International Journal of Learning, Teaching and Educational Research
Vol. 21, No. 4, pp. 150-167, April 2022
https://doi.org/10.26803/ijlter.21.4.9
Received Feb 11, 2022; Revised Apr 8, 2022; Accepted Apr 19, 2022

Killing Two Birds with One Stone? A Study on Achievement Levels and Affective Factors in Content and Language Integrated Learning (CLIL)

Hengzhi Hu*, Nur Ehsan Mohd Said and Harwati Hashim
Universiti Kebangsaan Malaysia, Bangi, Malaysia

Abstract. Despite the burgeoning research evidence on the multifarious effects of Content and Language Integrated Learning (CLIL) in academia, only limited empirical evidence can be retrieved from the Chinese context. The intent of this study conducted in a Chinese higher education provider was to probe into the effects of CLIL on English proficiency and learning motivation in the College English Teaching context. Characterised by a quasi-experimental pretest-posttest design, this study involved 60 undergraduates who were categorised into either the High-Group, Medium-Group or Low-Group based on their English proficiency. Data were collected from standardised English language tests and a questionnaire on motivation for English learning. Quantitative analyses involving the running of t-tests and ANOVA mainly indicated that: 1) CLIL had a generally positive effect on the participants' English proficiency and motivation; 2) the participants from the three subgroups all made significant progress in English proficiency with the low achievers showing increased considerable improvement; 3) CLIL negligibly affected the learners' positive attitude to language learning; 4) the higher achievers appeared to display stronger motivation towards English learning at the end of the study than those with a relatively lower level of English proficiency. It was concluded that CLIL has the potential to benefit language learners of different levels in both effective learning and affective learning, while a critical attitude should be held to the view that CLIL is the panacea for all with an appeal for more classroom-based research to enrich the performance and affective evidence of CLIL.

Keywords: College English Teaching; Content and Language Integrated Learning; English proficiency; motivation

1. Introduction

In China, English language education (ELE) has been regarded as significantly important ever since its legitimisation in the early twenty-first century. Since this

* Corresponding author: *Hengzhi Hu, p108937@siswa.ukm.edu.my*

point, English as a foreign language (EFL) has been a compulsory subject for Chinese students involved in primary, secondary and higher education with a publicly shared view that the proficiency in English as a lingua franca is a personal and national asset that can facilitate globalisation and internationalisation at home. In the upsurge of interest in ELE, College English (CE) is a topic of unfading interest and everlasting importance to Chinese policymakers, scholars and researchers (Qi, 2021).

CE is an integral part of the Chinese higher education agenda and "a required basic (EFL) course for undergraduate students" whose main subjects are not pertinent to English studies (Li & Xiao, 2020, p. 1720). Further, the delivery of CE courses is known as College English Teaching (CET). In the new century, dynamic reforms have been made in CET with a national endeavour to optimise curriculum design and implementation. A shared feature of the latest CE policies, such as *College English Curriculum Requirements* (Ministry of Education of the People's Republic of China, 2007) and *Guidelines for College English Teaching* (Ministry of Education of the People's Republic of China, 2020), is that CE should be regarded as not only an English course designed for the well-rounded development of language proficiency but also as a conduit for learners to acquire content knowledge and develop cross-cultural understanding. This specific nature reflects the philosophy of Content and Language Integrated Learning (CLIL) characterised by a dual teaching task as one of its properties (Coyle et al., 2010) and provides top-down support for the implementation of this pedagogical approach in CET.

Although considerable significance has been attached to CET, chances are that traditional English teaching still plays a dominant role in practice. Official reporting and domestic literature have revealed a failure to sufficiently provide students with proficiency in the English language (Qi, 2021; Wang & Xu, 2020), and both sources claim that Chinese university students' English competencies are far from satisfactory. Despite this issue, the fact that a number of Chinese students who have learned English for years and are able to achieve excellent performance in English assessments are demotivated for language learning, a study from He (2018) leads to the re-examination of the effects of CET pedagogies on emotions aimed at figuring out an approach to achieve both language learning and affective learning. Thus, given the political requirements and the issues faced by CE learners, the purpose of this study featured by pretest-posttest quasi-experimental designs is to examine the effects of CLIL on English proficiency and the motivation for language learning. This falls into the category of performance evidence and affective evidence identified by Coyle et al. (2010) for a CLIL research agenda, which refers to how students perform and how they feel in a CLIL programme respectively. This study is assumed to be significant and can bridge the gap that most previous and ongoing explorations of CLIL in China's academic agenda face, given their limit that exists at the theoretical level (Hu, 2021a). This study also prioritises the discussion of whether this is an ideal pedagogical approach for ELE through some classroom-based research that examines the multifarious effects of CLIL, thus yielding practical insight as to its implementation and popularisation.

2. Literature Review

According to Marsh (2003, p. 15), CLIL "refers to any dual-focused educational context in which an additional language, thus not usually the first language of the learners involved, is used as a medium in the teaching and learning of non-language content", and it is "dual-focused because whereas attention may be predominantly on either subject-specific content or language, both are always accommodated". In other words, CLIL emphasises both language learning and content learning in a classroom, and this feature distinguishes it from other pedagogical instructions, such as English-Medium Instruction that merely emphasises the learning of content knowledge through the medium of English and Content-Based Instruction and treats language learning as the ultimate goal for achievement through the vehicle of content knowledge (Brown & Bradford, 2017). From this perspective, scholars explain that CLIL offers learners an opportunity to learn the content subjects through the target language (TL) and to learn and use the TL meaningfully and purposefully through the conduit of content knowledge (Hu, 2021a; Maasum et al., 2012; Suliman et al., 2018).

Language learning is an essential part of CLIL. Although it falls into the category of Communicative Language Teaching and is characterised by a heavy focus on language meaning, the teaching and learning of linguistic forms as an essential part of language education should also be achieved in CLIL. A focus-on-form approach tends to invite language learners and teachers to take linguistic forms (e.g., grammatical rules) into prime consideration, as opposed to a focus-on-meaning method that facilitates language acquisition by exposing learners into abundant and authentic TL input and output (Celik, 2019). Tensions can easily occur in this dichotomy, while CLIL offers a solution in that "it is not a question of whether to focus on meaning or form but rather that it is fundamental to address both, the balance of which will be determined by different variables in specific CLIL settings" (Coyle et al., 2010, p. 67). Otherwise stated, the broad language teaching objective in CLIL has a dual focus on both form and meaning, the combination of which could contribute to the development of a learner's language proficiency. This assumption is embedded in the view that "form-focused instruction is therefore generally considered most effective when embedded in communicative contexts and is thus clearly distinguished from decontextualized grammar lessons" (Lyster, 2006, p. 40). When language learning needs are thoughtfully planned and attended to by teachers, student TL proficiency could be improved, thus facilitating the comprehensive development of listening, speaking, reading and writing skills (Dalton-Puffer & Smit, 2013).

Abundant research has proved the language benefits of combining language learning with content learning (e.g., Liu, 2019; Martyn, 2018; Suliman et al., 2020; Yunus & Sukri, 2017). However, although academia takes the positive effects of CLIL on language development for granted (Ostovar-Namaghi & Nakhaee, 2019), previous research has also revealed that CLIL was unable to help achieve the language learning objectives as anticipated. For instance, Kamal's (2020) latest study of a CLIL-oriented university programme revealed EFL learners' poor performance in language assessment and their sceptical attitudes to the effects of CLIL on language learning, a finding that aligns with Setiawan's (2013) research

that shows the double challenge of dealing with content and language learning could fail to improve EFL learners' language proficiency. Further, some studies have indicated the 'selective' benefits of CLIL, suggesting that CLIL only serves advanced learners. For example, both Mewald's (2007) and Zydatiß's (2012) studies have shown that academically weak learners were unable to make as much progress in language proficiency as their more advanced peers did and suggested that there should be a threshold for the admission of CLIL learners to ensure teaching efficiency. This contrasts with Hu's (2021b) latest research finding that demonstrates if CLIL programmes are differentiated as per learners' needs, significant academic progress can be seen across all proficiency levels. This contradiction of research findings necessitates the re-examination of the effects of CLIL on language learning and learners of differing levels.

Despite the potential of CLIL to promote language proficiency, it is also considered to be engaging and motivational for language learners (Marsh, 2003). In the field of language learning, motivation is defined as a combination of motivational intensity (i.e., the efforts made to learn the TL), desire to learn the TL and a positive attitude towards the TL (Gardner, 1985). Although CLIL places heavy emphasis on the integration of content and language learning, affective learning pertinent to a learner's positive attitudes is also a nonnegligible feature of it. According to Mehisto et al. (2008, p. 29), CLIL could create "a safe and enriching learning environment" wherein learners' confidence could be built through "the experiment with content and language". This has been confirmed in some empirical studies, such as the ones conducted by Martyn (2018) and Liu (2019), which have displayed learners' positive feelings in and through perceptions of CLIL. However, motivation is still an under-researched field in CLIL, despite it being one of the hottest topics in bilingual education (Lasagabster, 2020). In other words, despite the fact that most researchers and practitioners seem to have acquiescently agreed on the affective advantage of CLIL and on the interpretation that motivation enhancement is one of the cornerstones in CLIL embedded in the nature of the mutually beneficial integration of content and language "to ensure more learners are motivated to learn and use other languages in the future" (Coyle, 2013, p. 23), there is not yet sufficient empirical classroom-based evidence to vindicate this view, resulting in a gap yet to be bridged.

Bearing this brief review in mind, and to meet the research objective, the authors of this paper wish to answer the following questions in the study:

- What effects will CLIL have on students' English proficiency and motivation for English learning?
- At the end of the study, is there any statistical difference in English proficiency and motivation for English learning amongst the learners of different language levels?

3. Methodology

3.1 Research Design

To examine the effects of CLIL on English proficiency and motivation for English learning, quasi-experimental designs characterised by the non-random pre-selection of participants and the organisation of pretest and posttest were

employed to investigate the cause-effect relationship between variables (Indhiarti & Sudarwati, 2021). This gave the researchers the power over the variables to control them and was considered appropriate for this study to determine whether CLIL could affect students' English proficiency and motivation for language learning.

3.2 Research Site and Participants

This study was conducted in a comprehensive higher education provider situated in a major Chinese city, wherein various CLIL programmes were provided for undergraduates. Based on the eligibility criteria that the participants' main courses should not be pertinent to English studies and that they would not take any other English classes other than the ones provided in the studied programme, a sample of 60 students was recruited from the Faculty of Law with informed consent and put into a CLIL programme called *CE with American Law*. They were put into two classes as per the university's administrative policies but were taught by the same teacher to ensure that they were instructed by the same methodology. The sample consisted of 37 females and 23 males who were in the third year of their undergraduate studies and the fifth semester of CE learning. They had passed College English Test Band-4 (CET-4), a national standardised English test, and had similar demographic information, such as age (an average of 21 years old) and years of English learning (approximately 8 to 12 years).

3.3 Instruments

The first instrument used in the study was a mock College English Test Band-6 (CET-6), papers that were adapted from authentic test batteries (Wang, 2017). CET-6 is a national standardised test developed by China's educational authorities to measure undergraduates' English proficiency of listening, speaking reading and writing. The listening and reading sections were objective, while the writing and speaking sections were subjective and assessed based on certain rubrics. CET-6 is highly rated in academia for its assessment validity and reliability with regard to its authentic paper design, assessment structure and administration (Zhang et al., 2019). This means the score earned by a candidate could be a good representation of their English proficiency. Although using standardised tests to measure learners' learning outcomes in CLIL has been criticised for its lack of validity due to the misgiving that taught material may not be covered in assessment, Coyle et al. (2010) assume that the answer to the 'how to assess' question is context-dependent, giving practitioners the autonomy to use standardised tests in their own educational contexts and to justify their practice. Given that passing CET-6 is regarded as an important indicator of Chinese undergraduates' academic success from a personal, institutional and societal perspective (Chen & Webb, 2017), the implementation of it in this study could be an effective way to not only measure students' English ability but also offer referential implications. Prior to this research, a pilot study had been carried out, indicating that pretest and posttest papers were reliable and that the writing (K = 0.74, 0.71, respectively) and speaking (K = 0.77, 0.76, respectively) tests had acceptable inter-rater reliability. In original CET-6, the written test weighs 710 marks with the listening, reading and writing tests occupying 35%, 35% and 30% of the total score respectively; the full mark for the speaking test is 15, which is

calculated separately from the written test score. In this study, all the data were recorded and analysed as raw scores.

The other instrument was the *Questionnaire on English Learning Motivation* designed based on Gardner's (1985) Attitude and Motivation Test Battery, including 30 items on a six-point Likert scale ranging from 'strongly disagree' (scale point 1) to 'strongly agree' (scale point 6). Three constructs, namely motivational intensity (MI), desire to learn English (DTLE) and attitudes to English (ATE), were measured as per the definition of motivation for foreign languages. This questionnaire was originally prepared in English. To ensure the subjects could fully understand the items and thus respond to them properly, Chinese copies were prepared by three professional Chinese-English translators who implemented the back-to-back translation technique (Pradeep, 2021). The pilot study suggested the translated Chinese questionnaire had an acceptable internal consistency for each construct (Cronbach alpha = 0.72, 0.71, 0.75, respectively) and the entire questionnaire (Cronbach alpha = 0.81).

3.4 Research Procedures

The study lasted for nine weeks in an academic semester from March to May 2021. The administration of the pretests and posttests was done in regular class time. To ensure the assessment validity, reliability and fairness, the English proficiency tests were organised, invigilated and marked by professional staff at the research site in line with the CET-6 regulations. The administration of the questionnaire was carried out with the assistance of the CLIL teacher in this study. All the participants completed the pretests and posttests. The collected data were then organised and processed by Statistical Package for the Social Sciences 25.0, the results of which are recorded in the following tables with descriptive data and inferential data. Inferential data analyses were a principal part of this study, and t-tests and ANOVA were run to answer the research questions.

4. Results

4.1 English Proficiency of the Total Group

As shown in Table 1 and Table 2, a two-tailed, paired samples t-test with an alpha level of .05 was firstly run to compare the pretest (M = 456.80, SD = 39.16) and posttest (M = 471.17, SD = 39.72) scores of 60 individuals. On the whole, the participants' average posttest score of English proficiency was 14.37 higher than the pretest score. The difference was statistically different, $t(59) = -8.04$, $p < .001$, $d = .36$. The same analysis was also carried out to examine each section of the English tests, demonstrating that the participants significantly improved their proficiency in English listening ($p = .001$), reading ($p < .001$), writing ($p < .001$) and speaking ($p < .001$) after the treatment of CLIL.

Table 1: Paired samples statistics of total group's English proficiency tests

		Mean	Std. Deviation
Total Score	Pretest	456.803	39.164
	Posttest	471.168	39.724
Listening	Pretest	155.295	19.740
	Posttest	160.633	18.280
Reading	Pretest	151.783	20.764

	Posttest	155.975	21.126
Writing	Pretest	141.217	19.637
	Posttest	145.693	17.380
Speaking	Pretest	8.508	1.508
	Posttest	8.867	1.359

Table 2: Paired samples test of total group's English proficiency tests

	Mean (Pretest-Posttest)	Std. Deviation	t	Sig. (2-tailed)
Total Score	-14.365	13.836	-8.042	.000
Listening	-5.338	11.432	-3.617	.001
Reading	-4.192	4.425	-7.338	.000
Writing	-4.477	6.689	-5.184	.000
Speaking	-0.358	0.402	-6.901	.000

4.2 English Proficiency of the Subgroups

To further examine the effects of CLIL on students of different language levels, the participants were divided into three groups based on their pretest English proficiency scores, namely High-Group (H-Group) with a higher English proficiency and achieved the top 30% pretest scores (N = 18); Low-Group (L-Group) with a lower English proficiency and whose English pretest scores were within the bottom 30% of the total group's (N = 18); and Medium-Group (M-Group) consisting of the rest of the participants who had an average English proficiency (N = 24). A one-way ANOVA (see Table 3) was run, indicating the average pretest scores of the three subgroups were statistically different (p < .001). Post hoc analyses with Tukey's HSD use of an α of .05 (see Table 4) further confirmed the significant statistical difference of pretest scores amongst the subgroups (p < .001) and demonstrated that students had significantly different levels of English proficiency prior to the study.

Table 3: ANOVA of subgroups' average pretest scores

	Sum of Squares	Mean Square	F	Sig.
Between Groups	73925.625	36962.812	127.157	.000
Within Groups	16596.175	290.687		
Between Groups	90494.799			

Table 4: Multiple comparisons of subgroups' average pretest scores

Turkey HSD	(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.
	L	M	-47.613	5.316	.000
		H	-90.583	5.683	.000
	M	L	47.613	5.316	.000
		H	-42.971	5.316	.000

Each group's pretest and posttest average scores were analysed by a paired samples t-test. The descriptive data in Table 5 and the inferential data in Table 6 indicated that all the three subgroups improved their English proficiency after the intervention. Specifically, the L-Group's posttest score (M = 431.94, SD = 34.69) was 21.36 points higher than the pretest score (M = 410.58, SD = 20.40), and the difference was statistically significant, $t(17)$ = -4.59, p < .001, d = .78. Similarly, the M-Group's posttest score (M = 468.60, SD = 10.37) was 10.41 points higher than

the pretest score (M = 458.20, *SD* = 9.50), with a significant statistical difference, $t(23) = -5.35$, $p < .001$, $d = 1.05$. Likewise, the H-Group's posttest score ($M = 513.81$, $SD = 23.75$) was 12.64 points higher than the pretest score ($M = 501.17$, $SD = 20.89$). The difference was also statistically significant, $t(17) = -6.19$, $p < .001$, $d = .57$. Figure 1 demonstrated the proficiency gains of the subgroups and illustrated that they all made progress in the studied programme. However, it should be noted that the L-Group made more noticeable progress than the other groups, suggesting that the effects of CLIL could show increased effects for low achievers.

Table 5: Paired sample statistics of subgroups' English proficiency tests

Group		Mean	Std. Deviation
L	Pretest	410.583	20.401
	Posttest	431.944	34.687
M	Pretest	458.196	9.497
	Posttest	468.604	10.372
H	Pretest	501.167	20.891
	Posttest	513.811	23.749

Table 6: Paired samples test of subgroups' English proficiency tests

Group	Mean (Pretest-Posttest)	Std. Deviation	t	Sig. (2-tailed)
L	-21.361	19.726	-4.594	.000
M	-10.408	9.539	-5.346	.000
H	-12.644	8.667	-6.189	.000

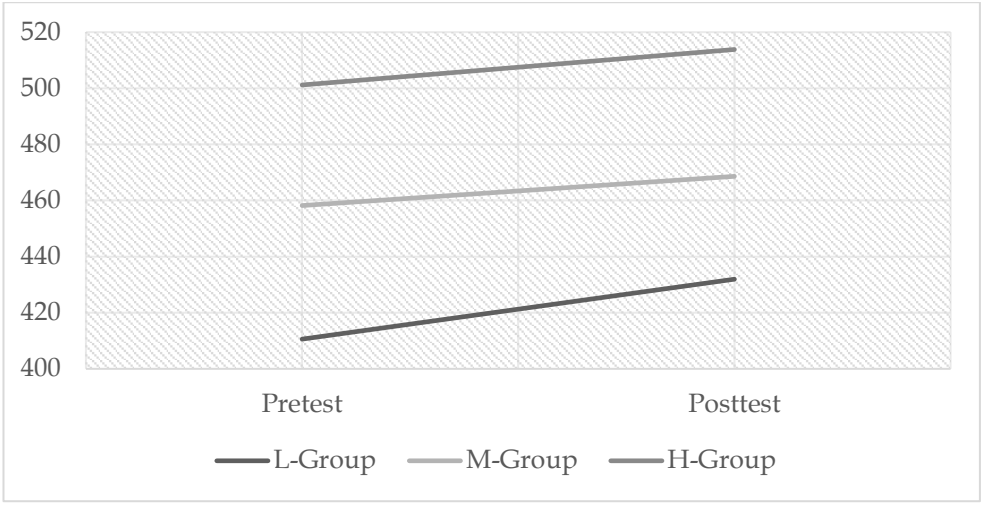

Figure 1. Proficiency gains of the subgroups

To further explore the effects of CLIL on the subgroups' English proficiency, a Welch's ANOVA (see Table 7) was run to analyse the average posttest scores under the conditions that the normality assumption was not violated but the homogeneity assumption was violated. It indicated that there were significant differences in the subgroups' posttest English proficiency scores ($p < .001$). Then, post hoc tests (see Table 8) were run to further explore the differences amongst the subgroups with Games-Howell (using an α of .05), and results indicated that there was a significant difference between the L-Group with the M-group (p =

.001) and the H-Group (p < .001), and between the M-group with the H-Group (p < .001). This illustrated that, by the end of the study, the H-Group still had the highest English proficiency, and that the L-Group had the lowest language proficiency.

Table 7: Welch's ANOVA of subgroups' average posttest scores

	Statistic	df1	df2	Sig.
Welch	40.182	2	26.792	.000

Table 8: Multiple comparisons of subgroups' English proficiency posttest

	(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.
Games-Howell	L	M	-36.660	8.446	.001
		H	-81.867	9.909	.000
	M	L	36.660	8.446	.001
		H	-45.207	5.985	.000

4.3 Motivation of the Total Group

Firstly, a paired samples t-test was run to compare the whole group's questionnaire data collected before and after this study. Regarding the participants' MI, as shown in Table 9 and Table 10, the average posttest score (M = 29.42, SD = 3.26) was 2.2 points higher than the pretest score (M = 27.22, SD = 2.79). The difference was statistically significant, $t(59) = -3.90$, p < .001, d = .73. Similarly, there was statistically significant difference in DTLE between the posttest (M = 27.32, SD = 4.01) and the pretest (M = 25.5, SD = 3.63), $t(59) = -2.48$, p = .016, d = .48. The descriptive data illustrated that the posttest score of ATE (M = 35.63, SD = 5.31) was higher than the pretest scores (M = 34.58, SD = 6.52), whereas inferential statistics did not demonstrate a significant difference of this construct, $t(59) = -1.48$, p = .144, d = .18. Generally, the data indicated that the participants had higher motivation score in the posttest (M = 92.37, SD = 9.81) than in the pretest (M = 87.3, SD = 8.20), $t(59) = -3.95$, p < .001, d = .56.

Table 9: Paired samples statistics of total group's motivation tests

		Mean	Std. Deviation
MI	Pretest	27.22	2.793
	Posttest	29.42	3.264
DTLE	Pretest	25.50	3.629
	Posttest	27.32	4.006
ATE	Pretest	34.58	6.523
	Posttest	35.63	5.314
Total Score	Pretest	87.30	8.199
	Posttest	92.37	9.813

Table 10: Paired samples test of total group's motivation tests

	Mean (Pretest-Posttest)	Std. Deviation	t	Sig (2-tailed)
MI	-2.2	4.376	-3.895	.000
DTLE	-1.817	5.682	-2.476	.016
ATE	-1.05	5.491	-1.481	.144
Total Score	-5.067	9.937	-3.95	.000

4.4 Motivation of the Subgroups

A paired samples t-test was also done separately for each subgroup to explore the level of changes in motivation towards English learning. Table 11 and Table 12 presented the descriptive statistics and the inferential statistics respectively.

For the L-Group, the test revealed that the participants had a more positive ATE with their posttest score (M = 32.61, SD = 5.22), a result that was higher than the pretest one (M = 29.11, SD = 5.37). The difference was statistically significant, $t(17)$ = -2.56, p = .02, d = .44. However, their MI decreased as the posttest score (M = 26.33, SD = 2.91) was significantly lower than the pretest one (M = 28.72, SD = 2.11), $t(17)$ = 2.84, p = .011, d = .95. Likewise, the average score of DTLE in the posttest (M = 23.39, SD = 2.48) was lower than that in the pretest (M = 26.44, SD = 4.26), $t(17)$ = 2.90, p = .01, d = .91. Regarding their overall motivation for English learning, however, there was no significant statistical difference between the pretest (M = 84.28, SD = 9.43) and posttest (M = 82.33, SD = 5.86), $t(17)$ = .84, p = .41, d = .26.

The M-Group had a higher level of MI after the intervention with their posttest score (M = 29.92, SD = 2.02) higher than the pretest one (M = 26.75, SD = 2.80), $t(23)$ = -6.26, p < .001, d = 1.31. In contrast, no statistical difference could be found between the pretest and posttest scores of DTLE (p = .23) and ATE (p = .87). However, the group generally displayed a stronger motivation towards English learning after the treatment of CLIL, as the posttest score (M = 91.63, SD = 6.04) was significantly higher than the pretest one (M = 87.92, SD = 8.51), $t(23)$ = -2.44, p = .023, d = .51.

For the H-Group, there was a significant difference between the pretest and posttest scores of MI (p < .001) and DTLE (p < .001). Although the posttest score of ATE (M = 39.50, SD = 2.28) was higher than the pretest one (M = 39.22, SD = 2.76), no statistical difference was found, $t(17)$ = -.45, p = .66, d = .11. On the whole, in relation to motivation, there was a significant statistical difference between their posttest score (M = 103.59, SD = 3.82) and the pretest score (M = 89.50, SD = 5.59), $t(17)$ = -9.90, p < .001, d = 1.50.

Table 11: Paired samples statistics of subgroups' motivation tests

			Mean	Std. Deviation
L-Group	MI	Pretest	28.72	2.109
		Posttest	26.33	2.910
	DTLE	Pretest	26.44	4.260
		Posttest	23.39	2.477
	ATE	Pretest	29.11	5.368
		Posttest	32.61	5.215
	Total Score	Pretest	84.28	9.430
		Posttest	82.33	5.861
M-Group	MI	Pretest	26.75	2.801
		Posttest	29.92	2.020
	DTLE	Pretest	25.96	3.432
		Posttest	26.71	1.967
	ATE	Pretest	35.21	6.447

		Posttest	35.00	5.405
	Total Score	Pretest	87.92	8.505
		Posttest	91.63	6.035
H-Group	MI	Pretest	26.33	2.910
		Posttest	31.83	2.526
	DTLE	Pretest	23.94	2.796
		Posttest	32.06	1.924
	ATE	Pretest	39.22	2.756
		Posttest	39.50	2.282
	Total Score	Pretest	89.50	5.586
		Posttest	103.39	3.822

Table 12: Paired samples test of subgroups' motivation tests

		Mean (Pretest-Posttest)	Std. Deviation	t	Sig. (2-tailed)
L-Group	MI	2.389	3.567	2.842	.011
	DTLE	3.056	4.478	2.895	.01
	ATE	-3.500	5.803	-2.559	.02
	Total Score	1.944	9.771	.844	.41
M-Group	MI	-3.167	2.479	-6.258	.000
	DTLE	-.750	2.982	-1.232	.23
	ATE	.208	6.366	.160	.874
	Total Score	-3.708	7.434	-2.444	.023
H-Group	MI	-5.500	3.204	-7.283	.000
	DTLE	-8.111	3.359	-10.246	.000
	ATE	-.278	2.630	-.448	.66
	Total Score	-13.889	5.950	-9.904	.000

To further examine the effects of CLIL on students of different English proficiency levels, a one-way ANOVA (see Table 13) was run to analyse their general levels of motivation for learning. It indicated that there was not any significant statistical difference between the subgroups' average scores regarding motivation in the pretest, F (2, 57) = 2.01, p = .144, η^2 = .07. However, Table 14 indicated that the three subgroups had rather different levels of motivation at the end of the study, F (2, 57) = 68.47, p < .001, η^2 = .71. Post hoc analyses with Tukey's HSD (see Table 15) further showed that the L-Group had a lower level of motivation than the M-Group (p < .001) and H-Group (p < .001) and that the M-Group had less motivation than the H-Group (p < .001). In other words, at the end of the study, the L-Group was the least motivated in contrast to the H-Group that was the most motivated, and the M-Group scored the middle. This meant the subgroups' general level of motivation for English learning was influenced differently by CLIL.

Table 13: One-way ANOVA of subgroups' motivation pretest

		Sum of Squares	F	Sig.
Total Score	Between Groups	260.656	2.005	.144
	Within Groups	3705.944		
	Total	3966.600		

Table 14: One-way ANOVA of subgroups' motivation posttest

		Sum of Squares	F	Sig.
Total Score	Between Groups	4012.031	68.473	.000
	Within Groups	1669.903		
	Total	5681.933		

Table 15: Multiple comparisons of subgroups' motivation posttest

Turkey HSD					
Dependent Variable	(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.
Total Score	L	M	-9.292	1.688	.000
		H	-21.056	1.804	.000
	M	L	9.292	1.688	.000
		H	-11.764	1.688	.000

5. Discussion

Given the above data analyses, the first major finding is that CLIL had a positive effect on the participants' general English proficiency and that their macro language skills all improved. This result corresponds to the data generated from previous research that CLIL is able to promote well-rounded ELE and language development (Goris et al., 2019; Ostovar-Namaghi & Nakhaee, 2019), as well as to the theoretical assumptions that CLIL could attend to language learners' needs of developing receptive and productive skills (Mehisto et al., 2008). Also, the result of this study rejects some contradictory voices which have arisen from empirical studies that CLIL "does not even demonstrate convincingly the foreign language benefits" (Bruton, 2013, p. 587). However, it must be acknowledged that no matter how carefully and sensitively the lessons are planned, the implementation of CLIL may not always provide the results expected in terms of language gains or necessarily guarantee that learners' language proficiency across all the skills can improve (Pižorn, 2017). Instead, a critical view should be held to the efficiency of CLIL, and it is not a panacea for all language learners or educational contexts. However, this study has at least reinforced the potential of CLIL in language learning and offered some, although few, educational implications for the application of CLIL in CET.

The participants with differing levels of language proficiency all benefited from CLIL and improved their English proficiency, though at the end of the intervention there was still a gap between the relatively advanced learners and the lower achievers. This is contradictory to the critical voice that CLIL only attracts and benefits high-achieving learners (Goris et al., 2019). Meanwhile, the research finding that academically weak learners actually made more remarkable progress in learning than their more advanced peers further confirm the very rare assumption that weaker English learners could benefit more from CLIL than their advanced peers (Blasco, 2014). Elitism in CLIL tends to be a vital issue in current academia and is characterised by the long-standing debate about whether CLIL should be selective. Both Bruton (2011) and Paran (2013) assume that the potential of CLIL can be only explored in elitist educational contexts when implemented selectively with overachievers and that the criterion-based selection of students is the prerequisite for the success of CLIL. Indeed, such a selective implementation

of CLIL still frequently occurs in practice, and it has attracted a number of highly motivated and advanced learners (Goris, 2019; Kamis et al., 2021), with empirical research findings rationalising and necessitating this phenomenon. For example, by analysing the reasons for students to drop out of CLIL programmes, Zydatiß (2012) suggests that there should be a threshold of academic competence for CLIL admission, below which learners may easily find difficulty with their learning. However, the positive finding is that the less advanced learners also made academic progress in this research, a result that aligns with the results yielded in a few but not many classroom-based studies (Dewi & Sudarmaji, 2020; Hu, 2021b; Karabassova, 2019). This illuminates the potential of CLIL to benefit a range of learners other than overachievers as well as foregrounding the need to re-examine the "commonly harboured beliefs vis-à-vis the elitism of" CLIL (Cañado, 2019, p. 1).

Another finding of this study is that CLIL had a productive effect on the participants' motivation for English learning, especially on MI and DTLE. This corresponds to previous findings that CLIL could considerably strengthen learners' motivation for language learning and create an engaging learning environment (Young, 2018). However, it should be noted the participants who had relatively advanced English proficiency benefited more from CLIL than those with low English proficiency with respect to motivation enhancement. Specifically, the L-Group learners held a more positive ATE at the end of this study but had decreased MI and DELE as well as an unchanged level of motivation in general. Although the M-Group learners generally had stronger motivation, their DTLE and ATE did not change. In contrast, the students involved in the H-Group benefitted more from CLIL with an increased level of motivation. In this sense, it could be said that the affective benefits of CLIL were not equal for all. This finding contradicts other research that showed low achievers could emotionally benefit more from CLIL (McDougald, 2015; Jäkel, 2015) and challenges the assumption that the motivational effect of CLIL works for all learners (Hamidavi et al., 2016). Instead, it is consistent with the finding reported by Yang (2015) that although CLIL students who had relatively low English proficiency did not fall behind in their academic performance, they did feel more discouraged and less motivated than their more advanced peers. It is also interesting to note that the participants' ATE did not change after the study, suggesting that CLIL had little effect on enhancing learners' positive views of English learning. This finding challenges the widely accepted view and evidence that CLIL learners could have more favourable attitudes to either learning in general or language learning in particular (Marsh, 2003; Urgal, 2019). CLIL scholars and practitioners seem to have agreed on the emotional benefits of CLIL and have always taken them for granted. However, the findings reported in this paper bring to light the very rare assumption that language learners' self-confidence and learning attitude in CLIL may be negatively affected (Coyle et al., 2010), necessitating a critical review of the effects of CLIL and the re-examination of the CLIL panacea in further research.

To a large extent, the aforementioned successful results both in language learning and affective learning stem from the nature of CLIL and the fact that it offers not

only increased exposure to the TL but also a high-quality solution. The natural use of language in CLIL could augment a learner's motivation. When they are interested in the topic of the CLIL class, they will be encouraged to acquire the TL to communicate. This makes CLIL methodology different from traditional language lessons in the manner that students "learn to use language and use language to learn" by replicating "the conditions to which infants are exposed when learning their first language" (Mehisto et al., 2008, p. 26). In this regard, the focus on content learning rationalises the aim for language learning and use and creates a safe and participatory learning environment wherein learners' motivation is greatly increased. In return, stronger motivation contributes to language learning (Liu, 2019), and a link can most likely be drawn between the participants' higher level of English proficiency and motivation, especially in the greater MI shown after the intervention. Theoretically, these benefits are achieved, because CLIL "provides a cognitively challenging situation which is associated with a meaningful use of the foreign language and an improved sense of achievement", "promote(s) fruitful discussions on pedagogical issues and practices" and "provides teachers and students with a sense of ownership of their teaching practice and the learning process" (Lasagabster, 2020, p. 348).

A possible explanation to the finding that learners having relatively lower English proficiency did not enhance their motivation as anticipated could be that the cognitive and linguistic challenge of learning EFL through the content subject might frustrate or demotivate CLIL learners. However, little information has indicated the intimate relationship between cognition and affectivity, and thus further research is needed. A more convincing explanation could be that language learners may easily feel disappointed when they realise their new language competencies do not match the time and efforts invested into learning (Mattheoudakis, 2019), which aligns relatively well with Lasagabster's (2020) research findings that show when CLIL learners are unable to manage linguistic demands, demotivation spontaneously occurs. Therefore, the way in which CLIL teachers encourage and motivate low achievers or those that lag behind is a closely related issue here that requires further investigation.

6. Limitation

The first limitation of this study was about the positivist nature of this research, which has only answered the 'what' and 'how much' questions related to the effects of CLIL but has not explored the phenomenon of interest from the perspective of interpretivism. This, to a large degree, falls into the research gap that most CLIL studies tend to focus on a few aspects of evidence (e.g., performance evidence, affective evidence, learning process evidence, materials and task evidence) and thus presents a narrow picture of CLIL programmes (Hu, 2021a). Hence, mixed-methods designs are more appropriate to examine and understand the effects of CLIL and should be utilised in further studies. Besides, a non-probability sampling technique was used, preventing the researchers from generalising the research findings to a larger population. Meanwhile, the sample size might not be large enough to detect more important differences with probability or generate more profound findings representative of a broader population, though the '$n \geq 30$ rule of thumb' was observed (Indhiarti &

Sudarwati, 2021). Therefore, future quantitative research involving more participants is desirable to percolate further understanding of the effects of CLIL on learning.

7. Conclusion

Since the introduction of CLIL into language education, many studies have been carried out to examine its effects from different perspectives. The present study that explores some of the performance and affective evidence in the case of a CLIL programme provided in a Chinese university has unlocked the potential of this pedagogical approach, and, though small, is an attempt to enrich the CLIL academia in China. The myth that CLIL should be selective and only accessible to advanced students is dispelled, at least based on the findings that the participants of different language proficiency levels made learning progress in the same classroom. However, the view that CLIL is a panacea for all should be re-examined, as it has been indicated that academically weaker learners in this research were motivated at a lower level than expected. To respond to the question whether CLIL can kill two birds with just one stone in CET — that is to say, if CLIL can improve Chinese EFL learners' language proficiency and motivation simultaneously — a critical view should be assumed. In this study, the win-win situation was true for proficient language learners but was not the case for those who were less academically competent. However, empirical research examining the effects of CLIL is required in any case, as it can provide vigorous evidence to support CLIL as the reason, or one of the reasons, for the gains language learners make in the learning process.

Acknowledgment

The authors received support from Universiti Kebangsaan Malaysia (UKM) with research grant codes GG-2019-011 and GGPM-2019-012 and from the Research Centre of Teaching and Learning Innovation, Faculty of Education, UKM.

8. References

Blasco, A. B. (2014). *L2 English young learners' oral production skills in CLIL and EFL settings: A longitudinal study* (Publication No. 9788449050886) [Doctoral Dissertation, Universitat Autònoma de Barcelona]. Theses and Dissertations Online.

Brown, H., & Bradford, A. (2017). EMI, CLIL, & CBI: Differing approaches and goals. In P. Clements, A. Krause & H. Brown (Eds.), *Transformation in language education* (pp. 328-334). JALT. https://doi.org/10.13140/RG.2.2.19525.04325

Bruton, A. (2011). Is CLIL so beneficial, or just selective? Re-evaluating some of the research. *System, 39*(4), 523–532. https://doi.org/10.1016/j.system.2011.08.002

Bruton, A. (2013). CLIL: Some of the reasons why…and why not. *System, 41*(3), 587-597. https://doi.org/10.1016/j.system.2013.07.001

Cañado, M. L. P. (2019). CLIL and elitism: myth or reality? *Language Learning Journal, 48*(1), 1-14. https://doi.org/10.1080/09571736.2019.1645872

Celik, B. (2019). A comparison of form-focused and meaning-focused instruction types: A study on Ishik University Students in Erbil, Iraq. *International Journal of English Linguistics, 9*(1), 201-228. https://doi.org/10.5539/ijel.v9n1p201

Chen, X., Webb, J. G. (2017). The CET-4: Encouragement or impediment of undergraduate English testing and learning. *International Journal of Education and Human Developments, 3*(4), 31-42. https://ijehd.cgrd.org/images/vol3no4/5.pdf

Coyle, D. (2013). Listening to learners: An investigation into successful learning across CLIL contexts. *International Journal of Bilingual Education and Bilingualism, 16*, 244–266. https://doi.org/10.1080/13670050.2013.777384

Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and Language Integrated Learning*. Cambridge University Press.

Dalton-Puffer, C., & Smit, U. (2013). Content and language integrated learning: A research agenda. *Language Teaching, 46*(4), 545-559. https://doi.org/10.1017/S0261444813000256

Dewi, N. I., & Sudarmaji, I. (2020). The effect of Content and Language Integrated Learning on reading comprehension at eleventh grade at Sma Al-Ijtihad in academic year 2018/2019. *Foremost Journal, 1*(2), 123-133. http://ejournal.unis.ac.id/index.php/Foremost/article/view/782

Gardner, R. (1985). *The Attitude Motivation Test Battery: Technical Report 1.* University of Western Ontario.

Goris, J. (2019). *Content and Language Integrated Learning in English as a foreign language: A European perspective*. Behavioural Science Institute, Radboud University Nijmegen.

Goris, J., Denessen, E., & Verhoeven, L. (2019). Effects of Content and Language Integrated Learning in Europe: A systematic review of longitudinal experimental studies. *European Educational Research Journal, 18*(6), 675-698. https://doi.org/10.1177/1474904119872426

Hamidavi, N., Amiz, M. S., & Gorjian, B. (2016). The effect of CLIL method on teaching reading comprehension to junior high school students. *Bulletin de la Société Royale des Sciences de Liège, 85*, 1642-1652. https://doi.org/10.25518/0037-9565.6361

He, D. (2018). *Foreign language learning anxiety in China: Theories and applications in English language teaching*. Springer.

Hu, H. (2021a). Mapping out the future: A proposal of a CLIL research agenda in China. *Journal of English Language Teaching and Applied Linguistics, 3*(9), 1-12. https://doi.org/10.32996/jeltal.2021.3.9.1

Hu, H. (2021b). Dispelling the myth of elitism and establishing the evidence of inclusion: A case of CLIL in online English education amid COVID-19 pandemic. In V. Akerson & M. Shelley (Eds.), *Proceedings of IConSES 2021 – International Conference on Social and Education Sciences* (pp. 29-50). ISTES Organization.

Indhiarti, T. R., & Sudarwati, E. (2021). *An introduction to research methods in applied linguistics: A practical guide*. Tim UB Press.

Jäkel, N. (2015). *Use and impact of language learning strategies on language proficiency: Investigating the impact of individual difference variables and participation in CLIL streams* (Unpublished doctoral dissertation). Ruhr-Universität Bochum.

Kamal, A. (2020). A case study on simultaneous teaching of English literature and language in a private university in Bangladesh. In S. Sultana, M. M. Roshid, M. Z. Haider, M. M. N. Kabir & M. H. Khan (Eds.), *The Routledge handbook of English language education in Bangladesh* (chap. 14). Routledge. http://https//doi.org/10.4324/9780429356803

Kamis, M. S., Ismail, M. J., Alias, M. N., Mikeng, D., Abidin, S. G. Z., & Yusof, R. (2021). CLIL approach in encouraging self-efficacy amongst Malaysian gifted students for Arabic tasks accomplishment. *Journal of Language and Linguistic Studies, 17*(2), 1001-1012. https://doi.org/10.52462/jlls.69

Karabassova, L. (2019). CLIL or "just good teaching" in Kazakhstan? *Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas, 13*(27), 55-71. https://doi.org/10.26378/rnlael1327341

Lasagabster, D. (2020). Motivation in Content and Language Integrated Learning (CLIL) research. In M. Lamb, K. Csizér, A. Henry & S. Ryan (Eds.), *The Palgrave handbook of motivation for language learning* (pp. 347-366). Palgrave Macmillan.

Li, F., & Xiao, L. (2020). A study on the teaching of professional-oriented English writing in applied-type university based on I Write 2.0. *Creative Education, 11*, 1720-1729. https://doi.org/10.4236/ce.2020.119125

Liu, H. (2019). An action research into task-based CLIL applied to education majors: From Chinese students' perspective. *English Language Teaching, 12*(3), 94-107. https://doi.org/10.5539/elt.v12n3p94

Lyster, R. (2006). Form-focused instruction in immersion classrooms. In S. Björklund, K. Mård-Miettinen, M. Bergström & M. Södergård (Eds.), *Exploring dual-focused education: Integrating language and content for individual and societal needs* (pp. 38-54). University of Vaasa.

Maasum, T. N. R. T. M., Maarof, N., Yamat, H., & Zakaria, E. (2012). An investigation of teachers' pedagogical skills and content knowledge in a content-based instruction context. *Indonesian Journal of Applied Linguistics, 1*(2), 75-90. https://doi.org/10.17509/ijal.v1i2.86

Marsh, D. (2003). *The relevance and potential of content and language integrated learning (CLIL) for achieving MT+2 in Europe*. University of Jyväskylä.

Martyn, E. R. (2018). Integrating Content and Language in business English teaching in China: First year students' perceptions and learning experience. *English Language Teaching, 11*(8), 86-102. https://doi.org/10.5539/elt.v11n8p86

Mattheoudakis, M. (2019). Content and Language Integrated Learning in Greece: It's here to stay. In M. Mattheoudakis & I. Ziaka (Eds.), *Promoting CLIL implementation in Europe* (pp. 6-27). IKY, Erasmus +.

McDougald, J. (2015). Teachers´ attitudes, perceptions and experiences in CLIL: A look at content and language. *Colombian Applied Linguistics Journal, 17*(1), 25-41. http://dx.doi.org/10.14483/udistrital.jour.calj.2015.1.a02

Mehisto, P., Marsh, D., & Frigols, M. J. (2008). *Uncovering CLIL: Content and Language Integrated Learning and multilingual education*. Macmillan Education.

Mewald, C. (2007). A comparison of oral foreign language performance of learners in CLIL and mainstream classes at lower secondary level in Lower Austria. In C. Dalton-Puffer & U. Smit (Eds.), *Empirical perspectives on CLIL classroom discourse* (pp. 139-178). Peter Lang.

Ministry of Education of the People's Republic of China. (2007). *College English Curriculum Requirements*. Higher Education Press.

Ministry of Education of the People's Republic of China. (2020). *Guidelines for College English Teaching*. Higher Education Press.

Ostovar-Namaghi, S. A., & Nakhaee, S. (2019). The effect of CLIL on language skills and components: A meta-Analysis. *Iranian Journal of Applied Linguistics, 22*(1), 109-144. https://ijal.khu.ac.ir/article-1-3030-en.pdf

Paran, A. (2013). Content and language integrated learning: Panacea or policy borrowing myth? *Applied Linguistics Review, 4*(2), 317–342. https://doi.org/10.1515/applirev-2013-0014

Pižorn, K. (2017). Content and Language Integrated Learning (CLIL): A panacea for young English language learners? In J. Enever & E. Lindgren (Eds.), *Early language learning: Complexity and mixed methods* (pp. 145-163). Oxford University Press.

Pradeep, K. (2021). *A short handbook of qualitative research*. Bluerose Publishers.

Setiawan, J. E. (2013). *Teacher's belief and student's perception regarding Content and Language Integrated Learning: A case study at SMA YPVDP Bontang* (Unpublished doctoral dissertation). Mulawarman University.

Suliman, A., Nor, M. Y. M., & Yunus, M. M. (2018). Gleaning into students' perspectives in learning science and mathematics using the English language. *The Journal of Social Sciences Research*, (2), 445-456. https://doi.org/10.32861/jssr.spi2.445.456

Suliman, A., Nor, M. Y. M., & Yunus, M. M. (2020). Illuminating students' understanding and acceptance towards the Dual Language Programme (DLP). *Cakrawala Pendidikan, 39*(3), 522-531. https://doi.org/10.21831/cp.v39i3.30760

Qi, J. (2021). Some important relations in College English Teaching. *Sino-US English Teaching, 18*(12), 359-362. https://doi.org/10.17265//1539-8072/2021.12.003

Urgal, C. C. (2019). Law and business students' attitudes towards learning English for specific purposes within CLIL and non-CLIL contexts. *Languages, 4*(2), 45-53. https://doi.org/10.3390/languages4020045

Wang, K. (2017). *Xinghuo Yingyu: Yingyu Liuji Zhenti* [Spark English: Authentic CET-6 papers]. Shanghai Jiao Tong University Press.

Wang, W., & Xu, H. (2020). *2018 Nian Zhongguo Waiyu Jiaoyu Niandu Baogao* [2018 China Foreign Language Education Annual Report]. Foreign Language Teaching and Research Press.

Yang, W. (2015). Content and Language Integrated Learning next in Asia: Evidence of learners' achievement in CLIL education from a Taiwan tertiary degree programme. *International Journal of Bilingual Education and Bilingualism, 18*(4), 361-382. https://doi.org/10.1080/13670050.2014.904840

Young, R. (2018). What is the effect of Content Language Integrated Learning on pupil motivation to learn a foreign language? *Scottish Languages Review*, (34), 29-38. https://scilt.org.uk/Portals/24/Library/slr/issues/34/34-03 Young.pdf

Yunus, N. M., & Sukri, S. I. A. (2017). The use of English in teaching mathematics and science: The PPSMI Policy vis-à-vis The DLP. *Advances in Language and Literary Studies, 8*(1), 133-142. http://dx.doi.org/10.7575/aiac.alls.v.8n.1p.133

Zhang, B., Qiao, M., Jin, X., & Wang, L. (2019). What are the experiences of College English Teaching in Heilongjiang Province, China? In T. D. Neimann & U. M. Stelson (Eds.), *Challenges and opportunities in global approaches to education* (pp. 95-114). IGI Global.

Zydatiß, W. (2012). Linguistic thresholds in the CLIL classroom? The threshold hypothesis revisited. *International CLIL Research Journal, 1*(4), 17–28. http://www.icrj.eu/14/article2.html