Effects of Content and Language Integrated Learning Instruction Model on Learning Outcomes in Chinese STEM Program: A Quasi-Experimental Comparative Study

Yangxi Han
University of Malaya, Kuala Lumpur, Malaysia
tva190013@siswa.um.edu.my
North China University of Water Resources and Electric Power,
Zhengzhou, Henan, China, P.O.: 450046
fao@ncwu.edu.cn

Zhao Shuo
Communication University of China, Beijing, P.R.C.
University of Malaya, Kuala Lumpur, Malaysia;
zhaoshuo@nwpu.edu.cn

Anne Benedict Nair
University of Malaya, Kuala Lumpur, Malaysia;
annebene@um.edu.my

Bio-Profiles:

Yangxi Han is a Ph.D. student in the Faculty of Languages and Linguistics, University of Malaya. She is also a lecturer in the School of Foreign Languages at North China University of Water Resources and Electric Power. Her research interests include blended education, technology-assisted language learning, and STEM education.

Zhao Shuo is a professor at the School of International Studies at the Communication University of China. He is also a guest professor at the University of Malaya and Krirk University. His research interests include comparative education, international education, and STEM education.

Anne Benedict Nair is a senior lecturer at the University of Malaya. Her research interests include language education and comparative education.

Abstract

English language acquisition is a major challenge for Chinese students in the fields of science, technology, engineering, and mathematics (STEM) cooperative educational programs. The general academic English courses in China serve as preliminary courses for future advanced Content and Language Integrated Learning with science and engineering in the programs. The present general academic English courses are often criticized for form-focused instruction and language skills training rather than focusing on developing cognitive skills and social awareness. Additionally, present English courses are faced with an excessive number of students and limited time and resources, thereby, constraining effective course implementation. This comparative research study investigated the effects of different instructional models (face-to-face, blended learning [BL], and flipped learning) in three groups of 100 Chinese undergraduate students each majoring in STEM. A quasi-experimental design and student engagement questionnaires were used for data collection. The Kruskal-Wallis test and an analysis of variance revealed a significant difference between experimental group 2 and the other two groups, while no significant differences were found between the first two groups. The findings indicate that flipped BL, along with modified teaching strategies and activities improved the participants' learning results and engagement levels.

Keywords: blended learning; CLIL framework; flipped learning; learning outcome; social knowledge gains; student engagement

1. Introduction

In Chinese cooperative educational programs, the English language barrier is a primary obstacle for non-native Chinese university students in understanding science subjects in English (Tagnin & Ní Ríordáin, 2021). The language barrier hinders students from understanding the course content and consequently impacts their learning quality. The general English academic courses thereby function as preparatory courses for future advanced subject study with science and engineering content in the programs. Therefore, examining the effectiveness of these general English academic courses is important since these courses lay the foundation for students' subsequent academic and career success.

Traditional general academic English courses in STEM disciplines in Chinese cooperative educational programs are often criticized by previous literature as following an overly form-focused instruction and language skills training rather than delivering language knowledge with meanings and focusing on developing students' critical thinking skills and social awareness.

Consequently, many students lack logical thinking and critical thinking skills in English and display a limited ability to successfully express themselves or comprehend the academic language of science.

Another drawback of current general academic English courses lies in the lack of development of sustainable awareness toward the community and society (He, 2020; Kunjie, 2021). A lack of sustainability awareness can have negative impacts on society, possibly leading to students' disregard for issues such as environmental degradation. Students in STEM disciplines are likely to pursue careers in industry, manufacturing, and construction after graduation. These fields are intimately related to economic, social, and environmentally sustainable development (Malik et al., 2019). Since the degree of sustainable awareness of students is likely to influence their future sustainable decisions, the development of their sustainable awareness is important at the undergraduate level. As Wen and Sun (2015) stated, the English course in STEM programs should focus on language education and developing critical thinking skills, awareness of social knowledge, and abilities in cross-cultural communication. The introduction of content on sustainability can help students develop a stronger understanding of sustainable development. Through this process, the courses can contribute to developing students' academic English language skills and fostering their awareness of sustainable development through both implicit and explicit integration of related content into the course design.

The ongoing curriculum reform aims to alleviate the adverse effects of form-focused language learning by introducing technology-mediated Content and Language Integrated Learning (CLIL) into the course design and its implementation (Li & Zhang, 2020). The integration of developing sustainable awareness is taken as one of the course considerations. With various technological advances, English language education, in general, has transformed from face-to-face classroom learning to various forms of blended learning (BL) that support different instructional models (Zhang & Zou, 2020).

Currently, specialized courses on sustainable development and courses on English for Academic Purposes (EAP) specifically related to ecological and environmental protection are available in Chinese universities. However, these courses enroll only a limited number of students. The general academic English courses are mandatory for all undergraduate STEM students in China, and therefore, encompass a much wider student population. However, owing to many unfavorable factors, successfully implementing the curriculum goal of imparting social knowledge in the general academic English course design is easier said than done. The general academic English courses in STEM programs are characterized by a limited duration

of classroom instruction, overcrowded classroom size, and lack of teaching resources. The ways of effectively conducting both language and content learning, therefore, merit further investigation. Therefore, the current research compares the effectiveness of face-to-face, BL, and flipped learning (FL) to evaluate the technology-mediated CLIL in improving students' academic English proficiency and raising their sustainable awareness.

Previous literature has overall indicated the benefits of adopting the BL and FL models as they provide additional learning opportunities through authentic English learning materials, online academic discussions in English, and facilitating self-regulated learning (Sahin-Kizil, 2014). Although a widespread discussion in existing literature has focused on BL and FL to improve language performance, most studies mainly regard English courses in terms of developing specific language skills (Adas & Bakir, 2013; Banditvilai, 2016; Sahin-Kizil, 2014). As the course goals expand from language education to a combination of language and social education, the investigation of the effectiveness of BL models in the knowledge transfer of English as a whole should be reconsidered. CLIL is an approach where learning a language is through the learning of a subject (Yang & Yang, 2022). CLIL shares a dual goal of improving language proficiency and imparting subject knowledge. Previous research has indicated the positive effects of CLIL on improving students' linguistic competence and motivating students to use the target language to communicate with their peers in face-to-face scenarios. However, there is limited empirical evidence to evaluate the effect of technology-mediated CLIL on students' learning performance and engagement. Further investigations are required to identify how CLIL should be integrated with technology-mediated environments to achieve the dual goal of improving English language proficiency and developing sustainable awareness. Therefore, this research evaluates the effects of face-to-face CLIL, blended CLIL, and flipped CLIL on writing performance and student engagement. The successful implementation of a comprehensive course would enable sustainable knowledge transfer in the future workplaces of students, which can better benefit companies, universities, and society in the long run.

2. Literature Review

2.1. Blended Learning, Flipped Learning, and Learning Outcomes

As defined by Garrison and Vaughan, BL is the organic integration of face-to-face and online learning. BL transforms and improves the personalized study experience for students (Picciano, Dziuban, & Graham, 2013). FL emphasizes the flipping of instructional designs and places more importance on online content learning rather than lectures and face-to-face learning, which focus on discussions and task-based activities (Zhang & Zou, 2020). According

to Bergmann and Sams (2012), FL is defined as a language pedagogy in which students attend lectures outside the classroom on their own time through online learning management systems. Additionally, students perform tasks and exchange ideas in the classroom, concentrating on the subjects' aspects that they find challenging.

Overall, BL in EFL classrooms has been reported to improve students' language proficiency and communicative skills (Adas & Bakir, 2013; Banditvilai, 2016; Sahin-Kizil, 2014) and yield a higher level of student motivation and supportive attitudes (Azamat Akbarov & Aydogan, 2018; Bakeer, 2018; Birbal, Ramdass, & Harripaul, 2018). For instance, Banditvilai (2016) validated the use of the BL model in English language learning by finding that BL supports greater opportunities for students to practice their English language skills—through listening, reading, writing, and speaking—with greater autonomy. By assessing the learning performances and attitudes of 60 EAP students in Thailand, Banditvilai revealed a positive correlation between technology and higher achievements in the participants' language skills (Banditvilai, 2016). Similarly, Pinto-Llorente, Sánchez-Gómez, García-Peñalvo, and Casillas-Martín (2017) assessed 358 students' perceptions of the efficacy of asynchronous tools (podcast, videocast, online tests, online glossary, and forums) as forms of blended designs for learning English grammar in a Spanish university. The findings show that BL benefits grammar learning through a higher level of autonomy and increased opportunities to practice English grammar in real-life situations. Therefore, most previous related studies have emphasized the role of BL in improving language proficiency. However, the role of BL in developing cognitive skills and social awareness has seldom been addressed despite its significance in developing students' comprehensive language abilities and positively influencing the community and society.

Previous research on evaluating the learning outcomes between face-to-face learning, BL, FL, and online learning has been inconclusive (Al-Qahtani & Higgins, 2013; Ceylan & Kesici, 2017; 15. Clark, Kaw, & Besterfield-Sacre, 2016; Khader, 2016; Talan & Gulsecen, 2019; Yapici, 2016). For instance, Clark, Kaw, and Besterfield-Sacre (2016) compared students' achievement scores in BL versus semi-flipped versus flipped classrooms with a total of 132 students who responded to the research. The findings revealed that the semi-flipped courses helped students achieve higher scores than other forms of instruction. However, the study indicated that BL resulted in the most desirable classroom environment compared to semi-flipped and flipped classrooms. In contrast, Al-Qahtani and Higgins (2013) compared 148 student achievement test results in face-to-face, BL, and e-learning instruction methods in Saudi Arabia. The findings show no significant difference in the test results between face-to-

face and e-learning, only a significant difference in achievement test results between the face-to-face and BL groups (Al-Qahtani & Higgins, 2013). Different contexts can be a possible factor in the contradicting results, and it is reasonable to analyze each technology-mediated model in its unique context and treat their integration with education individually as a socially constructed process (Arrosagaray, González-Peiteado, Pino-Juste, & Rodríguez-López, 2019).

Counter-studies argue that pedagogy appears to be a deciding factor in improved learning outcomes, regardless of how a course is delivered (Zhang & Zou, 2020). Nevertheless, as technology continues to mature, BL has revolutionized its forms by integrating state-of-the-art technologies such as mobile learning, digital-based game learning, augmented reality (AR) technology, and course management systems (Zhang & Zou, 2020). BL models have thereby gradually transformed teaching methodologies in various ways, complementing and even improving traditional teaching and learning paradigms (Wang, Chen, Tai, & Zhang, 2019). In terms of English teaching and learning, BL supports a shift in language pedagogy from teacher-centered to student-centered (Mahmud, 2018).

Several previous studies have found a positive correlation between technology-mediated FL and enhanced language skills by allowing students to practice the language in real-life environments while still receiving immediate input from the instructor (Fisher, Perényi, & Birdthistle, 2018; Murillo-Zamorano, Sánchez, & Godoy-Caballero, 2019). Despite the reported positive results, some studies argue that technology-mediated FL is highly demanding for students with lower language proficiency who lack autonomy (Lombardini, Lakkala, & Muukkonen, 2018). Additionally, an extra burden is placed on teachers to prepare countless PowerPoint files, videos, and online teaching materials, and update themselves about various technological innovations. Particularly, technology-mediated FL may cause challenges to teachers in developing regions, such as sub-Saharan Africa (Tachie, 2019). Because of limited access to computers and the lack of educational facilities, the implementation of technology-mediated FL demands preparatory educational facility investment and related professional support. Therefore, technology-mediated FL is not necessarily applicable to all contexts, learners, and courses, and should be analyzed on a case-by-case basis.

2.2. Blended Learning, Flipped Learning, and Student Engagement

Previous studies have supported the claim that both BL and FL can positively enhance student engagement (Baranova, Khalyapina, Kobicheva, & Tokareva, 2019; Henrie, Halverson, & Graham, 2015; Pilotti, Hardy, Murphy, & Vincent, 2017). However, there is a lack of consensus on the definition of student engagement, which leads to difficulties in measuring a

learner's degree of engagement. According to Fredricks (2012), student engagement may be categorized into three parts: behavioral, cognitive, and emotional. Nevertheless, as Halverson and Graham (2019) argued, behavioral and cognitive engagement include overlapping concepts. Thus, they proposed classifying student engagement into cognitive and emotional engagement. Behavioral engagement is considered a manifestation of engaging psychological processes. In the context of technology-mediated language learning, the pedagogy design is a significant challenge for the acceptance of any digital innovation (Arrosagaray et al., 2019). The theoretical framework justifies the use of questionnaires to measure student engagement. For instance, Wang et al., (2019) developed a questionnaire to examine 1603 students' perceptions of Small Private Online Course (SPOC) platforms that supported FL in the Chinese EFL context. Based on the questionnaire results, the findings revealed a positive correlation between SPOC-supported FL and student engagement by encouraging active practice of the English language.

In light of previous findings, this study aims to compare the effectiveness of different BL designs to support the transfer of knowledge of sustainability in general academic courses in STEM cooperative educational programs. The findings of this research highlight the importance of integrating pedagogical changes with the BL design to maximize technology-mediated EFL learning wherein both language learning and social knowledge transfer are emphasized.

3. Theoretical Framework

3.1. Content Language Integrated Learning and Technology

CLIL is a course framework that guides language learning where a foreign language is regarded as the mode of communication and as a learning subject. CLIL is distinctive from traditional EFL learning because of its dual focus on language and content (Coyle & de Larios, 2020). Most previous research has focused on applying CLIL in EAP or international programs in Chinese higher education (Chen, 2013; Shen, 2019). The adoption of CLIL in general academic English courses is relatively scarce. Some scholars argue that CLIL can place an additional burden on students who lack the preliminary language skills to understand the content and that the adoption of CLIL in general courses is another form of EAP course in essence (Bruton, 2011). However, CLIL does boost the scaffolding as a means and modifies course delivery forms to promote language learning and the development of social and cultural awareness rather than shifting the nature of the course (Li & Zhang, 2020). While some argue that CLIL and EAP are blurred in their concepts, EAP is intended for students with specific

English language needs regarding academic or employment opportunities, such as studying abroad, accessing texts written in English, or conducting business negotiations in international settings (de Zarobe & Lasagabaster, 2010) by strengthening the content of that specific field. While CLIL is an approach in EFL, no specific field of knowledge is prioritized. Instead, the learning of the language itself and various categories of the course content occur simultaneously. In CLIL, it is essential to balance the emphasis on language and content (Bruton, 2013). Previous empirical evidence has reported that the possible benefits of adopting CLIL include enhanced language competence, social and cultural awareness, and increased motivation (Coyle & de Larios, 2020; Mayo & de los Angeles Hidalgo, 2017). Similarly, positive outcomes of technology-facilitated CLIL have primarily been reported in language performance and content understanding (O'Dowd, 2018). However, as technology matures, technology-facilitated CLIL, which could have simply implied PowerPoint presentations in the past, has gradually become obsolete in language learning classrooms, where various types of technologies combine into one BL model (Bozdoğan, 2015). The gap in current technology-facilitated CLIL research is due to the lack of evaluation of the different BL designs under the CLIL framework, and the current study aims to bridge this gap in existing research.

3.2. Research Questions

RQ1: To what extent do different BL models affect participants' learning outcomes in the general academic English course under the CLIL framework in STEM programs?

RQ2: To what extent do different BL models influence student engagement in the general academic English course under the CLIL framework in STEM programs?

4. Methodology

4.1. Research Design

Figure 1 illustrates the design of the course procedure. The course content was designed based on the principle of the 4Cs: content, communication, cognition, and culture (Coyle, 2007). To address the aims of developing sustainability awareness, two units in the course explicitly covered topics on environmental and social sustainability in their primary readings. Additionally, links to the latest reports and videos on sustainability were shared in the BL group, while the control group mainly relied on text materials. Furthermore, debates on related sustainability issues were conducted online or in classrooms for students each week. Essays on topics relating to sustainability were assigned twice in the term, and activities such as

brainstorming, designing a "green fair," and debates over sustainable developmental issues were held every two weeks. The same textbook and self-designed materials were adopted, and the same five units of content were delivered in the autumn term from 2020 to 2021. Figures 2 and 3 illustrate how the two BL designs were used in this research among the three groups: a control group, which solely received face-to-face teaching without technology mediation in the learning process. As demonstrated in Figure 2, the technological tools in experimental group 1 enabled students to accomplish the pre-class tasks, exposed the students to extra videos and materials, facilitated an online board discussion, and set out exercises for students to practice online. The most distinctive difference between experimental groups 1 and 2 was the adoption of FL. Unlike the control group and experimental group 1, the main content of the course syllabus was placed online in the students' pre-class learning process for experimental group 2, which left more time for classroom activities and discussions during the classes. As shown in Figures 2 and 3, both the experimental groups adopted the same learning management system, mobile terminals, iCourses (www.icourses163.org), instant message tools, and WeChat to support students' autonomous learning.

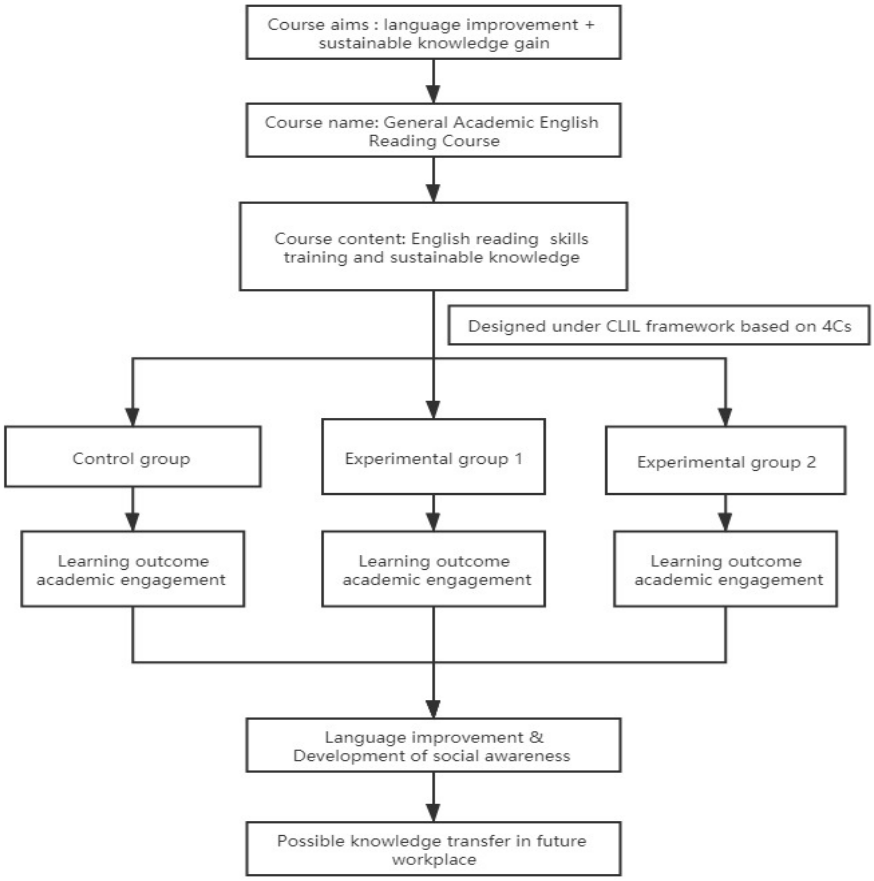

Figure 1: Course Procedure

CLIL: Content and Language Integrated Learning; 4Cs: content, communication, cognition, and culture

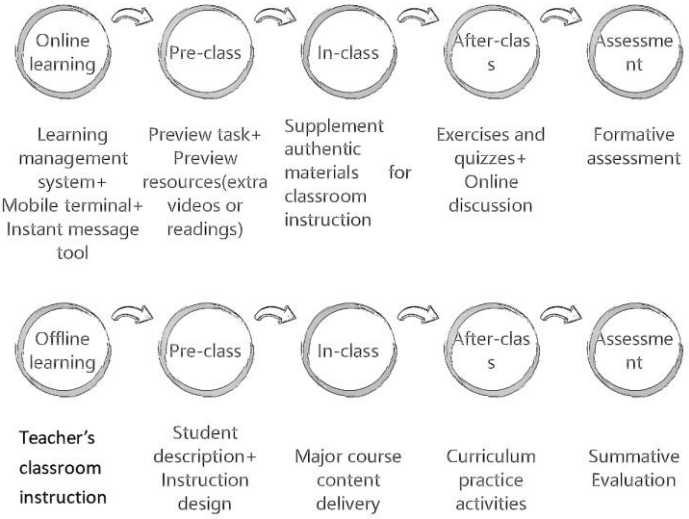

Figure 2: Blended Learning Model without Flipping: Experimental Group 1

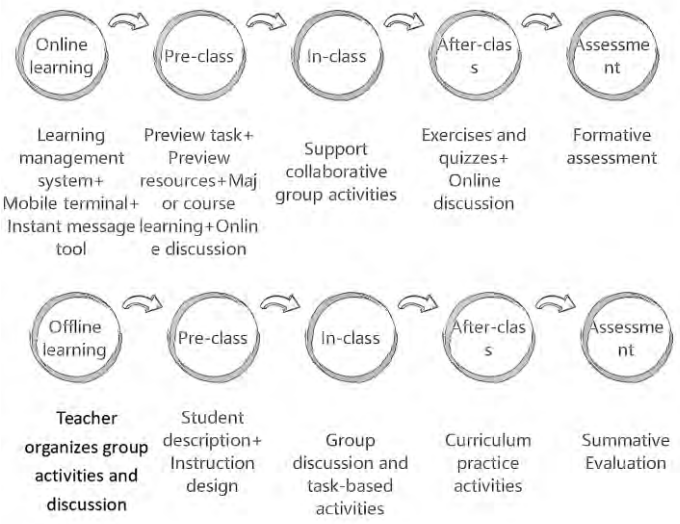

Figure 3: Blended Learning Model with Flipping: Experimental Group 2

The research was undertaken at three publicly funded universities ranked at the same level in Henan Province, China. The universities were neither "double first-class" universities, which refers to the top universities as classified by the Ministry of Education in China, nor were they professional colleges aimed at technical training. Three instructors who shared similar educational backgrounds and professional training experiences were chosen for this course. Their educational backgrounds are presented in Table 1.

Table 1: Instructors' Educational Backgrounds

	Instructor A	Instructor B	Instructor C
Degree	Masters	Masters	Masters
Academic title	Lecturer	Lecturer	Lecturer
Teaching experience	14 years	18 years	15 years

The three university lecturers had similar levels of professional training and had received previous training in the use of all online supporting systems and attended the summer seminars held by the publisher of the course textbooks. Additionally, the researcher formed a WeChat contact group to allow the three instructors to share their course materials and exchange their teaching experiences freely in the group.

4.2. Participants

Purposive sampling was adopted to ensure that all participants met the following conditions: 1) All participants in the experimental groups were familiar with operating the technological tools involved in this research, and 2) All the participants were pursuing STEM disciplines in the cooperative educational programs, such as information technology, engineering management, and water conservancy engineering. The participants were mainly from the Faculty of Management Engineering, the Faculty of Water Conservancy, the Faculty of Electrical and Information Engineering, and the Faculty of Material and Chemical Engineering in the three universities. A total of 300 sophomore Chinese students (100 per group), comprising 61.9 % men and 38.1 % women, with an average age of 19 years, participated in the study. In general, all the participants had 12 years of experience in learning English as a subject since their primary education. According to their English College Entrance Examination scores, all three groups shared similar levels of English proficiency, with a mean score of 107.2.

4.3. Instruments

RQ1: To what extent do different blended learning models affect participants' learning outcomes in the general academic English course under the CLIL framework in STEM programs?

A quasi-experimental design was used in this study. A pre-test and post-test were designed to directly compare the test results among the three groups. As the course aims included

improving language skills and fostering social and cultural awareness, the tests were designed in line with the teaching goals. The tests mainly included topics on social, economic, and environmental sustainability development. The test forms included multiple choices, reading comprehension, essay writing, and activity design. The scoring metrics for the essay writing were based on those of the CET-4 (College English Test Band-4). The draft tests were initiated and presented to the related instructors and experts in the EFL fields for review. With the consulting advice, additional test items were added, while some of the original test items were modified or deleted according to their content validity. The pilot test was conducted on 47 students outside the sample population. The test reliability was verified and Cronbach's alpha was 0.845.

RQ2: To what extent do different blended learning models influence student engagement in the general academic English course under the CLIL framework in STEM programs?

Regarding RQ2, an online student questionnaire was revised from the theoretical model of the National Survey of Student Engagement (NSSE)-China. The NSSE-China is an assessment tool for measuring student study quality in Chinese higher education using student engagement scales. The NSSE was developed by the American Higher Education Management System and translated into Chinese and adapted into the Chinese academic context to form the NSSE-China. The NSSE-China comprises the following five benchmarks: level of academic challenge, active and collaborative learning, student and faculty interaction, enriching educational experience, and supportive campus environment (Ross, Cen, & Shi, 2014), thus making it easier to fit it into the local academic context. The questionnaire used in this study adopted the same five-point Likert-type scale, ranging from 1 = never to 5 = all the time or 1 = totally disagree to 5 = totally agree. The questionnaire was designed in Mandarin to avoid any possible misunderstanding of the questionnaire items by students. Refer to Appendix A for the translated version of the questionnaire. The revised version of the questionnaire was modified further in response to the advice of related instructors and experts regarding the validity of the questionnaire. Concerning the questionnaire's reliability, a pilot test was conducted on 45 students and Cronbach's alpha was calculated using IBM SPSS 26. Cronbach's alpha was 0.951, which was considered appropriate for the study.

4.4. Data Collection and Analysis

Before conducting the research, all related ethical issues were addressed. First, formal institutional consent was obtained before the survey was conducted through a written consent letter. Second, all students were provided with an information letter with all the salient research information. They were required to sign the consent letter online before continuing with the tests and questionnaires. Third, all participants maintained anonymity during the entirety of the research process. The pre-and post-test and the questionnaire were delivered online by scanning QR codes. The pre-test was delivered at the beginning of the academic term before the

intervention, while the post-test was provided at the end of the term. The purpose of the pre-test was to assess whether students' English performance before the intervention reached a statistical significance ($p < 0.05$). Both the pre-test and post-test were 45 minutes long. Anti-cheating was used for the tests. The questionnaire was administered to all three groups at the end of the term and all participants responded to the tests and questionnaires. The first researcher and a senior teacher, who was experienced in technology-assisted language teaching, rated all the tests. The second researcher reviewed all the test results.

For data analysis, the statistical program SPSS 26.0 was applied. As Levene's Test proved to be violated, the non-parametric Kruskal-Wallis test was applied to determine any statistical difference among the three groups. A confidence level of 0.05 was adopted. Then, a pairwise comparison analysis was conducted to compare the statistical differences between the two models. The Bonferroni correction was adopted to avoid the possible "inflated Type 1 error." This required the significance level to be decreased to 0.0167. The effect size was determined by calculating the partial eta squared.

The questionnaires were delivered after the intervention online to the three groups by scanning QR codes. To analyze the questionnaire responses, a one-way between-groups analysis of variance (ANOVA) and Tukey's post-hoc tests were used to determine whether the results reached a significance level of $p < 0.05$. Descriptive data, homogeneity of variance test, and Tukey's multiple comparison results were obtained. The significance value was checked for mean differences and the effect size of the one-way ANOVA was calculated.

5. Findings

5.1. Test Performance

Tables 2 and 3 present the descriptive and inferential statistics of the pre-test included in this study. Regarding test performance before the intervention, students in all three groups achieved similar results. The mean pre-test scores of the control group (M = 73.3, SD = 9.331), experimental group 1 (M = 73.3, SD = 10.542), and experimental group 2 (M = 73.8, SD = 9.446) did not differ in terms of students' performance before the intervention. ANOVA was performed to identify whether the pre-test scores showed a statistically significant difference among the three groups, and a result above 0.5 (0.917) indicates no significant difference in students' pre-test performance, with a small effect size (η^2 = 0.001).

Table 2: Pre-test Descriptive Statistics

University	*N*	Mean[1]	*SD*	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
1.00	100	73.30	9.331	.933	71.45	75.15	52	95
2.00	100	73.30	10.542	1.054	71.21	75.39	49	91
3.00	100	73.80	9.446	.945	71.93	75.67	46	92
Total	300	73.47	9.758	.563	72.36	74.58	46	95

[1] 95% confidence interval for mean.

Table 3: Pre-test Analysis of Variance

	Sum of Squares	*df*	Mean Square	*F*	Sig.
Between Groups	16.667	2	8.333	.087	**.917**
Within Groups	28456.000	297	95.811		
Total	28472.667	299			

In terms of post-test performance, students in experimental group 2 (*M* = 80.69, *SD* = 15.425) outperformed the other two groups, as shown in the descriptive statistics in Table 4. According to the descriptive statistics, students in experimental group 1 (M = 74.89, *SD* = 14.338) slightly outperformed those in the control group (*M* = 73.79, *SD* = 9.727). As Levene's Test was violated ($p = 0.00$), the Kruskal-Wallis test was used to compare the group differences in post-test scores (Table 5; Figure 4). The Kruskal-Wallis test rejected the null hypothesis and revealed a significant difference among the three groups (χ^2 (2, n = 300) = 20.67, $p = 0.00$). However, the pairwise comparison indicated that the control and experimental groups showed no statistically significant differences (Table 6). After applying the Bonferroni correction with different comparisons, the corrected alpha level was 0.0167 (0.05/3). There were significant differences between the control group and experimental group 2 ($p = 0.000$) and also between the two experimental groups ($p = 0.001$).

Table 4: Post-test Descriptive Statistics

University	*N*	Mean	*SD*	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
1.00	100	73.79	9.727	.973	71.86	75.72	53	95

2.00	100	74.89	14.338	1.434	72.05	77.73	49	95
3.00	100	80.69	15.425	1.543	77.63	83.75	39	95
Total	300	76.46	13.688	.790	74.90	78.01	39	95

[1] 95% confidence interval for mean.

Table 5: Post-test Kruskal-Wallis Test

Null Hypothesis	Test	Sig.	Decision[1]
[1] Distribution of post-test is the same across categories of the university.	Independent-Samples Kruskal-Wallis Test	.000	Null hypothesis rejected.

[1] Asymptotic significance is displayed. Significance level was .050.

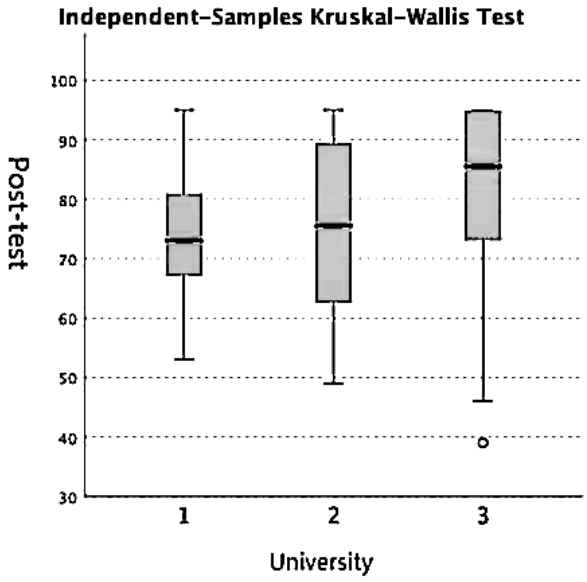

Figure 4: Independent-samples Kruskal-Wallis Test

Table 6: Post-test Pairwise Comparison

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.a[1]
1.00-2.00	−11.575	12.247	−.945	.345	1.000
1.00-3.00	−52.955	12.247	−4.324	.000	.000
2.00-3.00	−41.380	12.247	−3.379	.001	.002

[1] Each row tested the null hypothesis that Sample 1 and Sample 2 distributions were the same.

5.2. Student Engagement

To determine whether there was a significant difference in the students' engagement levels after the intervention, an ANOVA was performed (Table 9). Tables 7 and 8 indicate that equal variances can be assumed as Levene's Test result was not significant ($p = 0.065$). There was a statistically significant difference at the $p < .05$ level in engagement levels for the three groups: F (2, 300) = 8.2, $p = 0.00$. The effect size was calculated using eta squared ($\eta^2 = 0.05$). Post-hoc comparisons using the Tukey HSD test indicated that the mean score for experimental group 2 ($M = 3.33$, $SD = 0.71$) was significantly different from that of experimental group 1 ($M = 3.07$, $SD = 0.59$) and the control group ($M = 2.99$, $SD = 0.59$). Experimental group 1 and the control group did not differ significantly from each other (Table 10).

Table 7: Descriptive Statistics for Student Engagement

University	*N*	Mean	*SD*	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
1	100	2.9854	.59087	.05909	2.8682	3.1027	1.51	4.99
2	100	3.0732	.58758	.05876	2.9566	3.1898	1.73	5.00
3	100	3.3335	.70965	.07097	3.1927	3.4744	2.00	5.00
Total	300	3.1307	.64698	.03735	3.0572	3.2042	1.51	5.00

[1] 95% confidence interval for mean.

Table 8: Test of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
Engagement	Based on Mean	2.751	2	297	.065
	Based on Median	2.752	2	297	.065
	Based on Median and with adjusted df	2.752	2	292.931	.065
	Based on Trimmed Mean	2.771	2	297	.064

Table 9: Analysis of Variance for Student Engagement

	Sum of Squares	*df*	Mean Square	*F*	Sig.
Between Groups	6.556	2	3.278	8.208	.000
Within Groups	118.600	297	.399		

Total	125.156	299

Table 10: Multiple Comparisons for Student Engagement

(I) university	(J) university	Mean Difference (I−J)	Std. Error	Sig.	Lower Bound	Upper Bound
1	2	−.08772	.08937	.589	−.2982	.1228
	3	−.34810*	.08937	.000	−.5586	−.1376
2	1	.08772	.08937	.589	−.1228	.2982
	3	−.26038*	.08937	.011	−.4709	−.0499
3	1	.34810*	.08937	.000	.1376	.5586
	2	.26038*	.08937	.011	.0499	.4709

[1] 95% confidence interval for mean.

6. Discussion

RQ1: To what extent do different blended learning models affect participants' learning outcomes in the general academic English course under the CLIL framework in STEM programs?

The present study evaluated whether different BL models would promote students' learning outcomes. This study is in line with the study by Wang et al. (2019), which found that a SPOC that supported BL and FL enhances the learning outcome more than face-to-face learning. This study found that technology-mediated models outperformed non-technology modes when comparing the post-test means. Because of the intrinsic nature of CLIL, which is closely related to task-based learning, it requires students' participation in tasks and activities that convey authentic language for communicative needs (Bozdoğan, 2015). BL and FL, for instance, allow more flexibility and accessibility to authentic course-related resources, promote online communication, and provide traceable student learning data (Wang et al., 2019), which might partially explain the enhanced learning outcomes. The flexibility of online forums and activities supplements the deficit of limited time and resources in face-to-face classrooms. Similarly, the technological platforms and mobile apps used in this study enabled a much wider exposure to authentic videos and readings related to sustainable development. For instance, the video titled "Recycling" was assigned as one of the preview tasks in Week 4 that explicitly explained the meaning of recycling to the experimental group participants. While learning sustainable

development with technological mediation, the participants increased their vocabulary and simultaneously practiced English reading and writing skills.

However, no statistically significant difference was reported between the control group and experimental group 1, which blended the same technological tools as experimental group 2 but without FL. This finding indicates that technology alone does not sufficiently enhance the course learning outcome with a significant statistical difference when both language and content are emphasized. This finding is in line with Fisher et al.'s (2018) study stating that technological advancement innovates the presentation of knowledge for learning but does not necessarily ensure students' acquisition of knowledge. Similarly, Golonka, Bowles, Frank, Richardson, and Freynik (2014) stated that BL impacts the process of learning but does not necessarily influence learning outcomes. When using technology to improve literacy, pedagogical aims should take precedence over technical aspects. This finding contradicts the assertions that BL surpasses traditional face-to-face learning under any circumstances in general academic English learning courses in the Chinese higher education context (Hou, 2010). This finding suggests that BL should be fused with pedagogical changes and proper curriculum designs under the CLIL framework. Simply combining technology with the teacher-dominant instruction model did not significantly improve learning outcomes. The difference between experimental groups 1 and 2 is that experimental group 1 pedagogy did not allow much time for tasks and activities in the classroom. By designing practical tasks and activities that convey authentic language for communicative needs, flipped BL under CLIL affords students increased language practice and cognitive development opportunities. A previous theoretical study suggesting a strong relationship between in-class practice and learning outcomes under the CLIL framework supports this statement (Coyle & de Larios, 2020). According to Fisher et al. (2018), changes in course designs and model designs may contribute to possible changes in research findings. A change in course needs subsequently requires a shift in pedagogical strategies in course delivery. Therefore, the current research suggests that instructional models that are more closely integrated with class practice are more likely to consolidate CLIL and BL benefits.

Additionally, the significant differences between the control and experimental group 2 and between the two experimental groups support previous literature, which stated that the positive influence of flipped-mode BL on the grasping of knowledge and skills occurred by promoting active learning (Murillo-Zamorano et al., 2019). According to Murillo-Zamorano et al. (2019), flipped-mode BL is more than a rearrangement of activities. Their research findings indicated that their FL model directly affected motivating interactions between students and instructors,

and vice versa. Moreover, it creates more opportunities for students to practice in and outside the classroom, either online or face-to-face. Likewise, this study supports the claim that the flipped BL model benefits students more than traditional teacher-dominant classroom settings by promoting active learning.

RQ2: To what extent do different blended learning models influence student engagement in the general academic English course under the CLIL framework in STEM programs?

This research question aimed to assess whether different BL models improve participants' engagement levels. Results of the ANOVA show that the flipped BL group engaged the participants more than the other two forms. According to the students' responses, the flipped BL model promotes student engagement in the following aspects: increased practice opportunities for English listening, reading, writing, and speaking skills; improvement of critical thinking; promotion of collaborative peer learning; and increased student-student and student-lecturer interactions. This finding is consistent with Serrano Dea-Ayuela, Gonzalez-Burgos, Serrano-Gil, and Lalatsa's (2019) finding that FL facilitates student engagement by promoting active learning in higher education. The greater effectiveness in experimental group 2 can be attributed to the increased time and frequency for both technology-integrated and face-to-face practice activities offered by FL. The diverse forms of activities attracted the participants' attention and motivated them to engage in learning. Students reinforce their knowledge by applying them to practical use and practicing the target language for communicative purposes while fulfilling related tasks and activities.

No significant difference was found between the experimental group 1 and the control group, which shows that the traditional face-to-face teaching model and the BL model without pedagogical changes engaged the participants at a similar level. Morton et al. (2016) addressed the importance of a careful design of BL to optimize student engagement. Likewise, the findings of this research support the finding that BL under the CLIL framework is not simply an add-on technology for face-to-face learning. Instead, it should be combined with pedagogical changes to optimize students' engagement levels.

7. Conclusions

The present study evaluates the effects of face-to-face and BL under the CLIL framework on students' learning outcomes and engagement levels when multiple learning aims of language improvement and social knowledge are considered in STEM cooperative educational programs. The novelty of the study lies in that the practical use of English learning has been combined with learning social awareness. The study aimed to identify an effective way to

enable students to engage meaningfully in their own communities and communities abroad where their English skills would be needed. The findings reveal that technology-assisted models outperform non-technology-mediated models. A flipped BL model with modified teaching strategies and activities significantly improved the participants' learning outcomes and engagement levels. Simply adding technology to teacher-dominant classrooms seems insufficient to promote learning outcomes and engagement under the CLIL framework. Although some studies argue that flipped BL can be challenging for teachers by requiring extra effort to prepare online materials (Arrosagaray et al., 2019) and it requires preparatory facility investment in developing regions (Tachie, 2019), the current research findings indicate that it benefits students significantly. The findings provide pedagogical implications for teachers and other related stakeholders in that effective technology-enhanced CLIL must be ensured via adequate pedagogical support. It is necessary to adjust pedagogical strategies and practices to optimize the technology appropriate for EFL courses serving different purposes. Additionally, course instructors need to be efficient and well-prepared under the technology-enhanced CLIL framework since pedagogical changes and effective subject teaching and technological use demand additional related training. It is necessary to support teachers with corresponding professional development programs, counseling, and assistance services. Although this research was based in China, the structure of the study can be useful for emerging economies, which have large populations that require education. The limitations in this research are firstly, the research duration was one semester, which may not be enough to identify the longitudinal effect of the technology-mediated CLIL on students' learning outcome. Secondly, students may also used the technological resources to study the learning content in their free time during the experiment. Since their study time after classes are not fixed, this may possibly affect the experimental effects. Further research can contribute to maximizing the effects of BL models for non-native STEM students in various educational contexts to meet different teaching and learning needs.

References

Adas, D. B., & Bakir, A. (2013). Writing difficulties and new solutions: Blended learning as an approach to improve writing abilities. *International Journal of Humanities and Social Science*, *3*(9), 254–266.

Al-Qahtani, A. A., & Higgins, S. E. (2013). Effects of traditional, blended and e-learning on students' achievement in higher education. *Journal of Computer Assisted Learning*, *29*(3), 220–234. https://doi.org/10.1111/j.1365-2729.2012.00490.x

Arrosagaray, M., González-Peiteado, M., Pino-Juste, M., & Rodríguez-López, B. (2019). A comparative study of Spanish adult students' attitudes to ICT in classroom, blended and distance language learning modes. *Computers & Education*, 134, 31–40. https://doi.org/10.1016/j.compedu.2019.01.016

Azamat Akbarov, A. G. K., & Aydogan, H. (2018). Students' attitudes towards blended learning in EFL context. *Acta Didactica Napocensia*, *11*(1), 61–68. https://doi.org/10.24193/ADN.11.1.5

Bakeer, A. M. (2018). Students' attitudes towards implementing blended learning in teaching English in higher education institutions: A case of Al-Quds Open University. *International Journal of Humanities and Social Science*, *8*(6), 131–139. https://doi.org/10.30845/ijhss.v8n6a15

Banditvilai, C. (2016). Enhancing students' language skills through blended learning *The Electronic Journal of e-Learning*, *14*(3), 220–229.

Baranova, T., Khalyapina, L., Kobicheva, A., & Tokareva, E. (2019). Evaluation of students' engagement in integrated learning model in a blended environment. *Education Sciences*, *9*(138), 1–13. https://doi.org/10.3390/educsci9020138

Bergmann, J., & Sams, A. (2012). *Flip your classroom: Reach every student in every class every day*. Washington DC: International Society for Technology in Education.

Birbal, R., Ramdass, M., & Harripaul, C. (2018). Student teachers' attitudes towards blended learning. *Journal of Education and Human Development*, *7*(2), 9–26. https://doi.org/10.15640/jehd.v7n2a2

Bozdoğan, D. (2015). 13 Technology-enhanced CLIL classrooms. In S. Pokrivčáková (Ed.). *CLIL in foreign language education* (pp. 164–176). Nitra: Constantine the Philosopher University.

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

Ceylan, V. K., & Kesici, A. E. (2017). Effect of blended learning to academic achievement. *Journal of Human Sciences*, *14*(1), 308–320. https://doi.org/10.14687/JHS.V14I1.4141

Chen, Y. (2013). The feasibility analysis of CLIL teaching method in college English teaching in independent colleges. *Journal of Central China Normal University (Humanities and Social Sciences)*, *S4*, 164.

Clark, R. M., Kaw, A., & Besterfield-Sacre, M. (2016). Comparing the effectiveness of

blended, semi-flipped, and flipped formats in an engineering numerical methods course. *Advances in Engineering Education*, *5*(3), n3.

Coyle, D. (2007). Content and language integrated learning: Towards a connected research agenda for CLIL pedagogies. *International Journal of Bilingual Education and Bilingualism*, *10*(5), 543–562. https://doi.org/10.2167/beb459.0

Coyle, Y., & de Larios, J. R. (2020). Exploring young learners' engagement with models as a written corrective technique in EFL and CLIL settings. *System*, *95*, 102374. https://doi.org/10.1016/j.system.2020.102374

de Zarobe, Y. R., & Lasagabaster, D. (2010). *CLIL in Spain: Implementation, results, and teacher training*. Newcastle upon Tyne, UK: Cambridge Scholars Publishing.

Fisher, R., Perényi. Á., & Birdthistle, N. (2018). The positive relationship between flipped and blended learning and student engagement, performance, and satisfaction. *Active Learning in Higher Education*, 1469787418801702. https://doi.org/10.1177/1469787418801702

Fredricks, J. A. M. (2012). *The measurement of student engagement: A comparative analysis of various methods and student self-report instrument.* US: Springer.

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning*, *27*(1), 70–105. https://doi.org/10.1080/09588221.2012.700315

Halverson, L. R., & Graham, C. R. (2019). Learner engagement in blended learning environments: A conceptual framework. *Online Learning*, *23*(2), 145–178. https://doi.org/10.24059/olj.v23i2.1481

He, L. (2020). New requirements for college English teaching in the new era. *Foreign Language World*, *4*, 13–18.

Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring student engagement in technology-mediated learning: A review. *Computers & Education*, *90*, 36–53. https://doi.org/10.1016/j.compedu.2015.09.005

Hou, J. (2010). College English teaching practice and research based on blended learning. *E-Education Research*, *5*, 108–111.

Khader, N. S. K. (2016). The effectiveness of blended learning in improving students' achievement in third grade's science in Bani Kenana. *Journal of Education and Practice*, *7*(35), 109–116.

Kunjie, L. (2021). Design and Practice on EFL Course for Ideological and Political Education

Based on Sino-Foreign Cooperative Education. *Journal of Yellow River Conservancy Technical Institute 33*(2), 85-90.

Li, D., & Zhang, L. (2020). Exploring teacher scaffolding in a CLIL-framed EFL intensive reading class: A classroom discourse analysis approach. *Language Teaching Research*, 1362168820903340. https://doi.org/10.1177/1362168820903340

Lombardini, C., Lakkala, M., & Muukkonen, H. (2018). The impact of the flipped classroom in a principles of microeconomics course: evidence from a quasi-experiment with two flipped classroom designs. *International Review of Economics Education*, 29, 14–28. https://doi.org/10.1016/j.iree.2018.01.003

Mahmud, M. M. (2018). Technology and language–what works and what does not: A meta-analysis of blended learning research. *Journal of Asia TEFL*, *15*(2), 365.

Malik, M. N., Khan, H. H., Chofreh, A. G., Goni, F. A., Klemeš, J. J., & Alotaibi, Y. (2019). Investigating students' sustainability awareness and the curriculum of technology education in Pakistan. *Sustainability*, *11*(9), 2651. https://doi.org/10.3390/su11092651

Mayo, M. D. P. G., & de los Angeles Hidalgo, M. (2017). L1 use among young EFL mainstream and CLIL learners in task-supported interaction. *System*, *67*, 132–145. https://doi.org/10.1016/j.system.2017.05.004

Morton, C. E., Saleh, S. N., Smith, S. F., Hemani, A., Ameen, A., Bennie, T. D., & Toro-Troconis, M. (2016). Blended learning: How can we optimise undergraduate student engagement? *BMC Medical Education*, *16*(1), 1–8. https://doi.org/10.1186/s12909-016-0716-z

Murillo-Zamorano, L. R., Sánchez, J. Á. L., & Godoy-Caballero, A. L. (2019). How the flipped classroom affects knowledge, skills, and engagement in higher education: Effects on students' satisfaction. *Computers & Education*, *141*, 103608. https://doi.org/10.1016/j.compedu.2019.103608

O'Dowd, R. (2018). Innovations and challenges in using online communication technologies in CLIL. *Theory into Practice*, *57*(3), 232–240. https://doi.org/10.1080/00405841.2018.1484039

Picciano, A. G., Dziuban, C. D., & Graham, C. R. (2013). *Blended learning: Research perspectives* (Vol. 2). New York, US: Routledge.

Pilotti, M. A. S, Hardy, P., Murphy, P., & Vincent, P. (2017). Factors related to cognitive, emotional, and behavioral engagement in the online asynchronous classroom. *International Journal of Teaching and Learning in Higher Education*, *29*(1), 145–153.

Pinto-Llorente, A. M., Sánchez-Gómez, M. C., García-Peñalvo, F. J., & Casillas-Martín, S.

(2017). Students' perceptions and attitudes towards asynchronous technological tools in blended-learning training to improve grammatical competence in English as a second language. *Computers in Human Behavior*, *72*, 632–643. https://doi.org/10.1016/j.chb.2016.05.071

Ross, H., Cen, Y., & Shi, J. (2014). Engaging students in China. In H. Coates & A. C. McCormick (Eds.). *Engaging university students* (pp. 93–107). Springer.

Sahin-Kizil, A. (2014). Blended instruction for EFL learners: Engagement, learning and course satisfaction. *JALT CALL Journal*, *10*(3), 175–188. https://doi.org/10.29140/jaltcall.v10n3.174

Serrano, D. R., Dea-Ayuela, M. A., Gonzalez-Burgos, E., Serrano-Gil, A., & Lalatsa, A. (2019). Technology-enhanced learning in higher education: How to enhance student engagement through blended learning. *European Journal of Education*, *54*(2), 273–286. https://doi.org/10.1111/ejed.1233

Shen, K. (2019). Study on cultivating students' capacity of creation and innovation—Reform and effect based on CLIL. *Journal of Shanghai University of International Business and Economics*, *3*, 99–108.

Tachie, S. A. (2019). Challenges and opportunities regarding usage of computers in the teaching and learning of Mathematics. *South African Journal of Education*, *39*(1). https://doi.org/10.15700/saje.v39ns2a1690

Tagnin, L., & Ní Ríordáin, M. (2021). Building science through questions in Content and Language Integrated Learning (CLIL) classrooms. *International Journal of STEM Education*, *8*(1), 34. https://doi.org/10.1186/s40594-021-00293-0

Talan, T., & Gulsecen, S. (2019). The effect of a flipped classroom on students' achievements, academic engagement, and satisfaction levels. *Turkish Online Journal of Distance Education*, *20*(4), 31–60. https://doi.org/10.17718/tojde.640503

Wang, N., Chen, J., Tai, M., & Zhang, J. (2019). Blended learning for Chinese university EFL learners: learning environment and learner perceptions. *Computer Assisted Language Learning*, 1–27. https://doi.org/10.1080/09588221.2019.1607881

Wen, Q., & Sun, M. (2015). On key issues about critical thinking development in college EFL classroom. *Foreign Language Teaching*, *3*, 6–12.

Yang, W. & Yang, L. (2022). Evaluating learners' satisfaction with a distance online CLIL lesson during the pandemic. *English Teaching and Learning, 46,* 179-201.

Yapici, İ. Ü. (2016). Effectiveness of blended cooperative learning environment in biology teaching: classroom community sense, academic achievement and satisfaction. *Journal*

of Education and Training Studies, *4*(4), 269–280. https://doi.org/10.11114/jets.v4i4.1372

Zhang, R. & Zou, D. (2020). Types, purposes, and effectiveness of state-of-the-art technologies for second and foreign language learning. *Computer Assisted Language Learning*, 1–47. https://doi.org/10.1080/09588221.2020.1744666