

Learning foreign languages through content and language integrated learning in physical education: A systematic review

VÍCTOR GIL-LÓPEZ

SIXTO GONZÁLEZ-VÍLLORA

University Of Castilla-La Mancha

DAVID HORTIGÜELA-ALCALÁ

University Of Burgos

Received: 11 August 2020 / Accepted: 21 November 2020

DOI: <https://doi.org/10.30827/portalin.v0i35.15785>

ISSN paper edition: 1697-7467, ISSN digital edition: 2695-8244

ABSTRACT: The goal of this systematic review was to describe and assess studies that used content and language integrated learning (CLIL) programmes to teach a foreign language via physical education in a school setting. 35 articles met the selection criteria. Results consisted of two types of studies: (a) Low-intensity CLIL programmes (only the physical education subject was used) and (b) High-intensity CLIL programmes (various academic subjects were combined). No studies were found in the preschool education stage. Most of the research was implemented in Spain. Games and sports were the most frequently used contents, and English was the most commonly used foreign language. High-intensity CLIL programmes improved student's overall proficiency in a foreign language at a higher level. Finally, it is not clear if physical education classes conducted using CLIL have a positive or negative effect on students' moderate-to-vigorous physical activity (MVPA) levels.

Keywords: CLIL; Foreign Language; Language Learning; Multilingualism; Physical Education.

Aprendiendo lenguas extranjeras a través del aprendizaje integrado de contenidos y lenguas extranjeras en educación física: Una revisión sistemática

RESUMEN: El objetivo de esta revisión sistemática fue describir y evaluar estudios que utilizaron los programas de aprendizaje integrado de contenidos y lenguas extranjeras (AICLE) para enseñar una lengua extranjera a través de la educación física en el contexto escolar. 35 artículos cumplieron los criterios de selección. Los resultados mostraron dos tipos de estudios: (a) programas AICLE de baja intensidad (en los que solo la asignatura de educación física fue implicada) y (b) programas AICLE de alta intensidad (donde varias materias académicas fueron combinadas). No se encontraron estudios en la etapa de educación infantil. La mayor parte de las investigaciones se implementaron en España. Los juegos y deportes fueron los contenidos más utilizados, y el inglés la lengua extranjera más usada. Los programas AICLE de alta intensidad mejoraron el dominio general del estudiante en una lengua extranjera en mayor medida. Finalmente, no está claro si la educación física desarrollada con AICLE tiene un efecto positivo o negativo en los niveles de actividad física moderada-vigorosa de los estudiantes.

Palabras clave: AICLE; Lengua Extranjera; Aprendizaje de Lenguas; Multilingüismo; Educación Física.

1. INTRODUCTION

Learning both a foreign language (FL) and the contents of a specific academic subject at the same time is a framework that has become increasingly common in the education field. There are multiple methodological programmes to choose from: content based instruction, bilingual programmes, language X as a medium of instruction, game-based projects, language immersion programs, etc. (Merino & Lasagabaster, 2015; Pérez-Cañado, 2012). However, content and language integrated learning (CLIL) is the most favoured approach in Europe (Coral & Lleixà, 2016; Salvador-García et al., 2017).

CLIL has been described as a pedagogical approach that focuses on two goals: learning the academic subject content and learning the FL, which represents the medium of instruction for the content (Coyle et al., 2010). Philip Hood and David Marsh and drawing on their experience of CLIL in secondary schools, primary schools and English language schools across Europe, this book gives a comprehensive overview of CLIL. It summarises the theory which underpins the teaching of a content subject through another language and discusses its practical application, outlining the key directions for the development of research and practice. This book acknowledges the uncertainty many teachers feel about CLIL, because of the requirement for both language and subject knowledge, while providing theoretical and practical routes towards successful practice for all. --Provided by publisher. A comprehensive and up-to-date overview of the theory and practice of CLIL. This will be of use to practitioners (for example teachers and course developers). The CLIL approach was designed to help improve FL competence without having any pernicious effect on the students' L1 or the content learning (Lasagabaster & Ruiz de Zarobe, 2010). CLIL can be easily integrated through different academic subjects (Merino & Lasagabaster, 2017). However, Physical education (PE) is considered ideal, as it promotes learning in a playful and interactive manner (Chiva-Bartoll & Salvador-García, 2016).

Many researchers link the benefits of CLIL in PE to the communication and interaction that is provided through movement and play, creating an effective platform for learning a FL (Coral & Lleixà, 2013; Coral & Lleixà, 2016; Zurita-Ortega et al., 2019). Furthermore, PE via CLIL has been recognised as a comprehensive approach that embodies the principles of learning, teaching motor skills via a FL, and fostering cognition and citizenship. It also takes into consideration students' motivation for physical activity and provides support to develop both motor and language skills (Coral & Lleixà, 2016). PE and sports can also have a positive influence on the physical, cognitive, and social domains of a child's development, and in his/her lifestyle (World Health Organization, 2010), which, in turn, can be beneficial for FL learning.

A systematic review on learning a FL through PE has been conducted (Salvador-García et al., 2017). However, that review was more general, due to the fact that the study incorporated different teaching approaches in the learning of a FL through PE and it only concluded that CLIL was the most commonly used approach for teaching academic subjects through a FL in Europe (Coral & Lleixà, 2016; Salvador-García et al., 2017). Furthermore, the number and diversity of studies on CLIL and PE have increased exponentially since 2017 (25 studies), which calls for a recent and more extensive analysis (i.e., PE contents, outcomes). To promote this framework (CLIL + PE) and bring it closer to teachers and researchers, this systematic review aimed to describe and assess studies that used CLIL programmes to teach a FL via PE in a school setting.

2. METHOD

2.1. Search limits

A systematic search through six electronic databases (Web of Science, Pubmed, Scopus, SportDiscus-EBSCO, ERIC and Google Scholar) from January 2007 to April 2020 was conducted based on PRISMA Protocol for systematic reviews of the Cochrane Collaboration (Moher et al., 2015). Therefore, the starting date for the systematic search was chosen because Rottmann (2007) linked CLIL and PE to FL learning for the first time during the year of 2007 and further research has been conducted since then. Furthermore, these databases were selected because they included PE articles developed in the school context and published in peer-reviewed journals. The protocol for the systematic review was registered on PROSPERO (CRD42019126972) and is published elsewhere.

The search strategies used included a combination of the following keywords classified in four categories: (a) PE (physical activity, exercise, sports and PE), (b) CLIL, (c) FL (multilingual, bilingual, language teaching, language learning and FL), and (d) population (child, adolescent, pupil, youth and student). Additionally, the English Boolean data types 'and' and 'or' were used.

2.2. Selection criteria

All relevant articles included in this systematic review met the following criteria: (a) studies published in peer-reviewed journals, as these types of records have already been vetted by publishing bodies and experts in the field for their quality and relevance, (b) the intervention study had to promote and assess the learning or the teaching of a FL through CLIL, (c) studies that included qualitative and/or quantitative methods and findings, (d) research connected to school contexts (e.g. PE lessons, activity breaks or after-school programmes), (e) participants were students between three and 18 years of age and teachers, (f) studies which included CLIL programmes developed in an educational setting, and (g) studies published in English or Spanish, due to resource constraints. Duplicated documents, opinion articles, books, conference articles or thesis were eliminated at the first level of exclusion. Studies that did not meet the abovementioned criteria were excluded at the second level. Moreover, the evaluation of the methodological quality of the systematic review was performed using the 11-items checklist elaborated by AMSTAR, a measurement tool to assess the methodological quality of systematic reviews (Shea et al., 2007).

The inclusion criteria were first applied during a double-screening process; during which two reviewers independently screened each title and abstract and recorded the primary reason for rejection, if any. Inter-rater reliability for all screened records was then assessed via the Kappa statistic at 0.83, exceeding levels expected by chance. Disagreements regarding application of the inclusion criteria were resolved through co-author consultation. The full text of records retained after screening was then assessed for eligibility according to the inclusion criteria. Reasons for rejecting studies at this stage are documented in Figure 1.

To increase sensitivity and identify any additional relevant material, the bibliographies of all eligible records were examined through a hand-searching process. Unlike the original

database search, which was limited to peer-reviewed, scholarly records; the hand-search also considered working papers and report citations. Given the acknowledgment of these records within peer-reviewed, scholarly literature, as well as their clear relevance to our research question, excluding these grey literature studies may otherwise have been considered a methodological weakness. Records identified through hand-searching underwent the same screening and full text assessment process as records identified through database searching. The above-cited Kappa statistic for inter-rater reliability includes both database and hand-search records.

The systematic search process and the number of results in each database are shown in Figure 1. During the selection process, the database search found a total of 6,080 articles (6,288 with duplicates). Subsequent to the elimination of many works at the first level of exclusion, 1437 original, potentially relevant articles remained. After assessing the title, abstract, introduction and/or context, 1402 were discarded in the second level of exclusion. Finally, 35 articles were included in the final review. A narrative review method was employed to synthesize these studies. Data extraction was performed systematically by two authors to create two synthesis tables (table 2 and table 3) with comparable information related to the risk of bias of the Cochrane Collaboration protocol (CCP), the focus and assessed variables, the stage of education and age, the country where the research was developed, the number of schools involved, the sample size, the duration of the intervention, the PE contents involved, the FL used, the type of analysis (quantitative or qualitative), the measurement instruments and the outcome(s).

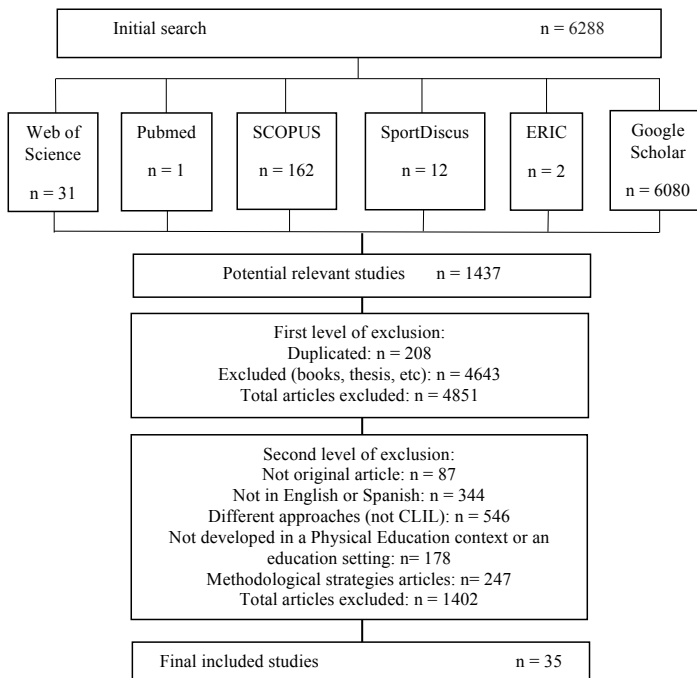


Figure 1. Flow diagram of the systematic search process

2.3. Data extraction and reliability

The information required by the reporting guideline STROBE (O'Brien et al., 2014; Von Elm et al., 2007) was included in the original studies of the extracted data: (a) bibliographic information (author, title, year and form of publication), (b) description of the study sample (age, number, sex and grade of the group of participants), (c) theoretical framework and aim of the study, (d) task description of the intervention and control group and intervention type (physical and theoretical components), (e) intervention name, risk of bias, didactical method, duration and frequency, (f) assessed variables, used measurement instruments and (g) study results. This data was analysed and described in order to obtain relevant information.

Based on the seven criteria of the Cochrane Collaboration protocol (Table 1), this systematic review assessed the risk of bias, which is defined as the risk of an occurring systematic error in study results or inferences (e.g. the study does not specify whether students are randomly assigned to an intervention or to a control group). This can lead to either under- or overestimation of the intervention effects. An assessment of the risk of bias can help to interpret result-variation and balance assumptions and conclusions drawn from these variations (Higgins et al., 2011). This protocol was designed in the field of health and health care research. However, different studies have proved the appropriateness and usefulness of this protocol in the social science field, concluding that there are certain types of observational studies on the effects of non-healthcare interventions for which there is in fact close agreement with the corresponding randomised controlled trials (Konnerup & Kongsted, 2012; Cook et al., 2008). In the process of judging the risk of bias, the study was assessed by two reviewers and discrepancies were resolved by agreement, citing relevant text passages, and offering a justified judgement. Studies were noted to have a low risk of bias if the requested criteria did not show any indication of bias. If there was insufficient information given about the criteria, it was described as an unclear risk of bias. High risk of bias included studies showing highly biased decisions in one or more of the domains considered and was thought to seriously influence the results of the study. Within the category "blinding of participants and personnel", it was considered positive and still a low risk of bias if teachers were not blinded.

Table 1. Criteria of risk of Bias of the Cochrane Collaboration protocol

Bias domain	Source of bias	Support for judgment
Selection bias: Systematic differences between baseline characteristics of the groups that are compared	Random Sequence Generation (RSG)	Description of the detailed method to generate the allocation sequence to allow an assessment of whether it should produce comparable groups. Judgment according to the adequacy with which the allocation sequence was generated
	Allocation concealment (AC)	Description of the detailed method to conceal the allocation sequence to determine whether intervention allocations could have been foreseen in advance of, or during enrolment. Judgement according to whether the allocation was concealed adequately

Performance bias (PB): Systematic differences between groups	Blinding of participants and personnel	Description of measures used to blind participants and personnel from knowledge of which intervention a participant received information on whether the intended blinding was effective
Detection bias (DB): Systematic differences between groups in how outcomes are determined	Blinding of outcome assessors	Description of measures used, to blind the outcome assessors from knowing to which intervention group a participant belonged. Judgement according to whether knowledge of the intervention means allocated was hidden from assessors
Attrition bias (AB): Systematic differences between groups in withdrawals from a study	Incomplete outcome data	Description of the completeness of outcome results for each main outcome, including attrition and exclusions from the analysis. Report of attrition and exclusions in each intervention group and description of reasons for attrition/exclusions
Reporting bias (RB): Systematic differences between reported and unreported findings	Selective outcome reporting	Description of how the possibility of selective outcome reporting was examined, and what was found. Judgement according to whether hints in selective outcome reporting were found
Other sources of bias (OSB)	Anything else, ideally prespecified	Concerns about bias that were not addressed in one of the other domains

3. RESULTS

Overall, 35 studies were included in the systematic review. Two types of studies were found in the systematic search process and are grouped in two different tables. Table 2 shows Low-CLIL interventions (low-intensity CLIL programmes; only the physical education subject was used). Table 3 groups High-CLIL interventions (high-intensity CLIL programmes; various academic subjects were combined: maths, natural sciences, social sciences (history and/or geography), arts and PE).

Table 2. Framework conditions of Low-CLIL interventions

Author(s)	Risk of bias (CCP)	Focus and assessed variables	SoE / age / country	Schools / sample size / duration	PE content / FL used	Analysis / measurement instruments	Outcome(s)
Baena-Extremera et al. (2018)	Low: RSG, AC, PB, DB, AB and RB. Unclear: OSB. High: RSG, AC, DB and OSB	Student's satisfaction in PE using CLIL	ScE / ages 13- 18 / Spain	Different schools / n=758 / one year	Not reported / English	QUANT: questionnaires	Autonomy is the best predictor of intrinsic motivation, and that this is the best predictor of satisfaction in PE using a FL through CLIL
Chiva-Bartoll et al. (2015)	Low: PB and RB. Unclear: OSB. High: RSG, AC, DB and AB	Possibilities of applying CLIL in PE	ScE / ages 14- 15 / Spain	One school / n=one group / six lessons	Games and sports: judo / English	QUANT: test, evaluation criteria, rubric and written composition	CLIL, encouraged motivation, participation and learning
Coral & Lleixà (2014)	Low: PB, DB, AB and RB. Unclear: OSB. High: RSG and AC	Create didactic approaches to solve dilemmas in PE through CLIL	PrE / ages 10- 11 / Spain	One school / n=27 / one year	Motor games / English	QUAL: teacher's diary, video recording and observation	Teachers should combine engaging and effective learning tasks
Coral & Lleixà (2013)	Low: PB, DB, AB and RB. Unclear: OSB. High: RSG and AC	Features of tasks which are most suitable to PE via CLIL	PrE / ages 10- 11 / Spain	One school / n=26 / not reported	Not reported / English	QUAL: teacher's diary	Successfulness when combining cooperative learning and oral skills with motor activity
Coral & Lleixà (2016)	Low: PB, DB, AB and RB. Unclear: OSB. High: RSG and AC	Teaching strategies to improve oral skills	PrE / ages 10- 11 / Spain	One school / n=26 / six months	Games and body language / English	QUAL: interviews, teacher's diary and video recording	Effectiveness of teaching strategies with a concrete purpose in the game
Coral et al. (2017)	Low: RSG, AC, PB, DB, AB and RB. Unclear: OSB	Measure engaged time and language learning physical activity	PrE and ScE / ages 6-16 / Spain	Different schools / n=15 teachers / two years	All contents involved / English	Mixed – QUAL: video recording QUANT: tests	Motor-engaged time was low (support materials)

Escobosa et al. (2019)	Low: RSG, AC, PB, DB, AB and RB. Unclear: OSB	Web-App to teach PE in CLJL	PrE / ages 6-12 / Spain	Different schools / n=15 teachers / not reported	Not reported / English	QUAL: focus group and interviews	Platform where users can create, share and explore PE lesson plans
Fernández-Barrionuevo & Baena-Extremera (2018)	Low: RSG, AC, PB, DB, AB, RB and OSB	Differences between gender	ScE / ages 12- 16 / Spain	One school / n=150 / not reported	Not reported / English	QUANT: different questionnaires	Significant differences between gender
García-Calvo (2019)	Low: RSG, AC, PB, DB, AB and RB. Unclear: OSB	Advantages and disadvantages of using translation	PrE / ages 11- 12 / Spain	One school / n=one group / not reported	Not reported / English	QUAL: teacher's diary and observation	Meaningfulness of translation regarding non-linguistic disciplines
García-Calvo et al. (2019)	Low: RSG, AC, PB, DB, AB and RB. Unclear: OSB	Study students' satisfaction levels	PrE / ages 10- 12 / Spain	Different schools / n=350 / one year	Not reported / English	QUANT: questionnaires	Students' satisfaction levels are related with fun.
Gil-López et al. (2019)	Low: RSG, AC, PB, DB, AB and RB. Unclear: OSB	CLJL vs traditional approaches	PrE / ages 8-9 / Spain	One school / n=51 / four lessons	Motor games and body language / English	QUANT: test, questionnaires and teacher observation	Improvement of the FL, motivation and reduces study time
Lamb & King (2019)	Low: RSG, AC, PB, DB, AB and RB. Unclear: OSB	Improving Speaking skills	ScE / ages 13- 14 / UK	One school / n=42 / 10 weeks	Handball / French	Mixed – QUAL: interviews QUANT: questionnaires	Improve motivation and reduces anxiety to speak
Martínez-Hita & García-Cantó (2017)	Low: RSG, AC, PB, AB and RB. Unclear: OSB. High: DB	Motor activities in CLJL and non-CLJL schools	PrE / ages 8-9 / Spain	Eight schools / n=8 teachers / one year	Motor games / English	QUANT: time record sheet and video recording	CLJL schools need more time to explain the tasks because a FL is used
Ni Chróinín et al. (2016)	Low: RSG, AC, PB, DB, AB, RB and OSB	Potential and pitfalls of CLJL in PE	PrE / ages 8- 11 / Ireland	Eight schools / n=200 / two years	Modified games / Irish	QUAL: interviews, teacher's diary and voice recording	PE learning was restricted by FL knowledge

Salvador-García et al. (2020a)	Low: AC, PB, DB, AB, RB and OSB. High: RSG	CLIL impact on PE lessons	ScE / ages 15- 17 / Spain	One school / n=8 teachers; n=4 students / n=49 accelerometers / one month	Athletics / English	Mixed – QUAL: interviews. QUANT: questionnaire and accelerometers	MVPA levels are higher when using CLIL (accelerometers). Social relationships remain the same way with or without CLIL
Salvador-García & Chiva-Bartoll (2017)	Low: AC, PB, DB, AB, RB and OSB. High: RSG	PE teacher's perspectives in CLIL	ScE / ages 12- 16 / Spain	Different schools / n=8 teachers / three years	All contents involved / English	QUAL: interviews and voice recording	No significant change in the PE curriculum using CLIL Increase of motivation
Salvador-García et al. (2016)	Low: PB and RB. Unclear: OSB. High: RSG, AC, DB and AB	PE features to develop CLIL through English	PrE / ages 10- 11 / Spain	One school / n=one group / six lessons	Orienteering / English	QUANT: initial questionnaire and rubric	Teachers should consider student's English level and the 4C's framework
Salvador-García et al. (2020b)	Low: AC, PB, DB, AB, RB and OSB. High: RSG	CLIL effects on MVPA levels	ScE / ages 13- 14 / Spain	One school / n=48 / one month	Athletics / English	QUANT: accelerometers	MVPA levels are higher using CLIL
Salvador-García et al. (2019)	Low: RSG, AC, PB, DB, AB and RB. Unclear: OSB	CLIL effects on MVPA levels	ScE / ages 15- 16 / Spain	One school / n=49 / two months-eight lessons	Athletics / English	Mixed – QUAL: interviews QUANT: accelerometers	MVPA levels are higher using CLIL
Salvador-García et al. (2018)	Low: RSG, AC, PB, DB, AB and RB. Unclear: OSB	CLIL in touch-rugby through Teaching Games for Understanding	ScE / ages 15- 17 / Spain	One school / n=20 / seven lessons	Games and sports: Touch-rugby / English	QUAL: teacher's diary and anecdotal record	Apprehension and fear for the FL turned into an increase of participation

Notes: RSG (Random Sequence Generation); AC (Allocation Concealment); PB (Performance Bias); DB (Detection Bias); AB (Attrition Bias); RB (Reporting Bias); OSB (Other Sources of Bias); LOW-CLIL (Low-intensity CLIL programmes); HIGH-CLIL (High-intensity CLIL programmes); SoE (Stage of education); PrE (Primary education); ScE (Secondary education); QUANT (qualitative data collect and analysis methods); QUANT (quantitative data collect and analysis methods).

Table 3. Framework conditions of High-CLIL interventions

Author(s)	Risk of bias (CCP)	Focus and assessed variables	SoE / age / country	Schools / sample size / duration	PE content / FL used	Analysis / measurement instruments	Outcome(s)
Alejo & Piquer-Piriz (2016)	Low: RSG, AC, PB, DB, AB and RB. Unclear: OSB	Influence of urban/rural settings	ScE / ages 14-15 / Spain	Two schools / n=48 / one year	Not reported / English	QUANT: questionnaires and tests	Urban learners had better conditions, were less anxious and motivated than rural learners
Coral et al. (2016)	Low: RSG, AC, PB, DB and RB. Unclear: AB. High: OSB	Analyse the English level of Catalonia, CLIL/non-CLIL schools and PE in CLIL	PrE / ages 6-12 / Spain	(a) 1175 schools, (b) n=11662, (c) n=96 / three years	Motor games / English	QUANT: FL tests marked by Catalan Ministry of Education	Better English level in Catalonia, in CLIL schools and when applying PE in CLIL
Fernández-Río et al. (2017)	Low: RSG, AC, PB, DB, AB and RB. Unclear: OSB	CLIL effects on social skills and goals	ScE / ages 12-16 / Spain	Different schools / n=502 / one year	Not reported / Spanish, English and French	Mixed – QUAL: interviews QUANT: questionnaires	Social goals, responsibilities, and relationships, seem to positively change after a year using CLIL
Forey & Cheung (2019)	Low: RSG, AC, PB, DB, AB and RB. Unclear: OSB	Impact on teachers and students of language teaching	ScE / ages 14-16 / UK	One school / n=two groups and teachers / 18 months	Not reported / English	QUAL: focus group, video recording, interviews and questionnaire	Improvement of students' written assignments and PE exam results
García-Calvo & Salaberrí (2018)	Low: RSG, AC, PB, DB, AB and RB. Unclear: OSB	CLIL vs grammatical translation	PrE / ages 10-12 / Spain	One school / n=39 / one year	Games and sports / English	Mixed – QUAL: interviews QUANT: questionnaires	Higher satisfaction using CLIL than grammatical translation
Hernando et al. (2018)	Low: RSG, AC, PB, DB, AB and RB. Unclear: OSB	PE teachers and coordinators perceptions towards CLIL programs in Spain	ScE / ages 12-18 / Spain	26 schools / n=26 coordinators; n=26 PE teachers / one year	Not reported / English	Mixed – QUAL: interviews QUANT: questionnaires	Using CLIL through English as a FL is not a factor which implies a lower specific PE curricular learning
Ito (2019)	Low: RSG, AC, PB, DB, AB and RB. Unclear: OSB	Effectiveness of CLIL	ScE / ages 13-14 / Japan	One school / n=23 / three lessons	Games and sports: Basketball / English	Mixed – QUAL: interviews QUANT: questionnaires	Improvement of the FL learning and teaching

Martínez-Adrián et al. (2017)	Low: AC, PB, DB, AB and RB. Unclear: OSB, RB. High: RSG	Communication strategies in FL production	PrE / ages 10-12 / Spain	One school / n=146 / one year	Not reported / English	QUANT: questionnaires and proficiency level test	Most used: paraphrasing and appeal. Less used: morphological creativity and miming
Méndez & Pavón (2012)	Low: RSG, AC, PB, DB, AB and RB. Unclear: OSB	Collaborative work among teachers using L1 and L2	Both stages / ages 6-16 / Spain	Two schools / n=14 teachers (n=1 PE teacher) / one year	Not reported / French	QUAL: semi-structured interviews and voice recording	Collaboration and the use of L1+L2 allowed changes in content teaching/learning
Merino & Lasagabaster (2017)	Low: RSG, AC, PB, DB, AB, RB and OSB	Intensity of CLIL and FL overall proficiency	ScE / ages 11-13 / Spain	13 schools / n=393 / one year	Not reported / English	QUANT: tests	Significant FL improvements in high-intensity programmes
Moore & Lorenzo (2015)	Low: RSG, AC, PB, DB, AB, RB and OSB	Provide materials to CLIL teachers	Both stages / ages 6-16 / Spain	Different schools / n=21 PE teachers (n=2 not reported)	Games / English, French and German	QUAL: interviews	Difficulties while integrating FL and content
Pavón (2018)	Low: RSG, AC, PB, DB, AB, RB. Unclear: OSB	Differences between rural/urban schools	Both stages / ages 11-16 / Spain	Different schools / n=295 / one year	Not reported / English	Mixed – QUAL: interviews. QUANT: tests	Differences cannot only be attributed to the school settings
Pavón & Ramos (2019)	Low: PB, DB, AB and RB and OSB. High: RSG, AC	Features and functions of using L1 in bilingual immersion	PrE / ages 10-11 / Spain	One school / n=22 / eight lessons	Not reported / English	QUAL: interviews and video recording	L1 communication strategies helped to learn FL
Pérez-Cañado (2018)	Low: RSG, AC, PB, DB, AB, RB and OSB	CLIL vs non-CLIL students' English linguistic competence	Both stages / ages 11-16 / Spain	53 schools / n=2024 / four years	Not reported / English	QUANT: tests and initial questionnaire	FL improvement in CLIL and non-CLIL with low effect sizes
Ruiz de Zarobe & Zenotz (2018)	Low: RSG, AC, PB, DB, AB and RB. Unclear: OSB	Student's reading competence in CLIL	PrE / ages 11-13 / Spain	One school / n=50 / two years	Not reported / English	QUANT: pre-test and post-test	Benefits in one skill and language can be applied to other languages or skills

Notes: RSG (Random Sequence Generation); AC (Allocation Concealment); PB (Performance Bias); DB (Detection Bias); AB (Attrition Bias); RB (Reporting Bias); OSB (Other Sources of Bias); LOW-CLIL (Low-intensity CLIL programmes); HIGH-CLIL (High-intensity CLIL programmes); SoE (Stage of education); PrE (Primary education); ScE (Secondary education); QUAL (qualitative data collect and analysis methods); QUANT (quantitative data collect and analysis methods).

4. DISCUSSION AND CONCLUSIONS

The aim of this systematic review was to describe and assess studies that used CLIL programmes to teach a FL via PE in a school setting. The number of studies has increased exponentially since 2017 (25 studies, 71.43%).

An analysis of the risk of bias revealed that all the studies had at least two fields with a low risk: performance bias (blinding of participants and personnel) and reporting bias (selective outcome reporting). The criteria selection bias (random sequence generation: 34.48% and allocation concealment: 27.58%) was highly biased in some studies. Most of the studies (30, 85.71%) revealed a low risk of detection bias (blinding of outcome assessment). Attrition bias (incomplete outcome data) revealed a low risk of bias in the majority of the studies (32, 91.43%). The other sources of bias assessed were the lack of certain English skills in the tests used to evaluate participant's overall FL proficiency, brief intervention programmes and small intervention groups. However, most studies (25, 71.42%) reported an unclear risk in other sources of bias.

Results showed two types of studies: (a) Low-CLIL (low-intensity CLIL programmes) and (b) High-CLIL (high-intensity CLIL programmes). Regarding the student's overall proficiency in a FL, Merino & Lasagabaster (2017) concluded that no significant differences emerged between learners that participated in non-CLIL and low-intensity CLIL programmes. However, significant benefits were found among students that participated in high-intensity CLIL programmes through the combination of different academic subjects (maths, natural sciences, social sciences (history and/or geography), arts and PE) (Coral et al., 2016; Merino & Lasagabaster, 2017). Taking into consideration all the academic subjects included within the curriculum, PE is ideal for developing CLIL and FL learning as it provides a comprehensive education and encourages motivation, participation and interactive learning amongst the students through movement and play (Coral & Lleixà, 2013; Coral & Lleixà, 2016; Zurita-Ortega et al., 2019).

Research seems to confirm that CLIL has a higher success rate in developing a FL, both written and oral skills, than traditional methods (García-Calvo & Salaberri, 2018; Gil-López et al., 2019).

It is important to highlight that no studies were found in preschool education. Incorporating a FL during the early stages of childhood education is a positive, feasible and recommended way for students to learn and advance (Coyle et al., 2010). It is also important to integrate FL learning with different school subjects, since the objective is for students to view the FL as something natural (Rodríguez & Varela, 2004).

Games and sports were the most frequently used PE contents when learning a FL through CLIL because this type of content encourages communication amongst students to improve output skills (writing and speaking) (Gil-López et al., 2019). Therefore, the same way Salvador-García et al. (2018) applied the Teaching Games for Understanding model through a lesson plan of Touch-rugby, applying pedagogical models through the incorporation of games and sports is a very powerful and innovative line of research in the learning of a FL through CLIL.

Although the original goal of CLIL was to promote multilingualism, English has proved to be the most commonly used language (Coyle et al., 2010; Merino & Lasagabaster, 2015).

In this systematic review, the majority of the research was implemented in Spain using the English language. It shows the growing importance of learning English as a FL through CLIL in this country (Coral & Lleixà, 2016).

Regarding urban/rural setting, Alejo & Piquer-Piriz (2016) highlighted the advantages that urban learners had over rural learners when learning a FL. Randhawa & Michayluk (1975) concluded that urban settings that offered a better physical environment, were more intellectually stimulating, and met the learners' needs to provide a satisfying learning experience. Nevertheless, Fan & Chen (1998) indicated that students from rural schools performed as well as, if not better than, the students in metropolitan schools in math, science, reading, and social studies. This research reinforces the idea that the differences between rural and urban students cannot only be attributed to the school settings (Pavón, 2018).

Regarding students' physical activity level, results are scarce, but contradictory. Coral et al. (2017) and Martínez-Hita & García-Cantó (2017) found lower MVPA levels than recommended (60 minutes of MVPA daily) to produce positive health effects (Martínez et al., 2012; World Health Organization, 2010), while Salvador-García et al. (2019), Salvador García et al. (2020a) and Salvador García et al. (2020b) concluded that physical activity levels were higher using CLIL. It is important that the use of a FL in PE does not reduce students' MVPA, and more research is needed.

Finally, Ní Chróinín et al. (2016) concluded that students' PE content learning was restricted by their FL knowledge. However, other researchers disagree. Hernando et al. (2018) concluded that using CLIL through English as a FL is not a factor which implies a lower specific PE curricular learning. Housen (2002) alluded to successful learning outcomes in content-subject areas (history, geography, arts, music, religion and economics) that were taught in a FL, and Jabrun (1997) found that after a year, FL immersion maths students outperformed mainstream students, and FL immersion science students performed as well as their mainstream counterparts. This study concluded that FL immersion students were the most efficient learners (Coyle et al., 2010). Therefore, more research is needed to assess students' PE content learning when using CLIL through a FL.

5. LIMITATIONS AND AREAS FOR FUTURE RESEARCH

This systematic review is subject to certain methodological limitations. While it attempted to capture an exhaustive list by using 14 different terms in the systematic search, some relevant records may not have been captured. The database search was also limited to records in English and Spanish, thus potentially excluding important publications in other languages. Finally, publication bias within the literature is well documented and can be a serious source of bias in systematic reviews (Petticrew & Roberts, 2006). This bias generally results in an underreporting of research with null or negative findings.

The magnitude of the impacts found in the studies included in this review were not expressed in such a way as to permit meta-analysis of overall impact. Overall, only some studies (Fernández-Barrionuevo & Baena-Extremera, 2018; Merino & Lasagabaster, 2017) provided information on effect sizes. Most of the studies only reported tendencies without a statistical approach. Therefore, it would be beneficial for researchers to include data on effect sizes for quantitative and mixed research to help provide sufficient information for a future meta-analysis.

With respect to our findings, the quality of some of the included studies is questionable despite the fact that they can be found in bibliographic databases and that they have even gone through a peer review process. Quality research (and evidence-based practice) should be deeply reviewed and analysed.

Lastly, the studies in our review were relatively geographically concentrated, primarily in Western Europe, due to the inclusion criteria and language restrictions. Different countries face unique challenges and the findings from one country may not be generalizable elsewhere. Therefore, more research is required in other continents to empirically deepen the knowledge of how to implement, understand and advance this teaching-learning approach.

6. REFERENCES

- Alejo, R., & Piquer-Píriz, A. (2016). Urban vs. rural CLIL: An analysis of input-related variables, motivation and language attainment. *Language, Culture and Curriculum*, 29(3), 245-262. <https://doi.org/10.1080/07908318.2016.1154068>
- Baena-Extremera, A., & Granero-Gallegos, A. (2015). Versión española del Sport Satisfaction Instrument (SSI) adaptado al aprendizaje de la Educación Física bilingüe en Inglés. *Porta Linguarum*, 24, 63-76.
- Baena-Extremera, A., Granero-Gallegos, A., Baños, R., & Ortiz-Camacho, M. M. (2018). Can physical education contribute to learning English? Structural model from self-determination theory. *Sustainability*, 10(10), 3613-3624. <https://doi.org/10.3390/su10103613>
- Castillo, I., Balaguer, I., & Duda, J. L. (2001). Perspectivas de meta de los adolescentes en el contexto académico. *Psicothema*, 13(1), 79-86.
- Chiva-Bartoll, O., Isidori, E., & Fazio, A. (2015). Bilingual Physical Education and Critical Pedagogy: an application based on Judo. *Retos. Nuevas Tendencias en Educación Física, Deporte y Recreación*, 28, 110-115.
- Chiva-Bartoll, O., & Salvador-García, C. (2016). *Aprendizaje integrado de educación física y lengua inglesa. Aplicación crítico-deliberativa del método AICLE*. INDE.
- Cook, T., Shadish, W., & Wong, V. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management*, 27(4), 724-750. <https://doi.org/10.1002/pam.20375>
- Coral, J., & Lleixà, T. (2013). Physical Education-in-CLIL tasks. Determining tasks characteristics through the analysis of the diaries. *Retos. Nuevas Tendencias en Educación Física, Deporte y Recreación*, 24, 79-84.
- Coral, J., & Lleixà, T. (2014). Teaching Physical Education under Clil: solving professional dilemmas encountered when conducting action research. *Movimento*, 20(4), 1447-1472.
- Coral, J., & Lleixà, T. (2016). Physical education in content and language integrated learning: Successful interaction between physical education and English as a foreign language. *International Journal of Bilingual Education and Bilingualism*, 19(1), 108-126. <https://doi.org/10.1080/13670050.2014.977766>
- Coral, J., Lleixà, T., & Ventura, C. (2016). Foreign language competence and content and language integrated learning in multilingual schools in Catalonia: an ex post facto study analysing the results of state key competences testing. *International Journal of Bilingual Education*

- and Bilingualism*, 21(2), 139-150. <https://doi.org/10.1080/13670050.2016.1143445>
- Coral, J., Urbiola, M., Sabaté, E., Bofill, J., Lleixà, T., & Vilà, R. (2017). Does the Teaching of Physical Education in a Foreign Language Jeopardise Children's Physical Activity Time? A Pilot Study. *International Journal of Bilingual Education and Bilingualism*, 1-18. <https://doi.org/10.1080/13670050.2017.1407289>
- Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL Content and Language Integrated Learning*. Cambridge University Press.
- Escobosa, G., Lleixà, T., & Coral, J. (2019). Design of the prototype of a Physical Education (PE) web-app in Content and Language Integrated Learning (CLIL). *Journal of Sport and Health Research*, 11, 1-16.
- Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American journal of theoretical and applied statistics*, 5(1), 1-4. <https://doi.org/10.11648/j.ajtas.20160501.11>
- Fan, X., & Chen, M. J. (1998). Academic achievement of rural school students: A multi-year comparison with their peers in suburban and urban schools. *Journal of Research in Rural Education*, 15(1), 31-46.
- Forey, G., & Cheung, L. M. E. (2019). The benefits of explicit teaching of language for curriculum learning in the physical education classroom. *English for Specific Purposes*, 54, 91-109. <https://doi.org/10.1016/j.esp.2019.01.001>
- Fernández-Barrionuevo, E., & Baena-Extremera, A. (2018). Motivation in Physical Education and Foreign Language Learning in CLIL Teaching: Gender Differences and Implications for Future Studies. *Porta Linguarum*, 30, 207-220.
- Fernández-Río, J., Hortigüela, D., & Pérez-Pueyo, A. (2017). Bilingual education, streaming and secondary education students' social goals. *Revista de Educación*, 377, 30-53. <https://doi.org/10.4438/1988-592X-RE-2017-377-352>
- García-Calvo, S. (2019). Can translation be useful in a second language teaching? A proposal for using translation in bilingual physical education through content and language integrated learning. *European Journal of Foreign Language Teaching*, 4(2), 113-125.
- García-Calvo, S., & Salaberri, M. S. (2018). Physical Education and English: constructivist vs traditional approaches in bilingual and non-bilingual education. *Sportis. Scientific Technical Journal of School Sport, Physical Education and Psychomotricity*, 4(2), 364-387. <https://doi.org/10.17979/sportis.2018.4.2.3330>
- García-Calvo, S., Banegas, D. L., & Salaberri, M. S. (2019). Satisfaction's study in Bilingual Physical Education after applying a schedule based on content and language integrated learning. *Sportis. Scientific Technical Journal of School Sport, Physical Education and Psychomotricity*, 5(2), 305-322. <https://doi.org/10.17979/sportis.2019.5.2.5235>
- Gil-López, V., Sierra-Díaz, M. J., & González-Villora, S. (2019). The effects of Content and Language Integrated Learning in the Physical Education subject in Primary Education. *Sportis. Scientific Technical Journal of School Sport, Physical Education and Psychomotricity*, 5(3), 562-585. <https://doi.org/10.17979/sportis.2019.5.3.5527>
- Granero-Gallegos, A., & Baena-Extremera, A. (2013). Análisis preliminar exploratorio del "Sport Motivation Scale" adaptado a la Educación Física. *Espiral. Cuadernos del profesorado*, 6(12), 3-14.
- Hernando, A., Hortigüela, D., & Pérez-Pueyo, A. (2018). Percepción de coordinadores de programas bilingües y docentes de Educación Física en inglés en secundaria sobre el proceso

- de implantación y desarrollo del bilingüismo en la Comunidad de Castilla y León. *Retos: nuevas tendencias en educación física, deporte y recreación*, 33, 63-68.
- Higgins, J. P. T., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., ... & Sterne, J. A. C. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, 343(7829), 1-9. <https://doi.org/10.1136/bmj.d5928>
- Housen, A. (2002). Processes and outcomes in the European schools model of multilingual education. *Bilingual Research Journal*, 26(1), 45-64. <https://doi.org/10.1080/15235882.2002.10668698>
- Ito, Y. (2019). The Effectiveness of a CLIL Basketball Lesson: A Case Study of Japanese Junior High School CLIL. *English Language Teaching*, 12(11), 42-54.
- Jabrun, P. (1997). Academic achievement in late partial immersion French. *Babel*, 32(2), 20-23.
- Konnerup, M., & Kongsted, H. C. (2012). Do Cochrane reviews provide a good model for social science? The role of observational studies in systematic reviews. *Evidence & Policy: A Journal of Research, Debate and Practice*, 8(1), 79-96. <https://doi.org/10.1332/174426412X620146>
- Lamb, P., & King, G. (2019). Another platform and a changed context: Student experiences of developing spontaneous speaking in French through physical education. *European Physical Education Review*, 1-20. <https://doi.org/10.1177/1356336X19869733>
- Lasagabaster, D., & Ruiz de Zarobe, Y. (Eds.). (2010). *CLIL in Spain: Implementation, results and teacher training*. Cambridge Scholars Publishing.
- Martínez-Adrián, M., Gallardo-del-Puerto, F., & Basterrechea, M. (2017). On self-reported use of communication strategies by CLIL learners in primary education. *Language Teaching Research*, 23(1), 39-57. <https://doi.org/10.1177/1362168817722054>
- Martínez-Hita, F. J., & García-Cantó, E. (2017). Influence of bilingualism on engagement motor time in Physical Education. *Retos. Nuevas Tendencias en Educación Física, Deporte y Recreación*, 32, 178-182.
- Martínez, J., Contreras, O. R., Aznar, S., & Lera, Á. (2012). Niveles de actividad física medido con acelerómetro en alumnos de 3º ciclo de Educación Primaria: actividad física diaria y sesiones de Educación Física. *Revista de Psicología del Deporte*, 21(1), 117-123.
- Méndez, M. C., & Pavón, V. (2012). Investigating the coexistence of the mother tongue and the foreign language through teacher collaboration in CLIL contexts: perceptions and practice of the teachers involved in the plurilingual programme in Andalusia. *International Journal of Bilingual Education and Bilingualism*, 15(5), 573-592. <https://doi.org/10.1080/13670050.2012.670195>
- Merino, J. A., & Lasagabaster, D. (2015). CLIL as a way to multilingualism. *International Journal of Bilingual Education and Bilingualism*, 21(1), 79-92. <https://doi.org/10.1080/13670050.2015.1128386>
- Merino, J. A., & Lasagabaster, D. (2017). The effect of content and language integrated learning programmes' intensity on English proficiency: A longitudinal study. *International Journal of Applied Linguistics*, 28(1), 18-30. <https://doi.org/10.1111/ijal.12177>
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., ... & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic reviews*, 4(1), 1-9. <https://doi.org/10.1186/2046-4053-4-1>
- Moore, P., & Lorenzo, F. (2015). Task-based learning and content and language integrated learning materials design: process and product. *The Language Learning Journal*, 43(3), 334-357. <https://doi.org/10.1080/09571736.2015.1053282>
- Ní Chróinín, D., Ní Mhurchú, S., & Ó Ceallaigh, T. J. (2016). Off-balance: the integration of physical education content learning and Irish language learning in English-medium primary

- schools in Ireland. *Education 3-13 International Journal of Primary Elementary and Early Years Education*, 44(5), 566-576. <https://doi.org/10.1080/03004279.2016.1170404>
- O'Brien, B. C., Harris, I. B., Beckman, T. J., Reed, D. A., & Cook, D. A. (2014). Standards for reporting qualitative research: a synthesis of recommendations. *Acad Med*, 89(9), 1245-1251. <https://doi.org/10.1097/acm.0000000000000388>
- Pavón, V. (2018). Learning Outcomes in CLIL Programmes: a Comparison of Results between Urban and Rural Environments. *Porta Linguarum*, 29, 9-28.
- Pavón, V., & Ramos, M. C. (2019). Describing the use of the L1 in CLIL: an analysis of L1 communication strategies in classroom interaction. *International Journal of Bilingual Education and Bilingualism*, 22(1), 35-48. <https://doi.org/10.1080/13670050.2018.1511681>
- Pérez-Cañado, M. L. (2018). CLIL and Educational Level: A Longitudinal Study on the Impact of CLIL on Language Outcomes. *Porta Linguarum*, 29, 51-70
- Pérez-Cañado, M. L. (2012). CLIL research in Europe: Past, present, and future. *International Journal of Bilingual Education and Bilingualism*, 15(3), 315-341. <https://doi.org/10.1080/13670050.2011.630064>
- Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Blackwell Publishing. <https://doi.org/10.1002/9780470754887>
- Randhawa, B. S., & Michayluk, J. O. (1975). Learning environment in rural and urban classrooms. *American Educational Research Journal*, 12(3), 265-279. <https://doi.org/10.3102/00028312012003265>
- Rodríguez, B., & Varela, R. (2004). Models of teaching foreign languages to young children. *Didáctica (Lengua y Literatura)*, 16(1), 163-175.
- Rottmann, B. (2007). Sports in English. Learning Opportunities through CLIL in Physical Education. In C. Dalton-Puffer and U. Smit (Eds.), *Empirical Perspectives on CLIL Discourse* (pp. 205-227). Peter Lang.
- Ruiz de Zarobe, Y., & Zenotz, V. (2018). Learning strategies in CLIL classrooms: how does strategy instruction affect reading competence over time? *International Journal of Bilingual Education and Bilingualism*, 21(3), 319-331. <https://doi.org/10.1080/13670050.2017.1391745>
- Salvador-García, C., Capella-Peris, C., Chiva-Bartoll, O., & Ruiz-Montero, P. J. (2020a). A Mixed Methods Study to Examine the Influence of CLIL on Physical Education Lessons: Analysis of Social Interactions and Physical Activity Levels. *Frontiers in Psychology*, 11(578), 1-10. <https://doi.org/10.3389/fpsyg.2020.00578>
- Salvador-García, C., & Chiva-Bartoll, O. (2017). CLIL in teaching physical education: views of the teachers in the Spanish context. *Journal of Physical Education and Sport*, 17(3), 1130 - 1138. <http://dx.doi.org/10.7752/jpes.2017.03174>
- Salvador-García, C., Chiva-Bartoll, O., & Capella-Peris, C. (2019). Bilingual physical education: the effects of CLIL on physical activity levels. *International Journal of Bilingual Education and Bilingualism*, 1-11. <https://doi.org/10.1080/13670050.2019.1639131>
- Salvador-García, C., Chiva-Bartoll, O., & Colomer-Diago, C. (2020b). The effect of bilingual Physical Education on students' Physical Activity. Things are not always as they seem. *Cultura, Ciencia y Deporte*, 15(43), 53-61.
- Salvador-García, C., Chiva-Bartoll, O., & Fazio, A. (2016). Features of Integrated Learning Contents of Physical Education and Foreign Language. *Retos. Nuevas Tendencias en Educación Física, Deporte y Recreación*, 29, 120-125.

- Salvador-García, C., Chiva-Bartoll, O., & Isidori, E. (2017). Learning a foreign language through physical education: A systematic review. *Movimento*, 23(2), 647-660.
- Salvador-García, C., Chiva-Bartoll, O., & Vergaz-Gallego, J. J. (2018). Perception of students on the use of CLIL method in Physical Education: A Case Study. *Retos. Nuevas Tendencias en Educación Física, Deporte y Recreación*, 33, 138-142.
- Shea, B. J., Grimshaw, J. M., Wells, G. A., Boers, M., Andersson, N., Hamel, C., ... & Bouter, L. M. (2007). Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC medical research methodology*, 7(10), 1-17. <https://doi.org/10.1186/1471-2288-7-10>
- Von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med*, 147(8), 573-577. <https://doi.org/10.7326/0003-4819-147-8-200710160-00010>
- World Health Organization. (2010). *Global recommendations on physical activity for health*. World Health Organization.
- Zurita-Ortega, F., Ortega-Martín, J. L., & Castro-Sánchez, M. (2019). Physical Activity and Self-concept in Bilingual and Non-bilingual Schools: An Analysis through Structural Equations. *Porta Linguarum*, 32, 23-35.