scale Statements in an E-Commerce Quantitative Study: A Cronbach Alpha Analysis Using SPSS Statistics. Proceedings of the 2024 Systems and Information Engineering Design Symposium (SIEDS); 03 May 2024; Charlottesville, VA, USA. IEEE: New York, NY, USA. pp. 90–95. DOI: https://doi.org/10.1109/SIEDS61124.2024.10534753

[49] Barkovska, I., 2012. CLIL Secondary School. In: Quartapelle, F. (ed.). Assessment and Evaluation in CLIL. Ibis: Como, Italy. pp. 117–126.

[50] Mashuri, S., Sarib, M., Rasak, A., et al., 2022. Semi-structured Interview: A Methodological Reflection on the Development of a Qualitative Research Instrument in Educational Studies. IOSR Journal of Research & Method in Education (IOSR-JRME). 12(1), 22–29. DOI: https://doi.org/10.9790/7388-1201052229

[51] Cardno, C., 2019. Policy Document Analysis: A Practical Educational Leadership Tool and a Qualitative Research Method. Educational Administration: Theory and Practice. 24(4), 623–640. DOI: https://doi.org/10.14527/kuey.2018.016

[52] Onwuegbuzie, A., Leech, N., Collins, K., 2015. Qualitative Analysis Techniques for the Review of the Literature. The Qualitative Report. 17, 56.

[53] Hamed, A., 2023. Qualitative Document Analysis of Teacher Appraisal Practice in Omani Schools: Identifying Challenges for Best Practices. Journal of Southwest Jiaotong University. 58(2), 713–727. DOI: https://doi.org/10.35741/issn.0258-2724.58.2.67

[54] Liu, S., Zu, Y., 2024. Evaluation Models in Curriculum and Educational Program - A Document Analysis Research. Journal of Technology and Humanities. 5(1), 32–38. DOI: https://doi.org/10.53797/jthkkss.v5i1.4.2024

[55] Melekhina, E., Kazachikhina, I., 2021. CLIL Works In Teaching ICT In Pre-Service Language Education. In: Kolmakova, O., Boginskaya, O., Grichin, S. (eds.). Language and Technology in the Interdisciplinary Paradigm, vol 118. European Proceedings of Social and Behavioural Sciences. European Publisher: London, UK. pp. 460–467. DOI: https://doi.org/10.15405/epsbs.2021.12.57

[56] Yang, W., 2019. Evaluating Contextualized Content and Language Integrated Learning Materials at Tertiary Level. Latin American Journal of Content & Language Integrated Learning. 11(2), 236–274. DOI: https://doi.org/10.1007/s42321-021-00091-5

[57] Llinares, A., Nikula, T., 2024. CLIL students' production of cognitive discourse functions: Comparing Finnish and Spanish contexts. Language and Education. 38(3), 381–400. DOI: https://doi.org/10.1080/09500782.2023.2211049

[58] Lasagabaster, D., Doiz, A., 2017. A Longitudinal Study on the Impact of CLIL on Affective Factors. Applied Linguistics. 38(5), 688–712. DOI: https://doi.org/10.1093/applin/amv059

[59] Tsuchiya, K., Pérez Murillo, M.D., 2019. Content and language integrated learning in Spanish and Japanese contexts: Policy, practice and pedagogy. Palgrave Macmillan: London, UK.

[60] Banegas, D.L., del Pozo Beamud, M., 2020. Content and Language Integrated Learning: A Duoethnographic Study about CLIL Pre-Service Teacher Education in Argentina and Spain. RELC Journal. 53(1), 151–164. DOI: https://doi.org/10.1177/0033688220930442

[61] Coyle, D., Meyer, O., 2021. Beyond CLIL: Pluriliteracies teaching for deeper learning. Cambridge University Press: Cambridge, UK.

[62] San Isidro, X., Lasagabaster, D., 2019. The impact of CLIL on pluriliteracy development and content learning in a rural multilingual setting: A longitudinal study. Language Teaching Research. 23(5), 584–602. DOI: https://doi.org/10.1177/1362168817754103

[63] Otwinowska, A., Foryś, M., 2017. They learn the CLIL way, but do they like it? Affectivity and cognition in upper-primary CLIL classes. International Journal of Bilingual Education and Bilingualism. 20(5), 457–480. DOI: https://doi.org/10.1080/13670050.

2015.1051944

[64] Lo, Y.Y., Lui, W.-M., Wong, M., 2019. Scaffolding for cognitive and linguistic challenges in CLIL science assessments. Journal of Immersion and Content-Based Language Education. 7(2), 289–314.

[65] Martínez-Adrián, M., Gutiérrez-Mangado, M.J., Gallardo-del-Puerto, F., et al., 2021. Language-related episodes by young CLIL learners. Language Teaching for Young Learners. 3(2), 214–245. DOI: https://doi.org/10.1075/ltyl.20005.mar

[66] Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). Biometrika. 52(3–4), 591–611. DOI: https://doi.org/10.1093/biomet/52.3-4.591

[67] Orcan, F., 2020. Parametric or Non-parametric: Skewness to Test Normality for Mean Comparison. International Journal of Assessment Tools in Education. 7(2), 255–265. DOI: https://doi.org/10.21449/ijate.656077

[68] Tukey, J.W., 1977. Exploratory Data Analysis. Addison-Wesley: Reading, MA, USA.

[69] Lasagabaster, D., 2010. Innovation in Language Learning and Teaching English achievement and student motivation in CLIL and EFL settings. Innovation in Language Learning and Teaching. 5(1), 3–18.

[70] Pérez-Cañado, M.L., 2018. Key variables in CLIL implementation and Research: Recent Perspectives on contextual, cognitive and affective variation in CLIL. Porta Linguarum. 29, 5–7.

[71] Ushioda, E., Dörnyei, Z., 2017. Beyond Global English: Motivation to Learn Languages in a Multicultural World: Introduction to the Special Issue. Modern Language Journal. 101(3), 451–454. DOI: https://doi.org/10.1111/modl.12407

[72] Lilliefors, H.W., 1967. On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. Journal of the American Statistical Association. 62(318), 399–402. DOI: https://doi.org/10.2307/2283970

[73] Field, A., 2013. Discovering statistics using IBM SPSS statistics. Sage Publications: London, UK.

[74] Pallant, J., 2020. SPSS Survival Manual, 7th ed. Routledge: London, UK.

[75] Daraini, A.M., Fauziyati, E., Rochsantiningsih, D., 2021. Students' Perception of The Implementation of Content and Language Integrated Learning (CLIL) at Senior High School. AL-ISHLAH: Jurnal Pendidikan. 13(1), 41–48. DOI: https://doi.org/10.35445/alishlah.v13i1.404

[76] Marsh, D., 2002. CLIL/EMILE – The European dimension: Actions, trends and foresight potential. European Commission: Brussels, Belgium.

[77] Lasagabaster, D., Doiz, A., 2016. CLIL students' perceptions of their language learning process: delving into self-perceived improvement and instructional preferences. Language Awareness. 25(1–2), 110–126. DOI: https://doi.org/10.1080/09658416.2015.1122019

[78] Pittas, E., Tompkins, L., 2024. A systematic review of student learning outcomes in CLIL in LOTE. Frontiers in Education. 9, 1447270. DOI: https://doi.org/10.3389/feduc.2024.1447270

[79] Dalton-Puffer, C., 2007. Discourse in Content and Language Integrated Learning (CLIL) Classrooms. John Benjamins: Amsterdam, Netherlands.

[80] Genesee, F., Hamayan, E., 2016. CLIL in context: Practical guidance for educators. Cambridge University Press: Cambridge, UK.

[81] Karina, R.M., 2023. Instructional scaffolding in CLIL: An overview of theory and research. Routledge: London, UK.

[82] Sidnyaev, N.I., Enkhzhargal, B., 2024. Methodology for detecting and removing outliers in statistical studies. Dependability. 24, 4–9.

[83] Mohseni, N., Nematzadeh, H., Akbarib, E., et al., 2023. Outlier Detection in Test Samples using Standard Deviation and Unsupervised Training Set Selection. International Journal of Engineering. 36, 119–129.

[84] Moore, D.S., McCabe, G.P., 1989. Introduction to the practice of statistics. W.H Freeman: New York, NY, USA.

[85] Aguinis, H., Gottfredson, R.K., Joo, H., 2013. Best-Practice Recommendations for Defining, Identifying, and Handling Outliers. Organizational Research Methods. 16(2), 270–301.

[86] Shier, R., 2004. Paired t-tests. Mathematics Learning Support Centre. Available from: www.statstutor.ac.uk/resources/uploaded/paired-t-test.pdf (cited 20 March 2025).

[87] Rumsey, D., 2010. Statistics Essentials For Dummies. Wiley: Indianapolis, IN, USA.

[88] Corral Robles, S., 2019. New Insights into Written Competence in CLIL and non-CLIL Programmes: Pedagogical Implications. ReiDoCrea: Revista electrónica de investigación Docencia Creativa. 8(3), 289–304. DOI: https://doi.org/10.30827/Digibug.58466

[89] Mcbride, S., 2017. Writing-to-Learn Content in a CLIL Spanish Compulsory Secondary Education Environment. An Exploratory Study. In: García Manzanares, N., Vinuesa Benítez, V. (eds.). El Bilinguismo a debate:Actas del IV Congreso Internacional de Enseñanza Bilingüe en Centros Educativos (CIEB). Servicio de Publicaciones: Murcia, Spain. pp. 1–26.

[90] Arribas, M., 2016. Analysing a whole CLIL school: Students' attitudes, motivation, and receptive vocabulary outcomes. Latin American Journal of Content & Language Integrated Learning. 9(2), 267–292. DOI: https://doi.org/10.5294/laclil.2016.9.2.2

[91] Camacho, A., Alves, R.A., Boscolo, P., 2021. Writing

Motivation in School: a Systematic Review of Empirical Research in the Early Twenty-First Century. Educational Psychology Review. 33(1), 213–247. DOI: https://doi.org/10.1007/s10648-020-09530-4

[92] Liu, Z.-M., Hwang, G.-J., Chen, C.-Q., et al., 2024. Integrating large language models into EFL writing instruction: effects on performance, self-regulated learning strategies, and motivation. Computer Assisted Language Learning. 1–25. DOI: https://doi.org/10.1080/09588221.2024.2389923

[93] Borenstein, M., Hedges, L.V., Higgins, J.P.T., et al., 2009. Introduction to Meta-Analysis. Wiley: Chichester, UK.

[94] Hedges, L.V., 1981. Distribution Theory for Glass's Estimator of Effect size and Related Estimators. Journal of Educational Statistics. 6(2), 107–128.

[95] Ping, C.X., Zahrin, N.H., Rusli, Y.A., 2024. Sentence Comprehension Performance in Malay School-Age Children: A Preliminary Study. GEMA Online Journal of Language Studies. 24(4), 43–61. DOI: https://doi.org/10.17576/gema-2024-2404-03

[96] Jexenflicker, S., Dalton-Puffer, C., 2010. The CLIL differential: Comparing the writing of CLIL and non-CLIL students in higher colleges of technology. In: Dalton-Puffer, C., Nikula, T., Smit, U. (eds.). Language Use and Language Learning in CLIL Classrooms. John Benjamins: New York, NY, USA. pp. 169–190. DOI: https://doi.org/10.1075/aals.7.09jex

[97] Lo, N., Chan, S., 2024. Gamification for Higher Education Applications: A Social Learning Theory Approach. In: Martindale, T., Amankwatia, T.B., Cifuentes, L., et al. (eds.). Handbook of Research in Online Learning: Insights and Advances. Brill: Leiden, Netherlands. pp. 576–610. DOI: https://doi.org/10.1163/9789004702813_025

[98] Sylvén, L.K., Thompson, A.S., 2015. Language learning motivation and CLIL. Journal of Immersion and Content-Based Language Education. 3(1), 28–50. DOI: https://doi.org/10.1075/jicb.3.1.02syl

[99] Lo, Y.Y., Lin, A.M.Y., 2015. Special issue: Designing multilingual and multimodal CLIL frameworks for EFL students. International Journal of Bilingual Education and Bilingualism. 18(3), 261–269. DOI: https://doi.org/10.1080/13670050.2014.988111

[100] Lasagabaster, D., 2017. Integrating content and foreign language learning: What do CLIL students believe? Journal of Immersion and Content-Based Language Education. 5(1), 4–29. DOI: https://doi.org/10.1075/jicb.5.1.01las

[101] Dalton-Puffer, C., 2018. Putting CLIL into Practice. ELT Journal. 1(72), 109–111. DOI: https://doi.org/10.1093/elt/ccx063

[102] Coyle, Y., Roca de Larios, J., 2020. Exploring young learners' engagement with models as a written corrective technique in EFL and CLIL settings. System. 95, 102374. DOI: https://doi.org/10.1016/j.system.2020.102374

[103] Sato, T., Hemmi, C., 2022. Development of second language productive skills through CLIL in a Japanese university: a pre-experimental longitudinal study. Language Learning in Higher Education. 12(1), 309–326. DOI: https://doi.org/10.1515/cercles-2022-2040

[104] Lee, J., 2020. Assessing the effects of CLIL on Korean high school students' writing. Linguistic Research. 37, 89–112. DOI: https://doi.org/10.17250/khisli.37.202009.004

[105] Swain, M., 2005. The Output Hypothesis: Theory and Research. In: Hinkel, E. (ed.). Handbook of Research in Second Language Teaching and Learning. Routledge: London, UK. pp. 471–483.

[106] Dalton-Puffer, C., 2011. Content-and-Language Integrated Learning: From Practice to Principles? Annual Review of Applied Linguistics. 31, 182–204. DOI: https://doi.org/10.1017/S0267190511000092

[107] Siepmann, P., Rumlich, D., Matz, F., et al., 2023. Attention to diversity in German CLIL classrooms: multi-perspective research on students' and teachers' perceptions. International Journal of Bilingual Education and Bilingualism. 26(9), 1080–1096. DOI: https://doi.org/10.1080/13670050.2021.1981821

[108] Lorenzo, F., Casal, S., Moore, P., 2009. The effects of content and language integrated learning in European education: Key findings from the Andalusian bilingual sections evaluation project. Applied Linguistics. 31(3), 418–442. DOI: https://doi.org/10.1093/applin/amp041

[109] Ruiz de Zarobe, Y., 2010. Written production and CLIL: An empirical study. In: Dalton-Puffer, C., Nikula, T., Smit, U. (eds.). Language use and language learning in CLIL classrooms. John Benjamins: Amsterdam, Netherlands. pp. 191–210.

[110] Lahuerta, A., 2020. Analysis of accuracy in the writing of EFL students enrolled on CLIL and non-CLIL programmes: the impact of grade and gender. The Language Learning Journal. 48(2), 121–132. DOI: https://doi.org/10.1080/09571736.2017.1303745