



МЕЖДУНАРОДНЫЙ ОПЫТ ИНТЕГРАЦИИ ОБРАЗОВАНИЯ / INTERNATIONAL EXPERIENCE IN THE INTEGRATION OF EDUCATION

УДК 37.043.2-055.-1

DOI: 10.15507/1991-9468.103.025.202102.192-213



Original article

A Review of European Research on Content and Language Integrated Learning

*I. Cimermanová**University of Presov, Presov, Slovakia,
ivana.cimermanova@unipo.sk*

Introduction. Subjects' integration in formal teaching can play an important role in addressing the issue of authentic and meaningful learning as opposed to rote memorisation. Content and language integrated learning has been the subject of educational studies for three decades. The scope of research is broad, and from the primary focus on foreign language performance, it has been slowly extended to the impact of Content and language integrated learning on content and mother tongue. The purpose of the research is to summarize selected research articles on Content and language integrated learning application and to estimate its summary average effect on content development in a group of students aged 10–16.

Materials and Methods. The article presents the systematic review of the studies published in the Web of Science database in the last decade (2010–2020) and surveys the selected empirical studies that focus on the impact of Content and language integrated learning implementation on the content subjects at primary and secondary schools. Sixteen studies met the inclusion criteria and were included in the analysis. Data from six studies were also statistically evaluated using Comprehensive meta analysis and RevMan software, and the synthesis is presented in the Results and Discussion parts.

Results. Based on the 16 discussed studies' results, Content and language integrated learning intervention produces positive added value; however, the statistical meta-analysis showed no statistically significant differences between the Content and language integrated learning and non-CLIL groups in their content knowledge and the results favouring non-CLIL groups. As the groups' size differed in terms of absolute value, the pooled standard deviation was used to reflect the sample sizes and standard deviation were averaged with more weight given to the larger sample groups.

Discussion and Conclusion. The practical significance and prospects of the study lie in pointing out the benefits of Content and language integrated learning and stressing the importance of its inclusion in teacher training study programmes along with the development of pre-service teachers' creativity, critical thinking and ability to create their materials.

Keywords: Content and language integrated learning, content subject, systematic review, meta-analysis, experiment

Funding: This study is a part of a research project financed by the Slovak Ministry of Education, Science, Research and Sport (KEGA 032PU-4/2019, project: Creating teaching materials for content and language integrated learning at elementary schools).

Acknowledgements: I am very grateful to the anonymous reviewers whose comments and suggestions helped improve and clarify this manuscript.

The author declares no conflict of interest.

© Cimermanová I., 2021



Контент доступен под лицензией Creative Commons Attribution 4.0 License.
The content is available under Creative Commons Attribution 4.0 License.

Оригинальная статья

Обзор европейских исследований по предметно-языковому интегрированному обучению

И. Цимерманова

*Прешовский университет, г. Прешов, Словакия,
ivana.cimermanova@unipo.sk*

Введение. Интеграция предметов в формальное обучение выполняет важную роль в решении проблемы аутентичного и осмысленного обучения в отличие от механического запоминания. В течение трех десятилетий предметно-языковое интегрированное обучение (CLIL) было объектом образовательных исследований. От основного внимания к уровню владения иностранным языком он постепенно расширился до влияния данного метода на содержание и родной язык. Цель настоящей работы – обобщить отобранные исследовательские статьи по применению предметно-языкового интегрированного обучения и оценить его суммарный средний эффект на развитие контента в группе учащихся в возрасте 10–16 лет.

Материалы и методы. В статье представлен систематический обзор исследований, опубликованных в базе данных Web of Science за последнее десятилетие (2010–2020), проанализированы отдельные эмпирические исследования, посвященные влиянию внедрения предметно-языкового интегрированного обучения на теоретические курсы (предметы) в начальных и средних школах.

Результаты исследования. Проанализированы 16 исследований, соответствующие критериям включения, 6 исследований были статистически оценены с использованием комплексного мета-анализа и программного обеспечения RevMan. Основываясь на результатах 16 обсуждаемых исследований, было доказано, что применение данного метода дает положительную практическую пользу. Однако статистический мета-анализ не обнаружил статистически значимых различий между группами, использующими интегрированные уроки и не применяющими их. Поскольку размеры групп различались по абсолютной величине, объединенное стандартное отклонение использовалось для отражения размеров выборки.

Обсуждение и заключение. Практическая значимость и перспективы исследования демонстрируют преимущества предметно-языкового интегрированного обучения и важность его включения в учебные программы подготовки учителей наряду с развитием творческого потенциала, критического мышления и способности создавать свои материалы.

Ключевые слова: предметно-языковое интегрированное обучение, теоретический предмет, систематический обзор, мета-анализ, эксперимент

Финансирование: данная статья поддержана проектом, финансируемым Министерством образования, науки, исследований и спорта Словакии (KEGA 032PU-4/2019, проект «Создание учебных материалов для предметного и языкового интегрированного обучения в начальных школах»).

Благодарности: автор выражает благодарность анонимным рецензентам за ценные замечания, которые позволили улучшить качество статьи.

Автор заявляет об отсутствии конфликта интересов.

Для цитирования: Цимерманова, И. Обзор европейских исследований по предметно-языковому интегрированному обучению / И. Цимерманова. – DOI 10.15507/1991-9468.103.025.202102.192-213 // Интеграция образования. – 2021. – Т. 25, № 2. – С. 192–213.

Introduction

Communication in a foreign language(s) belongs to the key competencies for life-long learning. Communication involves not only

skill(s) but also intercultural understanding. Foreign languages have been learnt and taught for centuries. The research into the history of English as a foreign language



teaching maps describes different methods and approaches. The methods have reflected the current knowledge and the theories as well as learners' needs. Content language integrated learning (CLIL) is a methodology when a foreign language functions as a means of communication rather than the aim of teaching, and language learning becomes meaningful and leads toward better long-term retention. It is based on dual aims (content / subject discipline aims and language teaching aims), and its main principles are often presented by the abbreviation 4Cs¹ – content, communication, cognition and culture, or extended 5C, which includes the development of competence². The competence in this context means that teachers should consider “can-do” statements³ they want their students to be able to do by the end of the lesson. The framework 4C+1 was suggested by Lynch [1] addressing also the theory presented by Coyle, Hood and Marsh⁴. The term CLIL came into existence in 1994 [2]; however, it has a much longer history and the signs of the CLIL principles have been present in different approaches; e. g. integrated thematic instruction (school model designed by Susan Kovalik⁵), immersion [3], content-based instruction [4], task-based language teaching [5], English for specific purpose [6–10] or bilingual education (see, e.g. [11]). Marsh refers to CLIL as to a “generic term” and describes it as an “educational approach in which diverse methodologies are used which lead to dual-focussed education where attention is given to both topic and language of instruction that includes a wide range of approaches” [12, p. 233]. This is one of the reasons why the studies on CLIL might be difficult or even inappropriate to compare. The studies where the applied approach is defined as CLIL may differ in the perception of what CLIL is, and it happens

that more rigorous researchers would evaluate or define it as immersion or perhaps CBI (content-based instruction), see [13; 14].

CLIL has substantially increased across Europe [15], and it is also reflected in the research. The earliest studies have primarily examined the impact of CLIL on foreign-language performance. Later, the focus was also shifted to the content learning impact and different aspects that may influence the results of CLIL application on language and content learning. The aspects that influence CLIL results, e.g. classroom interaction [16], strategies [17; 18], age [19], the influence of affective factors [20; 21], motivation [22], gender [22; 23], household structure [24], socioeconomic status [25; 26], became the subject of more in-depth focus. The experimental data are rather positive [27–31], especially concerning the positive impact of CLIL on language gains, motivation increase and positive change of attitudes towards foreign language learning; however, it is a more complex issue, and as such, the impact has to be studied and interpreted from a much broader context. Pérez Cañado studied the effects of CLIL on mother tongue (L1) and content learning, and she stresses that the CLIL students who receive instruction in L1 outperform their peers, especially in the long term, and she suggests that increased time and input are crucial for CLIL students “to achieve either the same or superior content results as their monolingual peers” [27].

In their study, Castellano-Risco, Alejo-González & Piquer-Píriz reported that CLIL students nearly doubled regular EFL learners' receptive vocabulary knowledge in their research sample [28]. The study presents the experiment results realised with the convenience sample of 44 third-grade secondary school students where the CLIL group was exposed to cca 810 more hours

¹ Coyle D. Developing CLIL: Towards a Theory of Practice. In: Monograph 6. Barcelona, Spain: APAC Barcelona; 2006. (In Eng.)

² Montalto S.A., Walter L., Theodorou M., Chrysanthou K. The CLIL Guide Book. Lifelong Learning Program. 2014. Available at: <https://www.languages.dk/archive/clil4u/book/CLIL%20Book%20En.pdf> (accessed 25.11.2020). (In Eng.)

³ Ibid.

⁴ Coyle D., Hood P., Marsh D. CLIL: Content and Language Integrated Learning. Cambridge: Cambridge University Press; 2010. Available at: <https://abdn.pure.elsevier.com/en/publications/content-and-language-integrated-learning> (accessed 25.11.2020). (In Eng.)

⁵ Kovalik S., Olsen K. ITI: The Model: Integrated Thematic Instruction; United States: S. Kovalik & Associates; 1993. (In Eng.)

of the foreign language than the non-CLIL group. The authors highlight the added value of CLIL application – an effect on the way learners perform when learning languages. CLIL students apply language learning strategies more effectively compared to their non-CLIL peers. The learning is influenced by the motivation and attitudes of students towards the subject. The students' attitude towards the target language was also the subject of numerous studies, some of which indicate that it declines in time [29]. Martínez Agudo [30] confirmed the results of research conducted by Somers & Llinares [31] and claimed that motivation observed at the initial stages of CLIL implementation starts to decline when CLIL becomes a regular practice and is no more a novelty for learners.

The impact on content and a possible influence (both positive and negative) on L1 should be studied systematically simultaneously with the effect on foreign language. Impressive results on using CLIL and L1 proficiency are presented in the studies, e.g. [27; 32–34] and indicate that the research in the field should be extended. Naturally, using a target language in content subject evokes the questions concerning L1 knowledge and its possible problems; e.g. attention divided between learning language and learning content; smaller L1 vocabulary size compared to learners' monolingual pairs; problems with subject terminology in L1. etc. L1 formed the central focus of a study by Navarro-Pablo & Gándara [32] in which the authors found that CLIL does not negatively influence the development of learners' L1 skills and knowledge, and they suggest that appropriate CLIL delivery based on the years of implementation and experience may benefit it. Similarly, Pérez Cañado claims, based on her research, that “CLIL is not detrimentally impacting L1 competence and is not watering down content learning, on which the positive impact of CLIL is particularly felt in the long term” [27, p. 18]. Esther Nieto-Moreno-de-Diezmás studied the effect of content and language integrated learning on reading competence development in the mother tongue and observed and concluded a significantly positive impact of bilingual education on vocabulary in the L1 [33]. The author claims that

“bilingual education (CLIL) does not harm the acquisition of reading competence in the mother tongue, since there are no significant overall differences between the CLIL experimental group and mainstream students” [33, p. 50]. What more, the results of the nationwide Spanish study [32] show not just a neutral effect but a positive one. The sample consisted of 271 primary (age 11) and secondary school (age 15) students from seven public schools and 38 EFL teachers, content teachers and language assistants. One of the study aims was to focus on the differences in Spanish Language and Literature results between CLIL and non-CLIL students. There were statistically significant differences in Spanish Language and Literature results between CLIL and non-CLIL students, favouring CLIL ones: both primary and secondary CLIL students outperformed non-CLIL students despite expectations derived from reduced L1 input in their end-of-year assessment in Spanish Language and Literature.

The research results, however, have to be carefully interpreted as, as it has been already mentioned, there are many moderating variables and factors that influence the strength or direction of an effect (of the educational process – age, gender, proficiency level, motivation, exposure time, intensity, social status and others) (see also [35]).

The replication studies are very important in research generally and in education especially. The experimenters should follow the same procedures; however, it is very difficult to minimise the factors that influence education results. There are not many studies that have examined the effects of CLIL on the content subjects and provide quantitative evidence of this impact. Not rarely, the experimental studies use relatively small convenience samples and the effect size referring to the value of a statistic calculated from a sample of data, similarly as the significance can be difficult to interpret. Systematic reviews and meta-analysis can be used to aggregate the effect size by integrating the results of different studies (selected based on defined criteria according to the research question) and meta-analyses enable the researchers to synthesise data from research with the same characteristics.



This study aims to establish comprehensive evidence about the effect of CLIL on the content subject. Therefore, we will conduct a review and meta-analysis of selected studies' data and compare the CLIL versus non-CLIL group. Concerning the data presented in this research, we have to admit, that we do not deal with the replication studies but tried to select the studies that work with a similar sample (age, type of school) and focus on the influence of CLIL on a subject discipline (content knowledge) based on the experimental data analysis.

The Effect of CLIL on Content Learning

Bonces defines CLIL as “a coherent way of doubling the amount of exposure to the language, without the necessity of adding more room in the timetable for language (only) lessons” [36, p. 183] what might seem to suggest that the primary focus of CLIL is language learning. This is not entirely the real situation as CLIL is content-driven⁶ and the dual aims should guarantee content development. Yet, Cenoz, Genesee & Gorter indicate that the proofs about the balanced pedagogic integration of content and language in CLIL are unconvincing [37]. Even after 10 years since the text was published we still miss the data that would indicate the change. To date, studies investigating CLIL and its impact on the content subject have produced equivocal results. Some studies have shown the beneficial effects of CLIL, but others have shown a deterioration in the results.

A systematic review realised by Graham et al. presents the results of 25 studies that were focused on language and/or content development as a result of CLIL application [38]. Six of them dealt with content (Mathematics (2), Physics (1); Accounting, Finance and History (1); World economy history and world economy (1); Science (1)). In one case there was no statistically significant difference between the groups; in 2 studies (both focused on Mathematics) CLIL groups performed better. Speaking about all 25 studies, there are only three cases with the results positive in favour of non-CLIL,

eight studies found no difference between the CLIL and non-CLIL groups and in the rest of the studies (14) CLIL students reached better results than the non-CLIL students. The authors stress the need to be cautious, as not all studies present the pre-test results. It is equally important to mention that various studies mention not only positive results in knowledge gains but they also present based on both quantitative and qualitative data, change of motivation [39], attitude [40; 41], use of strategies [42], way of thinking, critical thinking [43], analogical reasoning [44] etc.

This study synthesises research located from WoS database on CLIL teaching conducted in the last 10 years and examines the impact of CLIL on content knowledge gains in the groups of 10–16 years old learners in CLIL and non-CLIL settings.

Materials and Methods

Review question. This review focuses on studies exploring the effectiveness of a CLIL on content subject knowledge in a group of students aged 10–16. In the majority of European countries CLIL is implemented at the lower secondary education (ISCED 2) what corresponds to the 10/11–15/16 years. The review also aims to answer the questions which methods are common to assess content learning in CLIL and define the possible gaps in the research.

Selection of the studies. To identify the relevant studies, four databases from the Web of Science Core Collection were used as a source of high-quality peer-reviewed studies, namely (1) Science Citation Index Expanded, (2) Social Sciences Citation Index, (3) Arts & Humanities Citation Index and (4) Emerging Sources Citation Index. The timespan was limited to the studies published in the period from 2010 till 2020. The basic search (looking for “CLIL research” studies) resulted in 395 studies.

Inclusion criteria. All texts were screened following the PRISMA protocol steps. PRISMA statement consists of a checklist and a flow diagram, and it is a set of items for reporting systematic reviews. It is available online⁷ and has been published in several

⁶ Coyle D., Hood P., Marsh D. CLIL: Content and Language Integrated Learning.

⁷ PRISMA. Transparent Reporting of Systematic Reviews and Meta-Analyses [Electronic resource]. Available at: <http://www.prisma-statement.org/>.

journals. After the first screening, six articles were identified as duplications published in two different journals. To be included in the study the text had to (1) apply quantitative research methods, (2) the sample age corresponds to the research question (10–16 years) and (3) possibly provide statistical data (n, mean, SD) (4) comparing intervention and conventional group (5) with the focus on content/subject discipline knowledge (gain).

Forty-five articles were excluded after the titles screening. Some titles (e.g. Teaching linguistics to low-level English language users in a teacher education programme: an action research study; Languages of schooling in European policymaking: present state and future outcomes; Empowering Teachers, Triggering Change: A Case Study of Teacher Training through Action Research) define the sample age or the focus and thus it was possible to exclude them. Based on the abstract reading, the screening resulted in

112 articles, out of which 94 full-texts were retrieved. The majority of abstracts introduce the sample age and the focus and based on this information a relatively vast number of the articles were excluded. As the school systems and the terminology are not identical across the countries (e.g. term secondary school student – in Slovakia responds to a 15–19 years old student, while in Spain to a 12–16 years old learner), it was not possible to identify the age of the sample described what resulted in the subsequent exclusion of the articles due to the sample’s age (n = 25) after the full-text screening. Another 40 articles were excluded due to their focus not corresponding with the focus of the present systematic review, three articles were written in Spanish (even though the language was one of the criteria three Spanish texts were included), and four articles were critical reviews and did not present the research in the studied field of CLIL (figure).

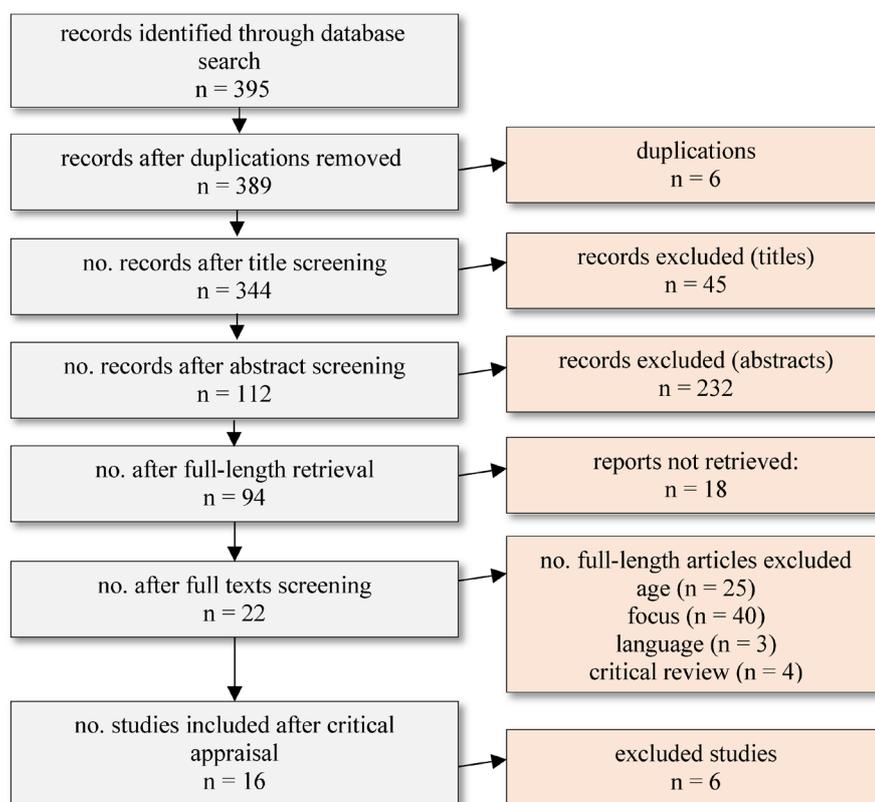


Figure. PRISMA flow diagram – articles selection



The list of studies was reduced to 22 out of which, after full-text reading and critical appraisal, sixteen following studies (table 1) were selected as the subject of the present analysis. Six studies that were excluded during the second reading described, e.g. the research based on one 40-minute lesson; data were collected through questionnaires to teachers to evaluate the impact on teaching and learning through this approach, etc.

Six articles presented data (n, mean and SD for both CLIL and non-CLIL groups) that allowed us to realise statistical meta-analysis. Comprehensive meta analysis (CMA version 3.3.070, trial/evaluation version) and RevMan (Review Manager 5) software were used to conduct a meta-analysis. The measures of the effect of the intervention were generally continuous data based on results obtained in a test and we used mean and standard deviation to compare the effect. Even though we tried to select the studies that met set criteria, the effect size could vary according to the not controlled variables or the broader set limit (e.g. age, different populations) and thus we applied a random-effects model. The level of statistical significance was set at $p < 0.05$. As the studies in the analysis did not use the same scale, it would be not appropriate to use raw differences in means and thus to assess the outcome the standardised mean difference (δ) and the unbiased estimate of δ (Hedges' g) were used. Hedges' g , also called corrected effect size, measures the effect size. Hedges' formula is

$$\text{Hedges' } g = \frac{M_1 - M_2}{SD^*_{pooled}}, \text{ where } M_1 - M_2 \text{ is a dif-}$$

ference in means SD^*_{pooled} is a pooled standard deviation, a weighted average of standard deviations for two or more groups. Pooled standard deviation reflects the sample sizes and standard deviation are averaged with more weight given to the larger sample groups. Glenn introduces three levels or categories of effects (a) small effect (cannot be discerned by the naked eye) = 0.2, (b) medium effect = 0.5 and (c) large effect = 0.8⁸.

Results and Discussion

The content subjects in the studies differ; there were five studies focused on natural science (2 Natural sciences, 1 Geography, 1 Biology and 1 Physics), four studies on Mathematics, 2 History, 1 Social science, 1 Science, 1 Music, 1 Physical Education and 1 Digital Skills. Almost one-third of the studies were conducted in Spain (6), 4 in Germany and 1 in Belgium, Cyprus, the Czech Republic, Greece, Norway and Switzerland.

Arts and Mathematics communicate the meaning in "universal language", they need no or little translation as they communicate meaning via symbols and images and thus many teachers opt to introduce CLIL via those subjects. This allows the teachers to apply multiple foci, create a safe and rich learning environment, realise authentic task, support active learning, visualise and scaffold the content⁹. Surmont et al. observed the group of first-year pupils of secondary education in the Dutch-speaking part of Belgium (n = 107; $M = 53$, $F = 54$, age 12.3 years) [45]. The pupils voluntarily selected the possibility to follow the CLIL course ($n_{CLIL} = 35$). The CLIL group followed the course for ten months and three testings were realised during the period (T0 at the beginning, T1 after three months and T2 after ten months at the end of the course) to observe the progress of individual groups and to compare the results. The researchers used three versions of the mathematical tests Mathematical Assessment Test-Help (MATH). After ten months, the mean scores of both groups showed statistically significantly different improvements compared to the mean scores at the beginning. Comparison between the groups showed that CLIL's pupil's progress was significantly better ($p < 0.5$) over the 10 months period when compared to a control group. The difference was evident even after three months (T1).

Fleckenstein et al. stress that the results of L2 math teaching in the immersion programs reported heterogeneous results and "immersion programs attain the same levels...

⁸ Glen S. Hedges' g : Definition, Formula. In: Statistics HowTo.com: Elementary Statistics for the rest of us! 2016. Available at: <http://www.statisticshowto.com/hedges-g/> (In Eng.)

⁹ Mehisto P. et al. Uncovering CLIL: Content and Language Integrated Learning in Bilingual and Multilingual Education. Oxford: Macmillan Education; 2008. (In Eng.)



Table 1. Details of included studies

Study ID	Study author(s)	Year	Source	Country	Sample size	Used in a subject
1	2	3	4	5	6	7
Belles-Calvera 2018	Belles-Calvera L.	2018	Teaching Music in English: A Content-Based Instruction Model in Secondary Education. <i>Latin American Journal of Content & Language Integrated Learning</i> . 11(1):109-140.	Spain	44	Music
Binterova 2013	Binterova H., Sulista M.	2013	GeoGebra Software Use within a Content and Language Integrated Learning Environment. <i>European Journal of Contemporary Education</i> . 4(2):100-116.	Czech republic	243 reduced to 175 for the study	Mathematics
Dallinger 2016	Dallinger S., Jonkmann K., Hollm J., Fiege C.	2016	The effect of content and language integrated learning on students' English and history competences – Killing two birds with one stone? <i>Learning and Instruction</i> . 41:23-31.	Germany	837	History
Fernández-Sanjujo 2017	Fernández-Sanjujo J., Fernández-Costales A., Blanco J.M.	2017	Analysing students' content-learning in science in CLIL vs. non-CLIL programmes: Empirical evidence from Spain. <i>International Journal of Bilingual Education and Bilingualism</i> . 22(6):661-674.	Spain	709	Natural Sciences
Fleckenstein 2019	Fleckenstein J., Gebauer S.K., Möller J.	2019	Promoting mathematics achievement in one-way immersion: Performance development over four years of elementary school. <i>Contemporary Educational Psychology</i> . 56:228-235.	Germany	877	Mathematics
Isidro 2018	Isidro X.S., Lasgabaster D.	2018	The impact of CLIL on pluriliteracy development and content learning in a rural multilingual setting: A longitudinal study. <i>Language Teaching Research</i> . 23(5):584-602.	Spain	44	Social Sciences
Lialikhova 2018	Lialikhova D.	2018	Triggers and constraints of lower secondary students' willingness to communicate orally in English in a CLIL setting in the Norwegian context. <i>Journal of Immersion and Content-Based Language Education</i> . 6(1):27-56.	Norway	27	History
Martínez 2020	Martínez J.d.D.	2020	To what extent does parental educational background affect CLIL learners' content subject learning? Evidence from research in Spain. <i>International Journal of Bilingual Education and Bilingualism</i> . p. 1-14.	Spain	318	Natural Sciences



End of table 1

1	2	3	4	5	6	7
Mattheoudakis 2014	Mattheoudakis M., Alexiou T., Laskaridou C.	2014	To CLIL or Not to CLIL? The Case of the 3rd Experimental Primary School in Evosmos. <i>Major Trends in Theoretical and Applied Linguistics</i> . Vol. 3. Berlin, Boston: De Gruyter.	Greece	51	Geography
Meyerhöffer 2018	Meyerhöffer N., Dreesmann D.C.	2018	The exclusive language of science? Comparing knowledge gains and motivation in English-bilingual biology lessons between non-selected and preselected classes. <i>International Journal of Science Education</i> . 41(1):1-20.	Germany	241	Biology
Nieto 2018	Nieto Moreno de Diezmas E.	2018	Exploring CLIL contribution towards the acquisition of cross-curricular competences: A comparative study on digital competence development in CLIL. <i>Revista De Lingüística Y Lenguas Aplicadas</i> . 13(1):75.	Spain	20 184	Digital skills
Piesche 2016	Piesche N., Jonkmann K., Fiege C., Keßler J.	2016	CLIL for all? A randomised controlled field experiment with sixth-grade students on the effects of content and language integrated science learning. <i>Learning and Instruction</i> . 44:108-116.	Germany	722	Physics
Salvador-Garcia 2020	Salvador-Garcia C., Capella-Peris C., Chiva-Bartoll O., Ruiz-Montero P.J.	2020	A Mixed Methods Study to Examine the Influence of CLIL on Physical Education Lessons: Analysis of Social Interactions and Physical Activity Levels. <i>Frontiers in Psychology</i> [online]. 2020; 11.	Spain	49	Physical Education
Serra 2017	Serra C.	2017	Assessing CLIL at Primary School: A Longitudinal Study. <i>International Journal of Bilingual Education and Bilingualism</i> . 2008; 10(5):582-602.	Switzerland	110	Mathematics
Surmont 2016	Surmont J., Struys E., Noort M.V., Craen P. V.	2016	The effects of CLIL on mathematical content learning: A longitudinal study. <i>Studies in Second Language Learning and Teaching</i> . 6(2):319-337.	Belgium (Dutch speaking part)	107	Mathematics
Xanthou 2011	Xanthou M.	2011	The impact of CLIL on L2 vocabulary development and content knowledge. <i>English Teaching: Practice and Critique</i> . 10(4):116-126.	Cyprus	77	Science

or even higher levels in mathematics achievement compared to conventionally schooled students” [46, p. 229]. The researchers conducted a longitudinal study in Germany that lasted from 2004 to 2009 with pupils ($n = 590$) aged six years at the beginning of the research in the immersion/CLIL programme and they compared the results with a group of pupils in regular classes ($n = 287$). The pupils were tested at the end of each year. It is crucial to notice that the test language was German, i.e. L1. The authors state, “The achievement growth in immersion programs is quite consistently high until the end of grade 3 and flattens out during grade 4, presumably due to an instructional ceiling effect. In contrast, achievement in the conventional program increases at a slower rate until the end of grade 3. Grade 4 has compensatory effects; conventionally schooled students show higher gains than immersion students until the end of that year” [46, p. 233].

A short term study [47] realised in the Czech republic focuses primarily on the students’ perceptions of teaching mathematics in CLIL. The authors applied quantitative and qualitative tools with a sample of 5th and 6th graders ($n = 55$) at three different schools. The CLIL method was applied in the Math classes at least four times per month in the experimental groups. Except for the attitude tests students filled before the experiment, selected lessons were video-recorded and analysed; researchers also conducted the interviews with teachers and pupils. The attitude test results indicated significantly different results between CLIL and non-CLIL groups, with more positive results in the CLIL group. It has been mentioned above that numerous studies reported the positive motion shortly after the CLIL integration followed by its decline. Even though the authors mention longitudinal observations, compared to other studies, we deal here with two one-month periods and these can be evaluated as a short-term period. Speaking about observations, the authors mention an interesting finding about the teachers in CLIL classes; namely, they state that teachers used more activating methods and communicated more with their pupils [47,

p. 110] and more intensive communication. The integration of mathematics content and Italian (2 groups) or Romansh (1 group) as a second language was a subject of study realised by Serra [48] in three Swiss primary schools. The researchers reported on a longitudinal study in which they observed learners in grades 1 to 6. The study inclusion criterion was the sample age should correspond to 10–16 years. The study was included even though the age of pupils in the sample was 6 to 12, considering the fact that pupils at the end of the study (last three years students were aged 10–12 who are considered eligible participants) met the set criterion. The study concerned the oral and written production in L2, the role of interaction for L2 acquisition and progress in mathematics that was evaluated based on the standardised math tests. The narrative analysis was focused on “the relevance of repair sequences to draw focus on form while negotiating meaning” [48, p. 583]. The researchers highlight that grammar instruction and search for accuracy were mostly to be connected to focus on content activities, this made teaching more authentic and meaningful as it was the response to the real need and not only an occasional shift to linguistic code features. Those situations “bring learners to notice the relationship that exists between meaning, forms and function in a highly context-sensitive situation” [48, p. 586]. The math results do not only present the comparison of the experimental and control groups but also a representative sample (cockpit reference sample, $n = 450$). The tests in the experimental groups were realised yearly, control groups in Grade 1, 2 and 4 and Control cockpit results were used to compare the results for Grades 3–6. In grade 1 all control groups performed better than the experimental ones. The results changed in Grade 2 when only 1 Italian group reached weaker results than the control group. In Grade 3, similarly as in Grade 6 all bilingual classes perform better (2 groups significantly better) than their monolingual counterparts. The Italian groups results in grades 4 and 5 were lower than the Cockpit reference sample but the control classes in Grade 4 perform better than the reference sample. The mean percentage



of the correct answers in the L2 mathematics test was 59.8% for Romansh group and 67.2% and 63.5% for Italian groups compared to the average of Cockpit reference sample – 50%. The important conclusion, concerning the content teaching and learning the authors state, is that rephrasings in L2 and in L1 involved “both the subject language and the everyday language to convey the intended meaning” [48, p. 600] what effectively supported the processing of subject content.

Piesche et al. studied the impact of CLIL on Physics knowledge in a group of German 6th graders ($n = 722$) [49]. It was (compared to the above-mentioned studies) a short-term study (5 lessons lasting for 90 minutes). The group comparisons showed that monolingual educated learners ($n_{\text{non-CLIL}} = 360$) outperformed those in CLIL groups. The research state a small effect size (-0.2), and considering the possible reasons they mention the lack of previous CLIL experience and the level of their language proficiency. They also mention the volume of the mother tongue and target language as possible factor that can influence the results.

A similar sample was the subject of Spanish research (6th graders, $n = 709$) realised by Fernández-Sanjurjo et al. [50] who also present the data where the non-CLIL group ($n = 357$) outperformed the experimental CLIL group in Natural Science ($p < 0.000$). The test for science content was specially designed according to the curricular content and included closed and open questions. Students in a CLIL group learn two subjects in L2 and have two English lessons per week.

The non-CLIL students outperformed their counterparts also in a study published by Mattheoudakis et al. [51] who present the data of year research where CLIL was applied in Geography (that is currently considered as a subject bridging the natural and social sciences) in the 6th grade ($n_{\text{CLIL}} = 26$; $n_{\text{non-CLIL}} = 25$) in Greece. CLIL learners had two classes of Geography per week instructed in English and both groups had eight classes of English weekly. Participants were tested three times and “CLIL learners scored higher in two out of the three tests; in content test 2 this difference reached statistical

significance ($p < 0.001$)” [51, p. 9]. In content test 3 the non-CLIL group scored higher than the CLIL group. The content tests topics were the same for both groups, but the languages were different (English for CLIL group, Greek for non-CLIL groups). The foreign language receptive skills were also tested and the researchers summarised that CLIL practice had a positive impact on foreign language learning (even though the result was not statistically significant).

Meyerhöffer & Dreesmann [52] applied CLIL in a group of slightly older students. They studied learning gains and motivation in the 9th grade CLIL and non-CLIL Biology classes in Germany. In their study, they highlight the importance of the selection that is applied in Germany. They explain that CLIL students are selected based on their previous academic results, their attitude towards foreign language learning, and their motivation in school. This is why they compared the groups of pre-selected students and bilingually inexperienced, non-selected students in their research. The sample consisted of 243 students (on average 14.3 years old, ranging from 12 to 16). The CLIL (experimental) group ($n = 168$) consisted of 85 bilingually inexperienced learners and 82 pre-selected learners for bilingual or gifted programmes. The authors of the study compared both total scores and gains of both control and experimental groups. They summarise that the increase in the content knowledge was similar between the groups and point out that the results “provide evidence against concerns that teaching non-selected students bilingually might lead to deficits in content knowledge acquisition” [52, p. 1].

Application of CLIL in Cypriot context was the subject of the study [53] with the focus on L2 vocabulary and content knowledge. Two quasi-experiments with different groups are described in the article. Both qualitative and quantitative data were collected. The pre- and post-treatment tests were administered in control and experimental groups to assess the vocabulary breadth and content knowledge. Video and audio recordings from experimental classrooms were analysed to interpret the quantitative data. There were no statistical differences

at the outset of the experiment. The subject matter tests were realised four days after the experiments with the L1 and L2 items. Both the CLIL and non-CLIL groups exhibited a significantly positive increase indicating a positive impact of CLIL on content knowledge. Although there was a positive mean difference between the groups, the difference was not statistically significant. This was the truth for both experiments described.

The correlation of affective variables and content learning achievement in CLIL programmes was studied by Martínez Agudo [30]. English level was evaluated by collecting learners' English grades; a battery of tests was used to assess learners' intellectual aptitudes, and the content knowledge was measured by learners' final grades (out of a total score of 10). The author stresses that "summative assessment may certainly generate heightened test anxiety in many cases due to added pressure on CLIL students to show both language and content related competences" [34]. Based on the discriminant analysis author summarises that lack of interest is the variable that had the greatest weight in explaining the differences between the achievement in natural sciences between the CLIL and non-CLIL groups.

The study written by Isidro & Lasagabaster [54] presents interesting data on teaching CLIL in Social Sciences classes in Spain. The experiment lasted for two years (what allowed them to observe students in different periods, after year 1 and after year 2) and the students in the sample were in their 3rd year of secondary education (14–15-year-olds). The specially designed test was prepared to measure previous knowledge, and at the beginning of the experiment the groups were homogeneous in terms of language and Social science performance. Interestingly, the means of non-CLIL students showed a slight decrease in their results in the different phases while the CLIL cohort did not show significant changes. Comparison between the groups did not show the statistical differences between the means, and the researchers state that CLIL "did not have any detrimental effect on CLIL students' learning of content" [54, p. 14].

History was the curricular area for CLIL implementation in the study conducted in

Norway [55]. The author focussed her attention on the learners' willingness to communicate orally and their motivation. It was a small-scale a short-term experiment (6-week intervention) resulting in a conclusion that "the CLIL intervention had reinforced most students' motivation and WTC [willingness to communicate] orally compared to their regular EFL lessons" [55]. Based on the studies stating the declination of motivation within time, it would be interesting to prolong the study [55] and observe WTC's possible impact. It is equally important to mention that Scandinavian countries have several common features that differentiate them from other European countries concerning foreign languages (learning). The fact that most television programmes are not dubbed can significantly influence EFL teaching. This is also indicated by Sylvén [56] who described the contextual differences of 4 European countries and analysed possible factors that may affect the success of CLIL. She mentions policy, teacher (education), age (and cognitive development) and extramural English (and the amount of exposure) as the key factors that may influence the result. She states that "regarding the amount of exposure to English outside of school there are huge differences" [56, p. 315] and Sweden compared to Finland, Germany and Spain also reached the highest scores in English language skills. This might also be one of the reasons why in some countries is CLIL not so successful.

Dallinger et al. focussed their attention on different aspects; besides language and content they also studied motivation, demography, cognitive abilities [57]. Regarding History, CLIL students reached significantly higher results; however, after the second year the differences indicated a (not significant) advantage for CLIL students.

Another small-scale and short-term experiment [58] was realised in Spain and studied the integration of Music and EFL teaching. The researcher used a questionnaire that involved factual and attitudinal questions. The researchers worked with qualitative data and concluded that teaching Music through English shows beneficial effects. On the other hand, the authors warn that some learners are stressed because of



their low English proficiency and thus using English as a medium of teaching in other subjects is still dubious. The quantitative and qualitative data analysis was the subjects of the research in Spain [59] in which researchers collected data based on the interviews with 12 participants, a sociometric questionnaire and a quasi-experiment. The impact of CLIL on physical activity in Physical education lessons was evaluated using accelerometry. This allowed researchers to measure sedentary-light physical activity and moderate to vigorous physical activity. Their findings show that the CLIL group obtained higher levels of MVPA than the non-CLIL group. An interesting remark is mentioned by the authors concerning integrating FL teaching and physical education, namely that teachers are concerned with the vocabulary they use, language structures they use and they “may overuse language learning materials such as flashcards and, consequently, the teacher talking time is increased while students’ activity time is diminished” [59, p. 7–8]. The study [59] was focused not on the integration of content subject and foreign language teaching but rather on the difference in learners’ digital competence. We included this study as the area of ICT, computer and digital skills are usually the compulsory subjects taught at elementary and/or secondary schools. The researchers tested 2nd year students in compulsory education (aged 13–14) regarding their (a) communicative digital competence and a year later (b) informational digital competence. In the first year 18 093 CLIL students and 2 152 non-CLIL students formed the sample and in the second year 2 581 CLIL and 17 553 non-CLIL students. It is essential to say that the testees in the two tests were not the same groups but two consecutive generations, students enrolled in the 2nd year of secondary education. It is apparent from the results that CLIL students significantly outperformed their peers in both tests, communicative competence and information competence. Six standards were the subject of the communicative competence evaluation (respect the rules of participation in virtual networks, handling network communication tools, using the internet as a source of information, sending email, understanding risks

of sharing personal data and managing files and folders). Information competence was evaluated based on fourteen standards (e.g. compressing folders, copying files to share, creating back-up copies, editing with a word processor, spreadsheets, selecting information critically, drawing and editing images etc.). The CLIL students reached better results in 07 out of 20 standards compared to their non-CLIL counterparts. The author suggests that one of the reasons can be that CLIL creates a methodological framework that naturally leads to developing cross-curricular competencies.

The table below summarises the data from the included studies that presented data about the influence of CLIL on the content subject. Studies in bold presented the statistical data that were used in the meta-analyses (see the text and tables 2).

As it can be seen, it is difficult to evaluate the results generally. The character of content subjects differ, similarly to teaching methods and approaches applied. The results presented above suggest there is a positive impact of CLIL on content learning. Eleven studies present data where CLIL students significantly (or not significantly) outperformed non-CLIL students. The studies that were selected had to focus their attention on content teaching and the impact of CLIL on the results of content subjects. However, some of the studies reported noteworthy limitations, e.g. the length of the study or the sample size. Another critical factor that has to be mentioned is publication bias, a tendency to write about the positive effects in education rather than about the negative impacts or results.

Out of 16 presented studies, six studies included the data that could be evaluated using RevMan software what allows us to combine, synthesise and analyse the selected studies. The studies that do not bring information on p value, SD or simply present data that are not statistically significant are not included in the following meta-analyses. The following table 3 with the forest plot summarises the means, SDs, effect sizes (described above) and confidence intervals of the studies. The confidence interval is the range of values which is likely to contain a population parameter. The table 3

shows that research results have been contradictory. Altogether, there were 3 303 pupils in the studied ($n_{\text{CLIL}} = 1\ 847$ and $n_{\text{non-CLIL}} = 1\ 456$) sample.

Differences of four studies (out of 6) are statistically significant – Fernández-Sanjurjo 2017 focusing on natural science, Fleckenstein 2019 – mathematics, Mattheoudakis 2014 – geography, and Piesche 2016 – physics. The confidence interval of 2 studies (Dallinger 2016, Surmont 2016), similarly to the summary result, include zero, i.e. the differences are not statistically significant. In both studies, CLIL group students outperformed the non-CLIL students. Four studies had a balanced influence on the overall estimate (19% and 18.9%), namely Dallinger 2016, Fleckenstein 2017, Fernández-Sanjurjo 2017 and Piesche 2016 (table 4).

The summary results show that the effect sizes fall in the range of -0.61 to 0.19 ; the proportion of observed variance (I^2) is very high (92%) what means we deal with substantial heterogeneity. The combined effect size is -0.14 (what can be evaluated as a small effect) with a 95% confidence interval of -0.40 to 0.13 . Confidence intervals are broader as we deal with the random model. The p-value for the summary effect is 0.31. The variance of dispersion (τ^2) that reflects the variance of the true effect is 0.09 what is a small effect.

“Criticism has recently been leveled at CLIL due to the plethora of models or variants which can be identified within it” [60, p. 14] and Banegas [61] stresses that CLIL shortcomings need to be addressed. Researchers are calling for further concise research covering and studying aspects that may influence the results and interpretation.

Spain is probably the most experienced European country in CLIL application and has authored numerous studies focusing on L2, content, different factors as well as L1. The government substantially supports the application of CLIL in Spain; in other countries it is realised systematically but offered as an option (e.g. in Germany, students apply to secondary schools with CLIL programme), and there are countries where it is implemented rather randomly depending on the capacities and willingness of teachers

and approval of the school management and parents. This is an important factor that may influence the results and interpretation of the research conducted in different countries. Even though we suggest there are numerous factors that may influence the result, we believe that replication studies and meta-analysis can shed more light on the positive or negative effects of CLIL application on both content and language learning. Three included studies were realised in Germany, one in Spain, Belgium and Greece. The length of treatment varied from 5 lessons to 4 years. Internal consistency of the majority of the tests used was tested and Cronbach alpha was presented. The majority of studies were based on pre-test – post-test research design showing the groups’ homogeneity and no statistical differences between the groups. In most of the studies, the researchers used tests out of which some were standardised, but researchers also used validated and non-validated tests. In four studies the non-CLIL students outperformed their CLIL peers; however, the data show that the difference between the groups was not statistically significant. This undoubtedly can be perceived positively; the CLIL does not negatively affect the content subject. This should be, however, evaluated along with language performance.

Limitations of the Study

Meta-analysis as an observational study of selected studies synthesises data from different (even small samples) where the results can even be from various reasons not statistically significant. On the other hand, there are some aspects that can be understood or perceived as threat, risks or drawbacks. Not all the studies are realised in the same conditions and do not control all the effects (see the text). The selection of the studies can also be understood as a limitation as “some studies have not been published, or have been published in a form to which the researcher has no access, or have been published in a language that the researcher cannot read, etcetera” [62]. The authors (ibid) also mention the problem with probability sampling, missing cases, the problem with pre-test and post-design and test differences.



Table 2. Details of included studies – Content subjects results

Study ID	Country	Sample grade (age)	No of CLIL/ non CLIL	Subject	Instrument // research design	Length	Results
Belles-Calvera 2018	Spain	12–14 years	22/22	music	final survey // experiment, comparison	month	positive attitude toward CLIL implementation (students)
Binterova 2013	Czech republic	6 th – 8 th grade	78/97	mathematics	questionnaire, long-term observations, video-recordings, interview// experiment	1 year (2 one-month periods)	learning climate: CLIL lessons more positively perceived by CLIL students; CLIL lessons indicated more activating methods, an increase in students' activity
Dallinger 2016	Germany	8 th grade (13,5 avg age)	483/354 (reduced)	history	The historical knowledge tests were developed by authors of this study who are trained History teachers ($\alpha = 0.72$) // comparison	data collected autumn 2012-summer 2013, CLIL applied since 5 th grade	CLIL 76.87 non-CLIL 76.01 not st. significant
Fernández-Sanjurjo 2017	Spain	6 th grade	352/357	natural sciences	Validated evaluation test for Science ($\alpha = 0.79$) context questionnaire // small-scale experiment		non-CLIL, $p = 0.000$
Fleckenstein 2019	Germany	mean age at T1 – 6 years	590/287	mathematics	DEMAT (Detscher Mathematiktest) – administered at the end of each school year – tested in L1 // comparison, experiment	4 years	CLIL
Isidro 2018	Spain	3 rd grade (14–15) 4 th grade (15–16)	20/24	social sciences	a test specially designed (in Galician) // comparison, experiment	2 years	CLIL ($M1 = 66.4, M2 = 64.79$ and $M3 = 67.29$) non-CLIL ($M1 = 65, M2 = 56.17$ and $M3 = 58.62$) not st. significant
Lialikhova 2018	Norway	14–15 years	27	history	classroom observations, teacher pre and post-interviews, student pre- and post-questionnaires // unstructured observation	18 CLIL lessons (6 weeks)	not all motivational variables exert the same effect on the CLIL learners' content achievement
Martínez 2020	Spain	6 th grade (11.4) 4 th grade (15.6)	156/162	natural sciences	verbal intelligence test, Pelechano's MA test, natural sciences grades // comparison		not all motivational variables exert the same effect on the CLIL learners' content achievement
Mattheoudakis 2014	Greece	6 th grade (11–12)	26/25	geography	A language test was designed by the English language teachers of the school aiming to test CLIL and non CLIL learners' reading and listening skills in English (the same test before and after intervention) // comparison, experiment	9 months, 1 ac. year	non-CLIL



End of table 2

1	2	3	4	5	6	7	8
Meyerhöffer 2018	Germany	9 th grade	168/73	biology	The knowledge test based on curricular standards featured eight close-ended items (five true/false, one multiple choice, two matching items) and eight short-answer items that covered content of all five modules of the unit // pre-test/post test design -experiment		CLIL, not. st. dif
Nieto 2018	Spain	13-14 years	2152/18093 2581/17553	digital skills	a test evaluating communicative digital competence, a test evaluating informational digital competence // comparison	2 years	CLIL students showed significantly better digital skills than non-CLIL learners, particularly regarding communicative digital competence non-CLIL, $p = 0.004^*$
Piesche 2016	Germany	6 th grade, mean age 11,5	362/360	physics	The students' physics preknowledge was assessed with 25 items (17 multiple-choice items and five free-response items). // a randomized controlled field experiment with three measurement points (pre/posttest, follow-up)	5 lessons lasting 90 minutes	
Salvador-Garcia 2020	Spain	13-14 years	49	physical education	sociometric questionnaire, accelerometers - measurement of the acceleration of a body // mixed methods experiment, comparison of C and E	two 50-min lessons	CLIL, higher levels of MVPA
Serra 2017	Switzerland	grades 1 to 7	46/64	mathematics	narratives, standardised mathematics tests // experiment, comparison qualitative analysis quantitative data		CLIL perform much better than the control sample. No significant differences have been found between the 3 bilingual classes: (ItP1 versus ItP2: $t 0.75$; $p 0.454$); (RomP versus ItP1: $t 1.04$; $p 0.308$); (RomP versus ItP2: $t 0.42$; $p 0.675$).
Surmont 2016	Belgium (Dutch speaking part)	CLIL M = 12.23; non-CLIL M = 12.36	M = 35/72	mathematics	Mathematical Assessment Test Help(MATH) // experiment, comparison C and E	10 months	CLIL
Xanthou 2011	Cyprus	6 th grade (11 ears)	exp 1: 15/16 exp2: 21/25	science	Vocabulary pre-post tests - to test vocab. Breadth Content tests assessed only cognitive objectives. // quasi experiment (2 times, $n = 31$; =46) observation	1 year	A significant effect ($p = .000$) of treatment on content knowledge was shown for both experimental and control groups. (No significant difference between the groups)

Table 3. Forest plot illustrating the results using a random-effect model

Study or Subgroup	CLIL		Non-CLIL		Weight	Std. Mean Difference IV, Random, 95% CI
	Mean	SD	Mean	SD		
<input checked="" type="checkbox"/> Dallinger 2016	76.87	8.83	76.01	9.48	19.0%	0.09 [-0.04, 0.23]
<input checked="" type="checkbox"/> Fernández-Sanjurjo 2017	14.99	5.5	17.41	5.07	18.9%	-0.46 [-0.61, -0.31]
<input checked="" type="checkbox"/> Fleckenstein 2019	3.69	0.69	3.56	0.69	19.0%	0.19 [0.05, 0.33]
<input checked="" type="checkbox"/> Mattheoudakis 2014	6	2.2	7.3	2	10.5%	-0.61 [-1.17, -0.05]
<input checked="" type="checkbox"/> Piesche 2016	48.37	10.13	51.63	9.61	18.9%	-0.33 [-0.48, -0.18]
<input checked="" type="checkbox"/> Surmont 2016	65.77	12.31	63.62	14.89	13.7%	0.15 [-0.25, 0.56]
Total (95% CI)					100.0%	-0.14 [-0.40, 0.13]
Heterogeneity: Tau ² = 0.09; Chi ² = 59.67, df = 5 (P < 0.00001); I ² = 92%						
Test for overall effect Z = 1.03 (P = 0.31)						

Table 4. Statistics for the studies and summary

Study ID	Statistics for each study				Hedges's g and 95% CI		
	Hedges's g	Standard error	Variance	Weight	Lower limit	Upper limit	p-Value
Dallinger 2016	0.094	0.070	0.005	19.0%	-0.043	0.231	0.178
Fernández-Sanjurjo 2017	-0.457	0.076	0.006	18.9%	-0.606	-0.308	0.000*
Fleckenstein 2019	0.188	0.071	0.005	19.0%	0.049	0.328	0.008*
Mattheoudakis 2014	-0.609	0.282	0.080	10.5%	-1.163	-0.056	0.031*
Piesche 2016	-0.330	0.075	0.006	18.9%	-0.477	-0.183	0.000*
Surmont 2016	0.154	0.227	0.051	13.7%	-0.290	0.599	0.496
Random model	-0.141	0.136	0.018	100.0%	-0.407	0.125	0.298

In the study presented we could not get enough information on the sample selection, the varying amount of target language and mother tongue use, the language used for testing, similarly the content of tests could not be evaluated. The not unified terminology also has to be mentioned as one of the limitations of the study. The term bilingual education today covers different models and is also introduced with different aims at schools and thus there are cases when the terms CLIL and bilingual education are used interchangeably.

The generalisability of the results is subject to a significant limitation, namely the study presents the research realised in rather a limited range of countries where the effects of CLIL were studied. In the majority of these countries (except for Greece and the Czech Rep), the mother tongue belongs to an analytical, rather than synthetic, language group. It can be a very serious factor as people speaking Germanic and Latin-based languages are believed to be more capable of mastering English compared to other languages.

We also have to mention publication bias that was not estimated in the present study as the number of studies was low, even though the search term was very broad. Selection of the studies indexed in WOS that was done intentionally to ensure the quality of studies can, however, also mean that we missed important data that can significantly influence the summary result.

Conclusion

This research aimed to contribute to the presentation and understanding the need for further study of CLIL through different subject-specific lenses. The number of studies focusing on content impacts is much lower than the studies focused on the impact of CLIL on (L2) development, which also influenced the number of studies included in the present review. Even though the search

started with a relatively high number of studies (n = 395) but after applying the inclusion/exclusion criteria 16 studies met the criteria and checking the homogeneity data that were available for pooled analysis only six studies could be used for meta-analysis. Thirty years of CLIL existence and its application in different countries indicates there are positives of its implementation. Difficulties caused by the vagueness of the term content and language integrated learning can be solved by narrowing the definition or possibly defining categories of CLIL. Similarly, this would also enable the replication of studies that can successively be synthesised, compared and evaluated. Whatmore, this would enable defining the principles of CLIL more precisely along with the conditions when its application can be effective. As to the pedagogical implications, even though the teacher's role has not been mentioned, their attitude, motivation, and self-confidence [63–67] play a crucial role in the quality of CLIL. The systematic preparation of pre-service teachers for possible subjects content integration (not only CLIL), development of their creativity [68; 69], ability to create materials and critical thinking [70] should be one of the main tasks of universities in case the CLIL is to be introduced to our schools.

As a final comment, I would like to mention that undoubtedly, there is a potential of CLIL methodology. But for further study there is a need for a clear definition of CLIL, CBI, bilingual education as they are similar but not synonymic and their proper use in research reports. The information on setting, controlled variables, the way of teacher collaboration (team teaching, co-teaching) is similarly missing in the studies, as well as the information on the percentage of CLIL teaching in a curriculum, language awareness, teaching time in a target language what makes difficult to compare and analyse them.

REFERENCES

1. Lynch G. Introducing Content and Language Integrated Learning in Study Abroad Programs. *Journal of Kanazawa Seiryō University*. 2015; 48(2):69-73. Available at: http://www.seiryō-u.ac.jp/u/education/gakkai/eronsyu_pdf/No125/09_lynch_Introducing_Content_125.pdf (accessed 25.11.2020). (In Eng.)



2. Hanesová D. History of CLIL. In: Pokrivčáková S. (ed.) CLIL in Foreign Language Education: E-textbook for Foreign Language Teachers. Nitra: Constantine the Philosopher University; 2015. p. 7-16. (In Eng.) DOI: <http://doi.org/10.17846/CLIL.2015.7-16>
3. Lasagabaster D., Sierra J.M. Immersion and CLIL in English: More Differences than Similarities. *ELT Journal*. 2010; 64(4):367-375. (In Eng.) DOI: <https://doi.org/10.1093/elt/ccp082>
4. Cenoz J. Content-Based Instruction and Content and Language Integrated Learning: The Same or Different? *Language, Culture and Curriculum*. 2015; 28(1):8-24. (In Eng.) DOI: <https://doi.org/10.1080/07908318.2014.1000922>
5. Ortega L. Researching CLIL and TBLT Interfaces. *System*. 2015; 54:103-109. (In Eng.) DOI: <https://doi.org/10.1016/j.system.2015.09.002>
6. Yang W. ESP vs. CLIL: A Coin of Two Sides or a Continuum of Two Extremes? *ESP Today*. 2016; 4(1):43-68. Available at: https://www.esptodayjournal.org/pdf/current_issue/3.6.2016/WENHSIEN-YANG-full%20text.pdf (accessed 25.11.2020). (In Eng.)
7. Taillefer G. CLIL in Higher Education: The (Perfect?) Crossroads of ESP and Didactic Reflection. *Asp*. 2013; 63:31-53. DOI: <https://doi.org/10.4000/asp.3290>
8. Tzoannopoulou M. Rethinking ESP: Integrating Content and Language in the University Classroom. *Procedia – Social and Behavioral Sciences*. 2015; 173:149-153. (In Eng.) DOI: <https://doi.org/10.1016/j.sbspro.2015.02.045>
9. Borzova E., Shemanaeva M. A University Foreign Language Curriculum for Pre-Service Non-Language Subject Teacher Education. *Education Sciences*. 2019; 9(3):163. (In Eng.) DOI: <https://doi.org/10.3390/educsci9030163>
10. Urgal C.C. Law and Business Students' Attitudes towards Learning English for Specific Purposes within CLIL and Non-CLIL Contexts. *Languages*. 2019; 4(2):45. (In Eng.) DOI: <https://doi.org/10.3390/languages4020045>
11. Nikula T. CLIL: A European Approach to Bilingual Education. In: Van Deusen-Scholl N., May S. (eds) Second and Foreign Language Education. Encyclopedia of Language and Education (3rd ed.). Springer, Cham; 2016. p. 1-14. (In Eng.) DOI: https://doi.org/10.1007/978-3-319-02323-6_10-1
12. Marsh D. Language Awareness and CLIL. In: Cenoz J., Hornberger N.H. (eds.) Encyclopedia of Language and Education. 2008. p. 233-246. (In Eng.) DOI: https://doi.org/10.1007/978-0-387-30424-3_152
13. Cenoz J., Zarobe Y.R. Learning through a Second or Additional Language: Content-Based Instruction and CLIL in the Twenty-First Century. *Language, Culture and Curriculum*. 2015; 28(1):1-7. (In Eng.) DOI: <https://doi.org/10.1080/07908318.2014.1000921>
14. Llinares A. Integration in CLIL: A Proposal to Inform Research and Successful Pedagogy. *Language, Culture and Curriculum*. 2015; 28(1):58-73. (In Eng.) DOI: <https://doi.org/10.1080/07908318.2014.1000925>
15. Pérez Cañado M.L. Evaluating CLIL Programmes: Instrument Design and Validation. Pulso. *Revista de Educación*. 2015; 39:79-112. (In Eng.)
16. Pastrana A., Llinares A., Pascual I. Students' Language Use for Co-Construction of Knowledge in CLIL Group-Work Activities: A Comparison with L1 Settings. *Zeitschrift Für Erziehungswissenschaft*. 2018; 21(1):49-70. (In Eng.) DOI: <https://doi.org/10.1007/s11618-017-0802-y>
17. Zarobe Y.R. Improving Reading Strategy Knowledge in Young Children: What Self-Report Questionnaires Can Reveal. *Elia*. 2017; 17:15-45. (In Eng.) DOI: <https://doi.org/10.12795/elia.2017.i17.02>
18. Straková Z. CLIL and Global Education: A Meaningful Match. *Society. Integration. Education. Proceedings of the International Scientific Conference*. 2020; 5:546-557. (In Eng.) DOI: <https://doi.org/10.17770/sie2020vol5.4819>
19. Pfenninger S.E. The Dynamic Multicausality of Age of First Bilingual Language Exposure: Evidence from a Longitudinal CLIL Study with Dense Time Serial Measurement. *Modern Language Journal*. 2020; 104(3):662-686. (In Eng.) DOI: <https://doi.org/10.1111/modl.12666>
20. Lasagabaster D., Doiz A. A Longitudinal Study on the Impact of CLIL on Affective Factors. *Applied Linguistics*. 2017; 38(5):688-712. (In Eng.) DOI: <https://doi.org/10.1093/applin/amv059>
21. Otwinowska A., Foryś M. They Learn the CLIL Way, but Do They Like It? Affectivity and Cognition in Upper-Primary CLIL Classes. *International Journal of Bilingual Education and Bilingualism*. 2017; 20(5):457-480. (In Eng.) DOI: <https://doi.org/10.1080/13670050.2015.1051944>
22. Fontecha A.F., Canga Alonso A. A Preliminary Study on Motivation and Gender in CLIC and NoN-CLIL Types of Instruction. *International Journal of English Studies*. 2014; 14(1):21-36. (In Eng.) DOI: <https://doi.org/10.6018/ijes/14/1/156681>

23. Canga Alonso A. The Receptive Vocabulary of Spanish 6th-Grade Primary-School Students in CLIL Instruction: A Preliminary Study. *Latin American Journal of Content and Language Integrated Learning*. 2013; 6(2):22-41. (In Eng.) DOI: <https://doi.org/10.5294/laclil.2013.6.2.2>
24. Mensel L.V., Hilgsmann P., Mettwie L., Galand B. CLIL, an Elitist Language Learning Approach? A Background Analysis of English and Dutch CLIL Pupils in French-Speaking Belgium. *Language, Culture and Curriculum*. 2020; 33(1):1-14. (In Eng.) DOI: <https://doi.org/10.1080/07908318.2019.1571078>
25. Perry L., Mcconney A. School Socio-Economic Composition and Student Outcomes in Australia: Implications for Educational Policy. *Australian Journal of Education*. 2010; 54(1):72-85. (In Eng.) DOI: <https://doi.org/10.1177/000494411005400106>
26. Sirin S.R. Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research*. 2005; 75(3):417-453. (In Eng.) DOI: <https://doi.org/10.3102/00346543075003417>
27. Pérez Cañado M.L. The Effects of CLIL on L1 and Content Learning: Updated Empirical Evidence from Monolingual Contexts. *Learning and Instruction*. 2017; 57:18-33. (In Eng.) DOI: <https://doi.org/10.1016/j.learninstruc.2017.12.002>
28. Castellano-Risco I., Alejo-González R., Piquer-Piriz A.M. The Development of Receptive Vocabulary in CLIL vs EFL: Is the Learning Context the Main Variable? *System*. 2020; 91:102263. (In Eng.) DOI: <https://doi.org/10.1016/j.system.2020.102263>
29. Chostelidou D., Griva E. Measuring the Effect of Implementing CLIL in Higher Education: An Experimental Research Project. *Procedia – Social and Behavioral Sciences*. 2014; 116:2169-2174. (In Eng.) DOI: <https://doi.org/10.1016/j.sbspro.2014.01.538>
30. Martínez Agudo J. de D. To What Extent Do Affective Variables Correlate with Content Learning Achievement in CLIL Programmes? *Language and Education*. 2021; 35(3):226-240. (In Eng.) DOI: <https://doi.org/10.1080/09500782.2020.1833910>
31. Somers T., Llinares A. Students' Motivation for Content and Language Integrated Learning and the Role of Programme Intensity. *International Journal of Bilingual Education and Bilingualism*. 2018. (In Eng.) DOI: <https://doi.org/10.1080/13670050.2018.1517722>
32. Navarro-Pablo M., Gándara Y.L. The Effects of CLIL on L1 Competence Development in Monolingual Contexts. *The Language Learning Journal*. 2020; 48(1):18-35. (In Eng.) DOI: <https://doi.org/10.1080/09571736.2019.1656764>
33. Nieto-Moreno-de-Diezmas E. Acquisition of Reading Comprehension in L1 in Bilingual Programmes of Primary Education. A Comparative Study. *Ocnos*. 2018; 17(1):43-54. (In Eng.) DOI: http://dx.doi.org/10.18239/ocnos_2018.17.1.1471
34. Kaushanskaya M., Yoo J., Marian V. The Effect of Second-Language Experience on Native-Language Processing. *Vigo International Journal of Applied Linguistics*. 2011; 8(1):54-77. (In Eng.)
35. Bruton A. Is CLIL so Beneficial, or Just Selective? Re-Evaluating Some of the Research. *System*. 2011; 39(4):523-532. (In Eng.) DOI: <https://doi.org/10.1016/j.system.2011.08.002>
36. Boncés J.R. Content and Language Integrated Learning (CLIL): Considerations in the Colombian Context. *Gist Education and Learning Research Journal*. 2012; (6):177-189. Available at: <https://latinjournal.org/index.php/gist/article/view/432> (accessed 25.11.2020). (In Eng.)
37. Cenoz J., Genesee F., Gorter D. Critical Analysis of CLIL: Taking Stock and Looking Forward. *Applied Linguistics*. 2014; 35(3):243-262. (In Eng.) DOI: <https://doi.org/10.1093/applin/amt011>
38. Graham K.M., Choi Y., Davoodi A., Razmeh S., Dixon L.Q. Language and Content Outcomes of CLIL and EMI: A Systematic Review. *Latin American Journal of Content & Language Integrated Learning*. 2018; 11(1):19-37. (In Eng.) DOI: <https://doi.org/10.5294/laclil.2018.11.1.2>
39. Doiz A., Lasagabaster D., Sierra J.M. CLIL and Motivation: The Effect of Individual and Contextual Variables. *The Language Learning Journal*. 2014; 42(2):209-224. (In Eng.) DOI: <https://doi.org/10.1080/09571736.2014.889508>
40. Ouazizi K. The Effects of CLIL Education on the Subject Matter (Mathematics) and the Target Language (English). *Latin American Journal of Content & Language Integrated Learning*. 2016; 9(1):110-137. (In Eng.) DOI: <https://doi.org/10.5294/laclil.2016.9.1.5>
41. Lasagabaster D., Sierra J.M. Language Attitudes in CLIL and Traditional EFL Classes. *International Journal of CLIL Research*. 2009; 1(2):4-17. Available at: http://www.laslab.org/wp-content/uploads/language_attitudes_in_clil_and_traditional_efl_classes.pdf (accessed 25.11.2020). (In Eng.)
42. Psaltou-Joycey A., Mattheoudakis M., Thomai A. Language Learning Strategies in CLIL and Non-CLIL Classes: Which Strategies Do Young Learners Claim They Use? In: Conference: Cross-Curricular Approaches to Language Education. Thessaloniki, Greece; 2012. (In Eng.) DOI: <https://doi.org/10.13140/RG.2.1.3240.2088>



43. Pozo E. del CLIL in Secondary Classrooms: History Contents on the Move. In: Tsuchiya K., Pérez Muri- llo M.D. (eds.). *Content and Language Integrated Learning in Spanish and Japanese Contexts*. London: Palgrave Macmillan; 2019. p. 125-151. (In Eng.) DOI: https://doi.org/10.1007/978-3-030-27443-6_6
44. Jäppinen A. Thinking and Content Learning of Mathematics and Science as Cognitive Development in Content and Language Integrated Learning (CLIL): Teaching Through a Foreign Language in Finland. *Language and Education*. 2005; 19(2):147-168. (In Eng.) DOI: <https://doi.org/10.1080/09500780508668671>
45. Surmont J., Struys E., Noort M.V., Craen P.V. The Effects of CLIL on Mathematical Content Learning: A Longitudinal Study. *Studies in Second Language Learning and Teaching*. 2016; 6(2):319-337. (In Eng.) DOI: <https://doi.org/10.14746/ssl.2016.6.2.7>
46. Fleckenstein J., Gebauer S.K., Möller J. Promoting Mathematics Achievement in One-Way Immersion: Performance Development Over Four Years of Elementary School. *Contemporary Educational Psychology*. 2019; 56:228-235. (In Eng.) DOI: <https://doi.org/10.1016/j.cedpsych.2019.01.010>
47. Binterová H., Šulista M. GeoGebra Software Use within a Content and Language Integrated Learning Environment. *European Journal of Contemporary Education*. 2013; 4(2):100-116. (In Eng.) DOI: <https://doi.org/10.13187/ejced.2013.4.100>
48. Serra C. Assessing CLIL at Primary School: A Longitudinal Study. *International Journal of Bilingual Education and Bilingualism*. 2007; 10(5):582-602. (In Eng.) DOI: <https://doi.org/10.2167/beb461.0>
49. Piesche N., Jonkmann K., Fiege C., Keßler J. CLIL for All? A Randomised Controlled Field Experiment with Sixth-Grade Students on the Effects of Content and Language Integrated Science Learning. *Learning and Instruction*. 2016; 44:108-116. (In Eng.) DOI: <https://doi.org/10.1016/j.learninstruc.2016.04.001>
50. Fernández-Sanjurjo J., Fernández-Costales A., Blanco J. M. Analysing Students' Content-Learning in Science in Clil vs. Non-Clil Programmes: Empirical Evidence from Spain. *International Journal of Bilingual Education and Bilingualism*. 2019; 22(6):661-674. (In Eng.) DOI: <https://doi.org/10.1080/13670050.2017.1294142>
51. Mattheoudakis M., Alexiou T., Laskaridou C. To CLIL or Not to CLIL? The Case of the 3rd Experimental Primary School in Evosmos. In: *Major Trends in Theoretical and Applied Linguistics*. Vol. 3. Berlin, Boston: De Gruyter; 2015. (In Eng.) DOI: <https://doi.org/10.2478/9788376560915.p13>
52. Meyerhöffer N., Dreesmann D.C. The Exclusive Language of Science? Comparing Knowledge Gains and Motivation in English-bilingual Biology Lessons between Non-selected and Preselected Classes. *International Journal of Science Education*. 2019; 41(1):1-20. (In Eng.) DOI: <https://doi.org/10.1080/09500693.2018.1529446>
53. Xanthou M. The Impact of CLIL on L2 Vocabulary Development and Content Knowledge. *English Teaching: Practice and Critique*. 2011; 10(4):116-126. Available at: <https://files.eric.ed.gov/fulltext/EJ962609.pdf> (accessed 25.11.2020). (In Eng.)
54. Isidro X.S., Lasagabaster D. The Impact of CLIL on Pluriliteracy Development and Content Learning in a Rural Multilingual Setting: A Longitudinal Study. *Language Teaching Research*. 2019; 23(5):584-602. (In Eng.) DOI: <https://doi.org/10.1177/1362168817754103>
55. Lialikhova D. Triggers and Constraints of Lower Secondary Students' Willingness to Communicate Orally in English in a CLIL Setting in the Norwegian Context. *Journal of Immersion and Content-Based Language Education*. 2018; 6(1):27-56. (In Eng.) DOI: <https://doi.org/10.1075/jicb.16013.lia>
56. Sylvén L.K. CLIL in Sweden – Why Does It not Work? A Metaperspective on CLIL across Contexts in Europe. *International Journal of Bilingual Education and Bilingualism*. 2013; 16(3):301-320. (In Eng.) DOI: <https://doi.org/10.1080/13670050.2013.777387>
57. Dallinger S., Jonkmann K., Hollm J., Fiege C. The Effect of Content and Language Integrated Learning on Students' English and History Competences – Killing Two Birds with One Stone? *Learning and Instruction*. 2016; 41:23-31. (In Eng.) DOI: <https://doi.org/10.1016/j.learninstruc.2015.09.003>
58. Bellés-Calvera L. Teaching Music in English: A Content-Based Instruction Model in Secondary Education. *Latin American Journal of Content & Language Integrated Learning*. 2018; 11(1):109-139. (In Eng.) DOI: <https://doi.org/10.5294/laclil.2018.11.1.6>
59. Salvador-García C., Capella-Peris C., Chiva-Bartoll O., Ruiz-Montero P.J. 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*. 2020; 11. (In Eng.) DOI: <https://doi.org/10.3389/fpsyg.2020.00578>
60. Pérez Cañado M.L. From the CLIL Craze to the CLIL Conundrum: Addressing the Current CLIL Controversy. *Bellaterra Journal of Teaching & Learning Language & Literature*. 2016; 9(1):9-31. (In Eng.) DOI: <https://doi.org/10.5565/rev/jtl3.667>
61. Banegas D.L. A Review of "CLIL: Content and Language Integrated Learning". *Language and Education*. 2011; 25(2):182-185. (In Eng.) DOI: <https://doi.org/10.1080/09500782.2010.539045>

62. Hak T., Rhee H.V., Suurmond R. How to Interpret Results of Meta-Analysis. *SSRN Electronic Journal*. 2018. (In Eng.) DOI: <https://doi.org/10.2139/ssrn.3241367>
63. Sepešiová M. New Perspectives in Pre-service Teacher Training in CLIL via VLE. In: *Language in Focus: Contemporary Means and Methods in ELT and Applied Linguistics*. LIF – Language in Focus. Tallinn; 2018. p. 515-533. (In Eng.)
64. Papaja K. The Role of a Teacher in a CLIL Classroom. *Glottodidactica. An International Journal of Applied Linguistics*. 2013; 40(1):147-154. (In Eng.) DOI: <https://doi.org/10.14746/gl.2013.40.1.11>
65. Banegas D.L. CLIL Teacher Development: Challenges and Experiences. *Latin American Journal of Content & Language Integrated Learning*. 2012; 5(1):46-56. (In Eng.) DOI: <https://doi.org/10.5294/lacil.2012.5.1.4>
66. Van Kampen E., Admiraal W., Berry A. Content and Language Integrated Learning in the Netherlands: Teachers' Self-Reported Pedagogical Practices. *International Journal of Bilingual Education and Bilingualism*. 2018; 21(2):222-236. (In Eng.) DOI: <https://doi.org/10.1080/13670050.2016.1154004>
67. McDougald J.S. Teachers' Attitudes, Perceptions and Experiences in CLIL: A Look at Content and Language. *Colombian Applied Linguistics Journal*. 2015; 17(1):25-41. (In Eng.) DOI: <https://doi.org/10.14483/udistrital.jour.calj.2015.1.a02>
68. Cross R. Creative in Finding Creativity in the Curriculum: The CLIL Second Language Classroom. *The Australian Educational Researcher*. 2012; 39(4):431-445. (In Eng.) DOI: <https://doi.org/10.1007/s13384-012-0074-8>
69. Hanesová D. Development of Critical and Creative Thinking Skills in CLIL. *Journal of Language and Cultural Education*. 2014; 2(2):33-51. (In Eng.)
70. Kosturková M. The Level of Critical Thinking in Students of Education. *Lifelong Learning*. 2014; 4(1):45-61. (In Eng.) DOI: <https://doi.org/10.11118/lifele2014040145>

Submitted 12.03.2021; approved after reviewing 15.04.2021; accepted for publication 23.04.2021.
 Поступила 12.03.2021; одобрена после рецензирования 15.04.2021; принята к публикации 23.04.2021.

About the author:

Ivana Cimermanová, Associate Professor of the Institute of British and American Studies, University of Presov (1, 17 novembra St., Presov 080 01, Slovakia), **ORCID:** <https://orcid.org/0000-0002-8063-8949>, **Scopus ID:** [57193775555](https://orcid.org/0000-0002-8063-8949), ivana.cimermanova@unipo.sk

The author has read and approved the final manuscript.

Об авторе:

Цимерманова Ивана, доцент Института британских и американских исследований Прешовского университета (080 01, Словакия, г. Прешов, ул. 17 ноября, д. 1), **ORCID:** <https://orcid.org/0000-0002-8063-8949>, **Scopus ID:** [57193775555](https://orcid.org/0000-0002-8063-8949), ivana.cimermanova@unipo.sk

Автор прочитал и одобрил окончательный вариант рукописи.