Heliyon 10 (2024) e39363

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Do concept maps prompt long-term memory in CLIL schoolchildren?

José Luis Gómez Ramos [a,*], Rosa María López Campillo [b], Isabel López Cirugeda [b], José Luis Palazón-Fernández [c]

[a] Department of Education and Psychology, University of Extremadura, Spain
[b] Department of Modern Philology, University of Castilla-La Mancha, Spain
[c] CUE Salus Infirmorum, University of Cádiz, Spain

ARTICLE INFO

Keywords:
Bilingual education
CLIL
Concept mapping
Instructional design
Long-term memory

ABSTRACT

To address the learning disadvantages that students might face in a foreign language (L2) environment, this research explores the impact of concept mapping on enhancing long-term memory retention among Content and Language Integrated Learning (CLIL) pupils. The research design is quasi-experimental and involves sixty Spanish primary education students non-randomly assigned from a Spanish-English semi-public bilingual school. The main hypothesis investigated is that concept maps serve as effective teaching instruments for enhancing long-term memory. During training, new information was systematically integrated with students' prior knowledge to aid long-lasting retention of instructional content. The study examines the use of concept maps in a CLIL classroom as a complement to traditional teaching strategies and instructional design, focusing on improving the recall of previously learned information. Findings show that concept maps are reliable tools for assisting CLIL students whose mother tongue (L1) is not English, significantly reducing forgetfulness because of the precision of the transmitted information. The study shows the potential of concept mapping to improve learning outcomes in L2 settings by providing visual, accurate, and structured learning aids.

1. Introduction

Concept mapping and rote and meaningful learning of syllabus material have a long history in education [1]. Student evaluations often reflect the type of learning—by rote, meaningful, or a mix of both—shaped by course instructional design, instructor's teaching methods and styles, and the accuracy with which curricular information is presented to students [2–5]. As a result, researchers have examined the acquisition of meaningful learning by students and the ongoing meaning-making processes in the learning of subject matter [6]. They have also explored how long-term retention of learned information and forgetfulness, also referred to by Ausubel [7] as *obliteration*, does affect students' performance [7–11], and how the benefits of activating prior knowledge make learners gain learning strategies to not struggle at school [12].

Although concept mapping remains authoritative for meaningful learning [13] and is frequently considered among educators [14, 15], there is a scarcity of evidence supporting its use in CLIL settings and certainty about its proper management. The goal is not to

* Corresponding author.
E-mail addresses: joseluisgr@unex.es (J.L. Gómez Ramos), rosa.lcampillo@uclm.es (R.M. López Campillo), isabel.lcirugeda@uclm.es (I. López Cirugeda), jluis.palazon@uca.es (J.L. Palazón-Fernández).

https://doi.org/10.1016/j.heliyon.2024.e39363
Received 26 September 2023; Received in revised form 4 October 2024; Accepted 12 October 2024
Available online 15 October 2024

exclude traditional teaching but to use concept maps correctly in CLIL or foreign language (L2) scenarios to help students (re-)construct knowledge, thereby strengthening their long-term memory [16] for later retrieval during, for example, end-of-unit exams. Since "word meanings are psychologically represented through the mapping of words to conceptual structures" ([17], p. 338), it is necessary to observe how long-term memory in bilinguals is affected (either positively or negatively) when concept mapping is integrated into everyday pedagogy.

Considering the above, this study investigates how concept maps might enhance learning in a Content and Language Integrated Learning (CLIL) classroom when used alongside conventional teaching methods. The study aims to improve students' long-term retention of information and examines how concept maps impact forgetfulness. It involves English as a Foreign Language (EFL) students from a semi-public Spanish bilingual school with a CLIL-based curriculum fully implemented in English. Participants are divided in two similar groups: one receives daily concept mapping training to support learning throughout the school year, while the other does not. The study also considers covariates, such as students' intelligence, learning styles, and English proficiency, and emphasises the value of incorporating visual and interactive tools into daily instruction.

1.1. *Content and Language Integrated Learning*

Content and Language Integrated Learning (CLIL) is an instructional approach—commonly through an L2 like English [18]—that integrates curricular contents with language learning [19]. The primary goal of CLIL is to help students learn both content and language simultaneously, which has been proved to enhance student engagement, improve language skills, and deepen subject-matter comprehension [19]. This interdisciplinary approach can be applied in many subjects, such as mathematics, natural and social sciences, or even physical education, employing instructional sub-strategies like project-based, inquiry-based, and problem-based learning [19]. While suggestions for using graphic organisers, particularly mind maps, are frequently found in CLIL literature [19–21], recommendations for concept maps are less common [22], revealing that the use of concept maps in CLIL settings is limited [23].

By developing both language competency and subject-matter knowledge simultaneously, students are expected to build up deeper understanding of curriculum concepts and ideas, enhance higher-order thinking skills [19,24], and improve L2 and teamwork competencies in a more comprehensive way [25]. In CLIL settings, increased student motivation leads to greater emotional engagement in learning activities [26]. CLIL does also support the expansion of students' word stock and language skills, thus promoting school success [19]. Furthermore, integrating language learning with curriculum's subject matter study in CLIL settings fosters cultural awareness and understanding by exposing students to diverse cultures and perspectives, which cultivates a global outlook, enhances comprehension, and encourages tolerance [19].

1.2. *Concept mapping in education settings*

Concept maps are visual representations designed to organise large and complex amounts of information (commonly hierarchically) from a particular domain, allowing for a better understanding of the mutual relations between ideas [27,28]. These visual organisers are not only useful to help concept mappers (cmappers) understand the relationships between familiar or unfamiliar concepts and ideas, but also to see how these fit together as part of a much larger system [28]. One benefit of using concept maps in the classroom would be that, apart from enabling learners to visualise the whole set of information to be learned, they would also provide apprentices with accurate reading skills, thus, improving pupils' learning outcomes and consecutive recall of previously learned information [29].

Rather than focusing on individual facts or isolated chunks of learning material, the use of concept maps helps students see how the propositions from a larger concept fit together, aiding in-depth understanding of such a curricular material and improving long-term retention of information [28]. Conceptualising or visualising information throughout concept maps does also reduce pupils' cognitive load and the emerging barriers they might face when managing large amounts of information [30]. Regarding this, concept maps help students understand complex ideas across a wide range of subjects, such as science, maths, history, and literature [28]. As a result, by encouraging learners to make their own concept maps, instructors may promote active learning, deeper understanding, and the creation of more dynamic and meaning-making learning environments [31].

1.3. *Cognitive processes in memory formation*

Although the learning of lexicon is essential in Second Language Acquisition (SLA), communication, and/or social interaction in the L2, vocabulary recall may be cognitively demanding because of the omission in considering learners' memory systems when developing "effective teaching strategies that support student learning and retention" by instructional designers ([32], p. 250). Word-meaning retrieval or elucidation by students in, for example, a final exam may make them struggle simply because they are not familiarised with the task [33]—in which case will not be the student's lack of retention but lack of task familiarisation. Word abstractness, test structure incomprehension, and the unawareness of the grammatical or lexical categories at hand may also affect memory at a certain point [34–36]. Although in the recalling of information meaning-giving appears to be more helpful than meaning-inferring [37], there is a need to know what educators teach and if they expect students to memorise, conceptualise, or both of them, and what is the final purpose of it best benefiting students.

While some researchers declare that "under some conditions of presentation, the degree of learning will be a function of total time, regardless of the duration of the individual trials or interitem times" ([38], p. 409); or that, assuming all other factors are constant, the extent of learning is determined by the quantity of time that a learner actively engages in the learning process [39]; to others, learning efficiency and memory retention can be optimised through the (re-)scheduling of proper study (S), test (T), and rest (R) varying programmes in type, mode, density, et cetera [40,41]. Thus, the retention of previously learned information not only depends on the time dedicated to learning but also on other factors like student's awareness of retention intervals, short and long-term overconfidence factors in their memories, and stability bias [42].

Considering the above, what occasionally causes students to struggle in learning slower, requiring a higher effort (e.g. when they have to learn to create a concept map to synthesise and learn information), may lead them to better memory retention in the long term [43]. Also, formative assessment techniques such as peer and self-assessment can be incorporated to promote student reflection and to develop metacognitive skills in them [44,45]. The constant feedback may help learners understand their strengths and areas for improvement to achieve learning outcomes [46,47]. What is more important, by analysing assessment results, instructional designers can identify gaps in learning and revise the instructional strategies or content to address these gaps [48] and contribute to better students' long-term memory construction.

1.3.1. Purpose of the present study

In this study, we analysed the usage patterns of concept mapping among CLIL students and assessed its effectiveness in promoting long-term retention of the information. We devised the research question: Does daily concept mapping practice enhance students' long-term retention of information, thus improving performance on end-of-unit exams? Yet, our investigation focuses on the impact of daily use of concept mapping on students' information retention in the CLIL classroom. It aims to determine whether the regular engagement with concept mapping improves students' inner ability to maintain information for long time. To achieve this, some other secondary objectives were considered: [i] to address the influence of students' English skills; [ii] intelligence quotient; and [iii] learning styles in their long-term retention of information and exam performance.

To provide an answer to the research question, participants were tasked with using concept maps during Natural Science class to aid comprehension and retention of course material throughout the school year. Our study focused on the dependent variable of long-term memory, with concept mapping as the independent variable of interest. Covariates such as intelligence quotient ("*g*"factor); learning styles, measured via the Group Embedded Figures Test (GEFT); and foreign language skills—(English) listening, reading, writing—were also considered. Through the statistical analysis of variables and covariates (ANCOVA), our research aimed to provide useful insights into the effectiveness of concept mapping as effective pedagogical tool to enhance long-term memory in CLIL students.

2. Method

2.1. Sample

While much educational research often relies on larger samples and broader instructional patterns, our study employed a smaller sample ($N = 60$) drawn from two parallel fifth-grade classes, with the instruction *precisely* designed and administered by one of the researchers. Despite the presence of tutor instructors in each group throughout the school year, the instructional content was uniform across both treatment and control groups. To manage individual differences in teaching, concept mapping training was exclusively delivered by the researcher to the treatment group. This training was consistently integrated into all lessons for the entire research period. The bilingual school chosen for this study offered nearly 30 % of its curriculum subjects in English, taught by staff trained in CLIL and the L2, but not in concept mapping. Instead of training the teachers, the researcher designed and implemented the concept mapping activities, ensuring students accurately understood the concept mapping process and maintaining privacy/integrity during data collection.

Thus, the sample chosen for this study comprised primary education students from a Spanish bilingual school. Because of the bilingual curriculum, students were taught several subjects, besides English, entirely in the L2. Among these subjects, Natural Science was considered the most suitable for the experiment because of the proneness and difficulty of its contents. Participants' characteristics did not differ significantly, except for one particular student with special needs (SEN) in the experimental group. The participants were 60 students aged nine to eleven ($M_{age} = 9.78$; $SD = 0.49$), enrolled in a semi-public school with an implemented CLIL-based English curricular programme. The age distribution among the groups was homogeneous ($t_{56} = 1.32$; $p = 0.191$). Table 1 displays the results of *t*-tests conducted to assess group homogeneity before the intervention for each of the study covariates chosen—where it shows that research starts from two equivalent groups and that pupils are homogeneous in terms of reading and writing skills and intelligence

Table 1
Means ± SD of scores in pre-tests as a function of treatment conditions.

Covariates	Control	Treated		
	$\overline{X} \pm SD$	$\overline{X} \pm SD$	t_{57}	p
Prior marks[a]	8.53 ± 1.38	8.13 ± 1.66	1.016	0.314
Intelligence	113.43 ± 14.54	109.90 ± 16.85	0.869	0.388
Learning Styles	5.13 ± 3.90	4.67 ± 4.20	0.446	0.658
Listening	7.71 ± 1.61	7.41 ± 1.82	0.694	0.491
Reading & Writing	6.23 ± 1.77	6.37 ± 2.32	−0.253	0.801

[a] Averaged marks on the sample from the previous school year to the start of the research.

quotient. It also shows how results obtained from the Group Embedded Figures Test (GEFT) are likewise similar in both groups.

Concerning the sample, participants were naturally given/assigned to both groups; however, the selection of treated and control groups was randomly assigned by flipping a coin. After this assignment, participants remained distributed as follows: one control group encompassing 21 males and 9 females, and one treated group encompassing 19 males and 11 females. Although the number of males was higher, no significant differences were found concerning the marks gotten by students during the previous school year based on gender variables, either between or within groups. Regarding the particular case referenced above, and despite the latent data bias respecting marks and daily performance, the scores of the SEN in the treated group were not excluded from the statistical analysis. Although including these data did not significantly alter results, it made the intervention less effective. The decision not to exclude this student to make the results more beneficial emerged from the belief that classrooms are real-life environments, not laboratories [8], and thus participants and collected data should remain unaltered in nature.

2.2. Instruments

2.2.1. General intelligence factor "g" Cattell

This test, adapted to Spanish by TEA, was used to assess individual intelligence in participants. The main reason for choosing it was its non-verbal nature, as well as its close similarity to the Group Embedded Figures Test (GEFT) for task completion. In both tests—Cattell's Factor "*g*" & Witkin's GEFT—, students had to perceive and discriminate shapes and figures. Thus, the IQ test was chosen for its suitability, easiness, compatibility with the GEFT, and the omission of covariant factors such as verbal fluency, cultural level, or individual behaviour. For the study, pupils were not expected to demonstrate oral or motor skills, but rather to interpret and perceive patterns involving perception or perceptual differences. In this sense, the test is a reliable psychometric instrument that reduces influences unrelated to intelligence and that is suitable because of its ease of administration and correction [49].

Although the same intelligence test has different scales adapted to various ages of participants (1, 2, 3), the one selected for our experiment was Scale 2, designed for learners aged 8 to 14. Scale 1 is intended for pupils aged 4–8, and Scale 3 for students in the final years of compulsory secondary education. For test administration, the material consisted of an Interpretation and Instruction Manual, problem booklets, and answer sheets (in our case, we used manual correction sheets). As supplementary material, we used a stopwatch, pencils, and erasers. It is worth noting that no inconvenience or incident arose during the test administration process. Since the researcher did not use the Spanish language at any time, test instructions were provided in Spanish by the teacher-tutor who shared the classroom with the researcher throughout the school year.

The test included the following.

1 *Series*. This task involves selecting, from among incomplete and progressive series, the appropriate response to the continuing series. Scale 2-A: 3 min to complete the series of 12 elements.
2 *Classification*. In this task, testees identify the distinct figures among the five figures presented in each of the elements. Scale 2-A: 4 min to complete the series of 14 elements.
3. *Matrices.* In this task, testees complete the grid of drawings or matrices presented to them. To do this, they choose one of the five possible solutions proposed in the test. Scale 2-A: 3 min to complete the series of 12 elements.
4. *Conditions.* In the last part, testees choose the alternative option from among the figures or squares provided as a reference. Scale 2-A: 2.5 min to complete the series of 8 elements.

2.2.2. Group Embedded Figures Test (GEFT)

Though the instructions provided in the GEFT booklets were concise and the tasks simple and visual, they were written in English. To ensure understanding by all participants, as was done with the IQ test, instructions were provided to students in Spanish by the tutor-teachers of the subject in each group. During task completion, the researcher provided support in both groups (treated and control) when necessary. To ensure proper administration, teachers and the researcher met beforehand to clarify any doubts about testing instructions and procedures. The GEFT test had a format similar to the Factor "*g*" test—both were administered individually. The materials required for the GEFT included pencils, erasers, and a stopwatch for time control.

Regarding time, both tests resulted dynamic to students and not overly prolonged, with a difference of only 7.5 min between them: 12.5 min for the Factor "*g*" test versus 20 min for the GEFT. The brevity of these tests was necessary to prevent discouragement and maintain students' attention during completion. However, since all pretests were administered during the first week of the school year and the posttests during the last week, some students showed signs of fatigue due to the testing process.

The test included the following.

- *Section 1*. Introduction to the tasks in the following sections.
- *Sections 2 and 3*. Each encompassing nine items whose difficulty increases as the task progresses (Section 3 implying a higher level of difficulty and abstraction). The activities in all sections keep the same colour representation of the figures (blue). Based on Gestalt principles, representations contribute to better identification of embedded figures [50–53].

2.2.3. Cambridge "Movers" English test

Movers knowledge test, designed and marketed by Cambridge Publishing [54], was used to assess participants' English level before and after the experiment. The purpose was to know pupils' English skills before the experiment and to verify if the experimental factors influenced L2 improvement. Although it is an official test designed to certify A1 level proficiency (Common European Framework of

Reference for Languages [55]), we adapted it omitting both speaking and conversation parts. Movers test constitutes the second initial element of the three marketed by the publisher (Pre A1: Starters; A1: Movers; A2: Flyers), and it is specifically designed for non-English-speaking learners age-ranged in the early ages. Though student preparation for these types of tests is typically over a year, in our study this did not happen. It was not relevant for us to instruct students to pass the test, but rather to know their English level on three specific skills of interest for the research: reading, writing, and listening.

The Starters, Movers, and Flyers tests are designed for pupils aged 6 to 12, and their design include drawings to keep examinees motivated during their completion through an attractive and visual layout [54]. As mentioned, through the test we measured three parts: *listening*, *reading*, and*writing*. The first part (listening) was conducted as a single test, with a duration of approximately 25 min. The following parts (reading and writing) were jointly applied in a single 30-min test.

The test included the following.

- *Listening.* Introduction to the tasks in the following sections. Task 1. Students link the names on the list with the drawings of the people depicted in the given photograph by drawing lines. Task 2. Students respond to questions they have previously heard by writing numbers and letters. Task 3. Students relate the people or objects belonging to one group to the representations and objects belonging to another different group by writing letters in the boxes. Task 4. Students choose which representation corresponds to the provided information by marking boxes. Task 5. By colouring the blank drawings, students choose which colour, number, or word to assign to each part of the drawing based on what they heard.
- *Reading* and *Writing*. Task 1. By copying words, students match the representations and words with the sentences that describe them. Task 2. By circling a letter from the multiple-choice set, students choose the correct answer. Task 3. By copying words and marking boxes, students choose the appropriate words or representations in a short story and select the appropriate title for it. Task 4. By copying words, students choose which ones are suitable to fill in the blanks in a short text. Task 5. By writing one, two, or three words, students decide which ones are most suitable to complete the sentences in a given story. Task 6. By writing words and phrases, students complete sentences, answer questions, and write sentences about a presentation.

2.3. Research design

The research is quasi-experimental, with participants selected non-randomly and all receiving the same instruction except for the treatment condition involving concept maps. An inferential hypothesis is included in the study, which examines the impact of using concept mapping in everyday teaching on the long-term memory of students with limited or uneven bilingual proficiency. The researchers compared two groups (treatment and control), using quantitative statistical analysis to identify differences between them, focusing on the treatment group and taking the control group as the unaltered group of reference.

2.4. Procedure

The experimental procedure was designed as potentially meaningful learning material to deductively integrate the varied types of curricular materials and the experiment methods over the course of a school year to ensure a thorough approach to concept mapping. The procedure for the yearly experiment involved the use of conventional, unconventional, and standardised materials, which were implemented systematically during the phases outlined below.

2.4.1. Preparatory phase

1. Identify and select the school and sample participants for the research (≈15 days)
2. Seek approval for the intervention from authorities (≈20 days)
3. Design the intervention (≈3 months)

2.4.2. Implementation phase

1. Pre-tests implementation (≈7 days)
2. Implementation of conventional materials aligned with the curriculum, consisting of contents and tests provided by the subject's book publisher, and administered after the end of each unit of work (≈8 months, or ≈1 month per unit of work). During this official testing, students in both groups were *informed* with sufficient time in advance (≈1 week) for them to study. Then, tests after the completion of each unit of work to measure students' knowledge of curricular contents revealed the punctuations that later will appear as the subject's final marks in their bulletins.
3. Simultaneously with 2 above, development of *ad hoc* activities for concept mapping throughout the school year.
4. Administration of an *uninformed* unofficial test covering syllabus contents from *all* previous units of work to test long-term retention of the information (≈1 h). This full test maintains the same structure as the previous official tests and encompasses queries on all instructed content from day one of the yearly course.
5. Post-tests implementation (≈7 days)

During the implementation of concept maps, peer assessment techniques were also used to provide ongoing feedback and measure training effectiveness.

2.5. Statistical analysis

A one-way analysis of covariance was conducted to evaluate the influence of concept mapping on the long-term retention of learned information. The between-subject factor comprised the group of students that were trained on concept mapping and those in the control group who were not. The covariates comprised students' previous marks to the experiment, intelligence quotient, learning styles, and foreign language competency. Nuisance variables in the analysis are to observe their potential influence on the null hypothesis (H_0).

2.6. Ethical praxes

The research described adheres to strict ethical standards by obtaining approval from the school's Management Team as the Research Ethics Committee, securing individual consent from legal guardians/parents of participating students, and adapting the intervention to the developmental stages of children while ensuring voluntary participation. To protect participants' privacy, identifiable data was fully anonymised, and only parents requesting feedback were informed of outcomes. Data confidentiality was maintained by not sharing identifiable information or research data files with the school or families, reflecting a commitment to participant welfare and ethical research conduct.

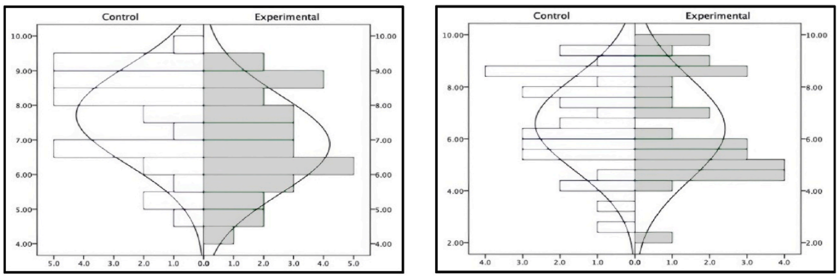
2.7. Results

Results from hypothesis testing reveal dichotomous findings between the INFORMED and UNINFORMED research conditions for both groups. Specifically, when the participants were informed of upcoming exam dates, significant statistical differences were observed across all evaluated yearly subject contents between the trained (experimental) and untrained (control) groups of participants ($F_{1.54} = 6.834$; $p = 0.012$; $\eta^2 = 0.112$). Paradoxically, and without a clear explanation from the researchers, these differences consistently favoured the control group throughout the study, which got better marks than the treated one (7.71 ± 1.41 *vs* 6.87 ± 1.42), respectively (Fig. 1a). On average, the difference was of 0.84 marks. From the total number of covariates included in the analysis, just reading and writing had a significant effect ($F_{1.54} = 18.057$; $p = 0.000$; $\eta^2 = 0.251$) in its relation with the dependent variable.

In contrast to the clear advantage of statistical results benefiting participants in the control group, differences just "vanish" when students in both groups are not informed of exam dates (Fig. 1b). Thus, non-significant differences were found between groups when they were *not* informed of upcoming exams and they did not have the opportunity to study for them ($F_{1.54} = 0.136$; $p = 0.714$; $\eta^2 = 0.003$). The effects due to IQ ($F_{1.54} = 0.503$; $p = 0.481$; $\eta^2 = 0.009$), and students' listening skills from the L2 ($F_{1,54} = 3.800$; $p = 0.056$; $\eta^2 = 0.066$) were also non-significant. There was, however, a significant effect of learning styles ($F_{1.54} = 5.085$; $p = 0.028$; $\eta^2 = 0.086$) and reading skills in English ($F_{1.54} = 12.266$; $p = 0.001$; $\eta^2 = 0.185$). As it happened before, the mean values pertaining the control group were slightly (but not significantly) higher than those for the treated group (6.58 ± 1.80 *vs* 6.39 ± 1.98), respectively. The variation in the arithmetic means between the groups was 0.19 marks.

The above results reveal that the control group performed better when *informed* of the exam date, indicating that their marks were dependent on having the opportunity to study. However, their performance declined when they were *uninformed* of the exam date, highlighting a significant gap in understanding the interaction between intervention strategies and study conditions. This discrepancy suggests that the advantages observed in the control group may not be solely attributable to the instructional method but rather to their ability to prepare for the exams. This raises important questions for future research regarding the impact of preparation time on exam performance and underscores the need to explore how different study conditions affect the efficacy of educational interventions.

In essence, participants in the control group experience a significant drop in marks when they lack the opportunity to study for an exam. This substantial fluctuation is not observed in the trained group, however, whose performance remains stable regardless of whether they are informed about upcoming exams. Specifically, the control group shows a notable variation of 1.13 marks between the *informed* and *uninformed* conditions (Fig. 2a). In contrast, the trained group exhibits minimal variation of only 0.48 marks between



(a) (b)

Fig. 1. This figure illustrates the impact of exam notifications on student performance. **(a)** Shows the results when students were informed of upcoming exam dates, with the control group achieving significantly higher marks despite not receiving training. **(b)** Displays the results when students were not informed of exam dates, revealing that the advantage of the control group diminished significantly.

these same conditions (Fig. 2b), indicating that their performance—and thus their long-term memory or information retrieval—is consistent regardless of exam awareness. Consequently, the statistical differences between the groups shift from significant in the *informed* condition to non-significant in the *uninformed* condition. Importantly, the results for the experimental group included data from the SEN student across both conditions, offering a comprehensive view of overall participant performance.

3. Discussion

As seen in other studies with pre-service teachers, which demonstrate the benefits of concept maps as visual organisers [56], our research does also provide experimental evidence supporting their advantages in the CLIL classroom. A significant part of this *perceptual* benefit arises from the hierarchical visualisation of connections between different word categories that students encounter or need to learn at a given point [57]. This conceptual representation also contributes to reinforcing students' long-term memory [58]. As noted by Novak [58], our annual research did also reveal that concept mapping *must* be implemented progressively to achieve satisfactory results; otherwise, the strong theoretical foundation underlying concept maps may be undermined. Specifically, the complexity of concept map creation should increase to help students understand their task and become familiar with the stages, theoretical complexities, and skills involved. To maximise the benefits of concept mapping, students need to grasp the purpose, process, and optimal timing for making concept maps. This aligns with Shavelson et al. ([57], p. 428), who declare that concept maps "help to know what and how students are learning and accommodating the new information learned."

Thus, from our study, three main interpretations related to CLIL, instructional design, and long-term retention of information emerge. The *first* interpretation is that while higher English skills may help students understand L2 content adequately, they do not necessarily enhance long-term memory. Specifically, concept mapping correlates with improved long-term memory even when students' L2 English skills are lower. This revelation emphasise the importance of accurate content transmission for bilingual memory. The *second* finding highlights the significance of instructional design and its explicitness for bilinguals. Conventional L1 teaching methods should be "cautiously" integrated into the CLIL classroom; when teaching is well-designed and adapted to students' specific needs and learning paths, students are more engaged and participative, and negative stimuli are minimised. The *third* key finding reveals that when learners are informed about an upcoming exam, they tend to review and memorise content previously covered in class, indicating that they may not fully rely on the knowledge acquired during their training period but focus more on content they expect to be tested on.

4. Theoretical-practical implications

Despite the challenges associated with producing accurate concept maps [59], the success of the yearly programme with children inexperienced in concept mapping underscores the importance of well-designed instruction and in-depth understanding of the *theoretical* foundations of these visual organisers. Either by learners or instructors, missteps in implementing concept mapping often emerge from a lack of understanding of its theoretical roots. In our study, as observed by Schroeder et al. [30], students became more focused on the learning material and less on the concept mapping process as their expertise grew, a pattern clearly seen among the most advanced cmappers.

Regarding the above information, we recommend that instructors provide students with a brief explanation of Ausubel [7,8,10] and Novak's [14,60,61] learning theories and the purpose of concept mapping before assigning them the task of making or interpreting concept maps by means of didactic transposition [62]. By doing so, we anticipate that adverse reactions towards concept mapping will be mitigated, thereby facilitating comprehension and fostering long-term retention of information. In our study, we observed that students' initial attempts at concept mapping produced detrimental rather than beneficial results. However, as students progressed, they demonstrated improved knowledge management and mapping skills. This underscores that concept mapping is a gradual and cognitively demanding process, producing satisfactory outcomes only when students understand its goals—there is compelling research suggesting that concept maps enhance users' metacognitive skills and awareness [63].

The *practical* implications from the study suggest the viability of integrating concept maps into daily CLIL classroom practices to boost long-term memory retention among EFL learners. Nonetheless, this approach demands considerable effort and dedication from

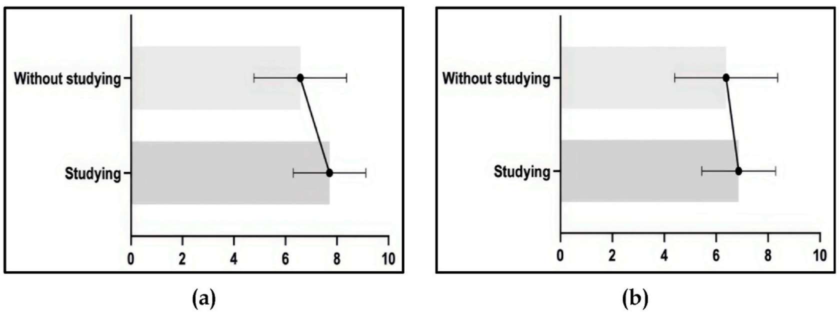


Fig. 2. Variation in exam marks across conditions. **(a)** The control group's performance, showing substantial variation of 1.13 marks between the *informed* and *uninformed* conditions. **(b)** The trained group's performance, revealing minimal variation of 0.48 marks between the same conditions.

all parties involved. For example, teachers must meticulously design instructional materials to seamlessly integrate concept map activities, ensuring optimal learning outcomes. The implementation should also be gradual, allowing students to progressively develop their understanding and proficiency over time. From the students' perspective, the initial stages of concept mapping may lead to cognitive challenges, whereas later stages may become less demanding. Hence, finding a balance in the purpose of concept mapping is essential because long-lasting tasks do not necessarily lead to enhanced learning outcomes.

In sum, the research hypothesis is both explained and supported by the study, which shows that trained students demonstrated higher rates of information retrieval. This finding suggests that both the conscious use of concept mapping and the unconscious visualisation and management of curricular content through concept maps contributed to the prolonged retention of new content in the trained group. The results imply that long-term knowledge can be effectively retained by students without additional study before exams. Additionally, it was observed that high English proficiency is not essential for meaningful learning and understanding of curricular content. This is especially relevant for students with lower proficiency in the foreign language, as allowing responses in either their native language (L1) or the L2 could potentially improve their performance. These insights underscore the effectiveness of concept mapping as a teaching strategy, offering considerable benefits for bilingual education and supporting the scholarly success of students.

5. Conclusions

In conclusion, concept maps, as effective in L1 educational settings, also prove viable, reliable, and effective in L2 classrooms, promoting long-term retrieval and memory retention among students. As instructional tools, concept maps assist both learners and educators in objectively assessing the quality of learning and teaching and their impact on learning outcomes. While paper-based concept maps offer certain benefits, adopting digital methods could further enhance their effectiveness. However, overemphasising routine procedures, implementing poorly structured intervention programmes, or failing to explain the fundamental theory of concept mapping to students may lead to adverse results. In English-Spanish bilingual schools, addressing students' English proficiency or other limitations through well-designed concept-mapping-based learning programmes can optimise student engagement and comprehension. Additionally, aligning instruction with both formative and summative assessment formats remains crucial for effective teaching and learning practices in CLIL settings.

6. Limitations

Despite the study's numerous strengths, some limitations were noted. First, the researcher's exclusive intervention in the experimental group, without parallel instruction in the control group, *might* have introduced inconsistencies. As ideally, both groups should have received uniform intervention only from him. However, this may be difficult to change in future research, because of schools' policies. Additionally, while English was consistently used in the experimental group, it is unclear whether this standard was constant in the control group. Student engagement did also present a challenge, with some participants showing minimal involvement in concept mapping, probably due to fatigue or lack of interest, which could have reduced the effectiveness of the activities. This could have been solved with a reinforcing programme of more catchy activities. Furthermore, the end-of-course exam, being standardised rather than tailored to the intervention, might have led to lower scores for the trained group. These limitations underscore the need for future research to address these pedagogical issues and improve the validity of the findings.

CRediT authorship contribution statement

José Luis Gómez Ramos: Methodology, Investigation, Conceptualization, Data curation, Formal analysis, Resources, Writing – original draft, Writing – review & editing. **Rosa María López Campillo:** Supervision, Conceptualization. **Isabel López Cirugeda:** Methodology, Resources. **José Luis Palazón-Fernández:** Formal analysis, Data curation, Writing – original draft, Writing – review & editing.

Data availability

The dataset supporting the current study has not been deposited in a public repository but is available upon reasonable request to the lead author.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] R.A. Berk, Top 10 Flashpoints in Student Ratings and the Evaluation of Teaching: what Faculty and Administrators Must Know to Protect Themselves in Employment Decisions, Routledge, 2013, https://doi.org/10.4324/9781003448211.

[2] S.K. Clark, S.R. Helfrich, L. Hatch, Examining preservice teacher content and pedagogical content knowledge needed to teach reading in elementary school, J. Res. Read. 40 (3) (2017) 219–232, https://doi.org/10.1111/1467-9817.12057.
[3] J.A. Fishman, R.I. Cooper, R. Ma, Bilingualism in the Barrio, Indiana University, 1971.
[4] Y.Y. Lo, Development of the beliefs and language awareness of content subject teachers in CLIL: does professional development help? Int. J. Biling. Educ. BiLing. 22 (7) (2017) 1–15, https://doi.org/10.1080/13670050.2017.1318821.
[5] H. Roothooft, Spanish lecturers' beliefs about English medium instruction: STEM versus Humanities, Int. J. Biling. Educ. BiLing. 25 (2) (2019) 1–14, https://doi.org/10.1080/13670050.2019.1707768.
[6] E. Kostiainen, T. Ukskoski, M. Ruohotie-Lyhty, M. Kauppinen, J. Kainulainen, T. Mäkinen, Meaningful learning in teacher education, Teach, Teach. Educ. 71 (2018) 66–77, https://doi.org/10.1016/j.tate.2017.12.009.
[7] D.P. Ausubel, Educational Psychology: A Cognitive View, Holt, Rinehart and Winston, 1968.
[8] D.P. Ausubel, The Acquisition and Retention of Knowledge: A Cognitive View, Springer Science+Business Media, Dordrecht, 2000, https://doi.org/10.1007/978-94-015-9454-7.
[9] D.P. Ausubel, The facilitation of meaningful verbal learning in the classroom, Educ. Psychol. 12 (2) (1977) 162–178, https://doi.org/10.1080/00461527709529171.
[10] D.P. Ausubel, The Psychology of Meaningful Verbal Learning, Grune & Stratton, 1963.
[11] D.P. Ausubel, A subsumption theory of meaningful verbal learning and retention, J. Gen. Psychol. 66 (2) (1962) 213–224, https://doi.org/10.1080/00221309.1962.9711837.
[12] A.L. Bruhn, T.S. Hasselbring, Increasing student access to content area textbooks, Interv. Sch. Clin. 49 (1) (2013) 30–38, 10.1177%2F1053451213480030.
[13] J.D. Novak, A.J. Cañas, The development and evolution of the concept mapping tool leading to a new model for mathematics education, in: K. Afamasaga-Fuata'i (Ed.), Concept Mapping in Mathematics: Research into Practice (3–16), Springer US, 2009, https://doi.org/10.1007/978-0-387-89194-1_1.
[14] D.P. Ausubel, J.D. Novak, H. Hanesian, Educational Psychology: A Cognitive View, Holt, 1978. Rinehart, and Winston.
[15] J.D. Novak, A.J. Cañas, The theory underlying concept maps and how to construct and use them, Technical Report IHMC CmapTools 2006-01 Rev 01-2008, Florida Institute for Human and Machine Cognition (2008) available at: https://acesse.dev/JFxNn.
[16] R.M. López Campillo, J.L. Gómez Ramos, Argumentation skills for the design of formative assessment queries, in: J.L. Gómez Ramos, I.M. Gómez Barreto (Eds.), Design and Measurement Strategies for Meaningful Learning, IGI Global, 2022. https://l1nq.com/HouKF.
[17] G.L. Murphy, The Big Book of Concepts, The MIT Press, 2002.
[18] P. Romanowski, CLIL models in Polish lower-secondary schools, Kwartalnik Neofilologiczny, LXV (4/2018. https://doi.org/10.24425/kn.2018.125006, 2023.
[19] D. Coyle, P. Hood, D. Marsh, CLIL: Content and Language Integrated Learning, Cambridge University Press, 2010.
[20] A.M.Y. Lin, Language across the Curriculum & CLIL in English as an Additional Language (EAL) Contexts: Theory and Practice, Springer, 2016. https://link.springer.com/book/10.1007/978-981-10-1802-2.
[21] Y.Y. Lo, Theoretical underpinnings of cross-curricular collaboration: how and why would it work in CLIL? in: Y.Y. Lo (Ed.), Professional Development of CLIL Teachers Springer Singapore, 2020, pp. 29–47, https://doi.org/10.1007/978-981-15-2425-7_3.
[22] D. Coyle, O. Meyer, S. Staschen-Dielmann (Eds.), A Deeper Learning Companion for CLIL: Putting Pluriliteracies into Practice, Cambridge University Press, 2023, https://doi.org/10.1017/9781009043755.
[23] J.L. Gómez Ramos, J. Lirio Castro, I.M. Gómez Barreto, Organisers, concept maps, and text summarisation in the CLIL classroom, in: A. Cañas, P. Reiska, J. Novak (Eds.), Concept Mapping: Renewing Learning and Thinking, Proceedings of the Eight International Conference on Concept Mapping, 2020, pp. 181–191, https://doi.org/10.31219/osf.io/be485.
[24] J. Biggs, C. Tang, Teaching for Quality Learning at University, Open University Press, 2011.
[25] I. Shevchenko, O. Kordyuk, Tasks and planning of English lessons according to content and Language integrated learning (CLIL) methodology in general secondary education institutions of Ukraine, Polonia University Scientific Journal 57 (2) (2023), https://doi.org/10.23856/5714.
[26] A.M. Puente, M. Rojas, J.C. Corral, Making emotions count through CLIL, Latin American Journal of Content & Language Integrated Learning 15 (2) (2023) e1521, https://doi.org/10.5294/laclil.2022.15.2.1.
[27] C. Limongelli, A. Marani, F. Sciarrone, M. Temperini, Measuring the similarity of concept maps according to pedagogical criteria, IEEE Access 10 (2022) 27655–27669, https://doi.org/10.1109/access.2022.3156662.
[28] J.D. Novak, Learning, Creating, and Using Knowledge: Concept Maps as Facilitative Tools in Schools and Corporations, second ed., Routledge, 2010.
[29] M. Wilson, C. Howell, K. Martin-Morales, S. Park, Concept mapping and reading comprehension, J. Polit. Sci. Educ. 19 (3) (2023) 408–431, https://doi.org/10.1080/15512169.2023.2164861.
[30] N.L. Schroeder, J.C. Nesbit, C.J. Anguiano, O.O. Adesope, Studying and constructing concept maps: a meta-analysis, Educ. Psychol. Rev. 30 (2) (2018) 431–455, https://doi.org/10.1007/s10648-017-9403-9.
[31] C. Dalton-Puffer, A construct of cognitive discourse functions for conceptualising content-language integration in CLIL and multilingual education, Eur. J. Appl. Ling. 1 (2013) 216–253, https://doi.org/10.1515/eujal-2013-0011.
[32] D.A. Sadiq, Enhancing vocabulary learning through understanding the human memory system: episodic memory, Rev. Amazonia Investiga 12 (62) (2023) 250–257, https://doi.org/10.34069/ai/2023.62.02.25.
[33] T. Merisuo-Storm, M. Soininen, Primary school pupils deriving word meaning from written context, Procedia Soc. Behav. Sci. 2 (2) (2010) 1625–1629, https://doi.org/10.1016/j.sbspro.2010.03.248.
[34] R.M. Best, R.G. Floyd, D.S. McNamara, Differential competencies contributing to children's comprehension of narrative and expository texts, Read. Psychol. 29 (2) (2008) 137–164, https://doi.org/10.1080/02702710801963951.
[35] K. Cain, J. Oakhill, K. Lemmon, Individual differences in the inference of word meanings from context: the influence of reading comprehension, vocabulary knowledge, and memory capacity, J. Educ. Psychol. 96 (4) (2004) 671–681, https://doi.org/10.1037/0022-0663.96.4.671.
[36] R.G. Fukkink, Deriving word meaning from written context: a process analysis, learn, Instructor 15 (1) (2005) 23–43, https://doi.org/10.1016/j.learninstruc.2004.12.002.
[37] J.-A. Mondria, The effects of inferring, verifying, and memorizing on the retention of L2 word meanings: an experimental comparison of the "meaning-inferred method" and the "meaning-given method,", Stud. Sec. Lang. Acquis. 25 (4) (2003) 473–499. http://www.jstor.org/stable/44486696.
[38] B.R. Bugelski, Presentation time, total time, and mediation in paired-associate learning, J. Exp. Psychol. 63 (4) (1962) 409–412, https://doi.org/10.1037/h0045665.
[39] R.P. Carver, A test of a hypothesized relationship between learning time and amount learned in school learning, J. Educ. Res. 64 (2) (1970) 57–58, https://doi.org/10.1080/00220671.1970.10884091.
[40] C. Izawa, A new temporary maintenance system, working cognition: a construct to incorporate working memory in unified theories of cognition [review of the book models of working memory: mechanisms of active maintenance and executive control, in: A. Miyake, P. Shah (Eds.), Am. J. Psychol. 114 (1) (2001) 115–126, https://doi.org/10.2307/1423383.
[41] C. Izawa, Total time and efficient time management: in search of optimal learning and retention via study-test-rest presentation programs, Am. J. Psychol. 113 (2) (2000) 221–248, https://doi.org/10.2307/1423729.
[42] N. Kornell, Failing to predict future changes in memory: a stability bias yields long-term overconfidence, in: A.S. Benjamin (Ed.), Successful Remembering and Successful Forgetting: A Festschrift in Honor of Robert A, Psychology Press, Bjork, 2011, pp. 365–386.
[43] H.L. Roediger III, J.D. Karpicke, Intricacies of spaced retrieval: a resolution, in: A.S. Benjamin (Ed.), Successful Remembering and Successful Forgetting: A Festschrift in Honor of Robert A, Psychology Press, Bjork, 2011, pp. 23–47.
[44] P. Black, D. William, Developing the theory of formative assessment, Educ. Assess. Eval. Account. 21 (2009) 5–31, https://doi.org/10.1007/s11092-008-9068-5.
[45] D. Wiliam, What is assessment for learning? Stud. Educ. Eval. 37 (2009) 3–14, https://doi.org/10.1016/j.stueduc.2011.03.001.

[46] J. Hattie, Visible Learning for Teachers: Maximizing Impact on Learning, Routledge, 2012.
[47] J. Hattie, H. Timperley, The power of feedback, Rev. Educ. Res. 77 (1) (2007) 81–112, https://doi.org/10.3102/003465430298487.
[48] L.A. Shepard, The role of assessment in a learning culture, Educ. Res. 29 (7) (2000) 4–14.
[49] R. Cattell, A. Cattell, Tests de Factor "g", Escalas 2 y 3: Manual [Measuring Intelligence with the Culture Fair Tests, Scales 2 and 3], TEA, Madrid, Spain, 2009.
[50] H.A. Witkin, D.R. Goodenough, Cognitive styles: essence and origins, Field Dependence and Field Independence (1981).
[51] H.A. Witkin, Cognitive Styles in Personal and Cultural Adaptation, Clack University Press, 1978.
[52] H.A. Witkin, R.B. Dyk, H.F. Paterson, D.R. Goodenough, S.A. Karp, Psychological Differentiation: Studies in Development, Wiley, 1962.
[53] H.A. Witkin, Individual differences in ease of perception of embedded figures, J. Pers. 19 (1) (1950) 1–15, https://doi.org/10.1111/j.1467-6494.1950.tb01084.x.
[54] Cambridge Assessment English, Cambridge English: Young Learners - Movers, Cambridge University Press, 2024.
[55] Council of Europe. https://acesse.dev/7bbby, 2024.
[56] T. Hartsell, Visualization of knowledge with concept maps in a teacher education course, TechTrends 65 (5) (2021) 847–859, https://doi.org/10.1007/s11528-021-00647-z.
[57] R.J. Shavelson, M.A. Ruiz-Primo, E.W. Wiley, Windows into the mind, High Educ. 49 (4) (2005) 413–430, https://doi.org/10.1007/s10734-004-9448-9.
[58] M. Kiefer, F. Pulvermüller, Conceptual representations in mind and brain: theoretical developments, current evidence and future directions, Cortex 48 (2012) 805–825, https://doi.org/10.1016/j.cortex.2011.04.006.
[59] A.J. Cañas, P. Reiska, J.D. Novak, Is my concept map large enough? in: A. Cañas, P. Reiska, J. Novak (Eds.), Innovating with Concept Mapping, Cham Springer International Publishing, 2016, pp. 128–143, https://doi.org/10.1007/978-3-319-45501-3_10.
[60] J.D. Novak, Learning, Creating, and Using Knowledge: Concept Maps as Facilitative Tools in Schools and Corporations, Lawrence Erlbaum Associates, 1998.
[61] J.D. Novak, A Theory of Education, Cornell University Press, 1977.
[62] Y. Chevallard, M. Bosch, Didactic Transposition in Mathematics Education, Springer, 2020, pp. 214–218, https://doi.org/10.1007/978-3-030-15789-0_48.
[63] B.A. Schwendimann, Comparing expert and novice concept map construction through a talk-aloud protocol, in: A. Cañas, P. Reiska, J. Novak (Eds.), Innovating with Concept Mapping. CMC 2016. Communications in Computer and Information Science, vols. 16–28, Springer, Cham, 2016, p. 635, https://doi.org/10.1007/978-3-319-45501-3_2.