TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

Investigating the Relationship Between CLIL and L1 Ability in a Finnish Education Context

Peter Launonen

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. B OSA – TOM. 743 | HUMANIORA | TURKU 2025

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

INVESTIGATING THE RELATIONSHIP BETWEEN CLIL AND L1 ABILITY IN A FINNISH EDUCATION CONTEXT

Peter Launonen

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. B OSA – TOM. 743 | HUMANIORA | TURKU 2025

University of Turku

Faculty of Humanities
School of Languages and Translation Studies
Applied Language Studies
Doctoral Programme in Languages and Translation Studies (Utuling)

Supervised by

Professor Minna Maijala
University of Turku
Turku, Finland

Senior Lecturer Anssi Roiha
University of Jyväskylä
Jyväskylä, Finland

Reviewed by

Professor Tarja Nikula-Jäntti
University of Jyväskylä
Jyväskylä, Finland

Professor Ana Llinares García
Universidad Autónoma de Madrid
Madrid, Spain

Opponent

Professor Tarja Nikula-Jäntti
University of Jyväskylä
Jyväskylä, Finland

ISBN 978-952-02-0376-4 (PRINT)
ISBN 978-952-02-0377-1 (PDF)
ISSN 0082-6987 (Print)
ISSN 2343-3191 (Online)
Painosalama, Turku, Finland 2025

UNIVERSITY OF TURKU
Faculty of Humanities
School of Languages and Translation Studies
Applied Language Studies
PETER LAUNONEN: Investigating the Relationship Between CLIL and L1 Ability in a Finnish Education Context
Doctoral Dissertation, 163 pp.
Doctoral Programme in Languages and Translation Studies (Utuling)
November 2025

ABSTRACT

This study analyses differences in first language (L1) ability between students in content and language integrated learning (CLIL) and students in mainstream education (non-CLIL). The study provides new insights by using multiple methods to analyse this relationship in a Finnish secondary education context. The dissertation comprises three sub-studies: a small-scale case study ($n = 4$) (*Article I*), a study concerning L1 oral production ($n = 31$) (*Article II*), and a study focusing on L1 writing and speaking ($n = 30$) (*Article III*).

The participants in all sub-studies were either 14 or 15-year-old L1 Finnish speakers who were enrolled in either CLIL or mainstream education in Finland. The data were analysed using quantitative methods in order to assess group-level differences in L1 ability and identify correlations between L1 ability and individual differences in socioeconomic status (SES) and academic achievement in L1 as a subject. Specifically, *Article I* explores approaches to assessing differences in CLIL and non-CLIL students' performance in both Finnish and English. *Article II* examines lexical access and retrieval in L1 oral production using a verbal fluency task and a picture-naming task. *Article III* analyses students' L1 writing via a short written task based on a prompt question concerning the environment and their L1 speaking via an elicitation task based on a short comic strip.

The results suggested that CLIL students are not disadvantaged in terms of L1 ability in Finnish secondary education CLIL contexts, where L1 use is typically not avoided or forbidden. The correlation analyses also underscored the importance of accounting for SES in the analysis of L1 ability in both CLIL and mainstream education. While the non-CLIL students' results were suggestive of a positive relationship between SES and L1 ability, the CLIL students' results in L1 ability did not clearly reflect differences in SES in either direction. These findings add to the growing body of research suggesting that CLIL may help promote equality by mitigating the effect that SES has on the development of language skills.

KEYWORDS: CLIL, L1, bilingual education, bilingualism, multilingualism, socioeconomic status

TURUN YLIOPISTO
Humanistinen tiedekunta
Kieli- ja käännöslaitos
Soveltava kielitiede
PETER LAUNONEN: Kaksikielisen opetuksen ja ensikielen taidon välisen suhteen tarkastelu suomalaisessa koulukontekstissa
Väitöskirja, 163 s.
Tohtoriohjelma Utuling
Marraskuu 2025

TIIVISTELMÄ

Tämän väitöskirjan tavoitteena on analysoida eroja ensikielen (L1) taidoissa kaksikieliseen opetukseen (*content and language integrated learning*, CLIL) ja perinteiseen opetukseen (ei-CLIL) osallistuvien oppilaiden välillä. Tutkimus tuo uusia näkemyksiä yhdistämällä useita menetelmiä tämän ilmiön analysoimiseksi suomalaisessa koulukontekstissa. Väitöskirja koostuu kolmesta osatutkimuksesta. Ensimmäinen osatutkimus on pienimuotoinen tapaustutkimus (n = 4) (*Artikkeli I*). Toinen osatutkimus käsittelee ensikielen suullista tuottamista (n = 31) (*Artikkeli II*) ja kolmas osatutkimus ensikielellä kirjoittamista ja puhumista (n = 30) (*Artikkeli III*).

Osallistujat kaikissa osatutkimuksissa olivat joko 14- tai 15-vuotiaita suomea ensikielenään puhuvia oppilaita, jotka osallistuivat joko CLIL-opetukseen tai perinteiseen opetukseen Suomessa. Aineisto analysoitiin määrällisin menetelmin, joilla tutkittiin ryhmätason eroja ensikielen taidoissa sekä korrelaatioita niiden ja yksilöllisten erojen välillä esimerkiksi sosioekonomisessa asemassa (*socioeconomic status*, SES) ja akateemisessa suoriutumisessa äidinkielen oppiaineessa. Artikkeli I:ssä kartoitettiin menetelmiä, joiden avulla voidaan arvioida eroja CLIL- ja ei-CLIL- oppilaiden suoriutumisessa sekä suomen että englannin kielessä. Artikkeli II:ssa tutkittiin ensikielen suullisen sananhaun ja tunnistuksen tehokkuutta käyttäen sanasujuvuustehtävää ja kuvannimeämistehtävää. Artikkeli III:ssa analysoitiin oppilaiden ensikielellä kirjoittamista ja puhumista lyhyellä ympäristöä käsittelevällä kirjoitustehtävällä ja lyhyellä sarjakuvaan perustuvalla puhetehtävällä.

Tulokset osoittivat, että CLIL-oppilaat eivät ole heikommassa asemassa L1-taitojen osalta suomalaisessa yläkoulussa toteutuvassa CLIL-opetuksessa, jossa L1-käyttöä ei tyypillisesti rajoiteta tai kielletä. Korrelaatioanalyysien tulokset korostivat myös SES:n huomioimisen merkitystä sekä CLIL- että ei-CLIL-oppilaiden L1-taitojen analyysissa. Vaikka ei-CLIL-oppilaiden tuloksissa havaittiin positiivinen suhde SES:n ja L1-taitojen välillä, CLIL-oppilaiden L1-taidot eivät selkeästi heijastaneet eroja SES:ssä kumpaankaan suuntaan. Tutkimuksen tulokset täydentävät kasvavaa tutkimustietoa, jonka mukaan CLIL saattaa edistää tasa-arvoisuutta lieventämällä SES:n vaikutusta kielitaidon kehittymiseen.

ASIASANAT: CLIL, L1, kaksikielinen opetus, kaksikielisyys, monikielisyys, sosioekonominen asema

Acknowledgements

Completing a doctoral degree is probably never a linear process, and mine is no different. Having worked full-time as a teacher during my doctoral studies and research, I can genuinely say that my PhD would not have been completed had it not been for the guidance, encouragement and opportunities afforded to me throughout this journey. To begin with, I would like to thank the external examiners of this dissertation, Professor Tarja Nikula-Jäntti and Professor Ana Llinares García. Their critical reflections prompted me to revisit various theoretical considerations and clarify my thinking concerning elements of the discussion and conclusions. I am especially grateful to Professor Nikula-Jäntti for agreeing to act as my opponent in the public defence.

I would also like to extend my deepest gratitude to my supervisors, Professor Minna Maijala and Senior Lecturer Anssi Roiha. Minna and Anssi not only co-authored two articles with me, but they were instrumental in giving me critical and constructive feedback on my data analysis and writing, while also providing timely practical suggestions when challenges emerged along the way. Minna, thank you for helping me find ways to keep moving forward during difficult moments – one *nytkähdys* at a time. Anssi, I would like to thank you for always taking the time to provide invaluable input on all matters concerning CLIL and pedagogy, to name only two areas. I hope to continue collaborating with both of you for many years to come.

Another important contributor that helped me find a path through this non-linear process is the Utuling doctoral programme itself. During my doctoral research, I have had the opportunity to take courses in university pedagogy, research methods, ethics and even a course designed to help candidates write the summary section of the dissertation. Having this kind of support, education and opportunities for self-development has given real meaning to the term *doctoral training*. I would also like to extend a special thank you to Meri Heinonen – the coordinator of the Utuling programme – who works hard to ensure that we, the doctoral students, have everything we need when we need it. I also owe a debt of gratitude to the students, teachers and staff at the secondary schools where I collected data. This research would not have been possible without their participation and support.

On a personal note, I am grateful to those close to me for their support in reaching this important milestone in my academic journey. Firstly, I could not have made it to this point without the love and encouragement of my wife, Gülnara (Gulja). Throughout my time as a doctoral researcher, you have always been there to listen to my grievances, offer words of wisdom and motivation, and help me stay focused on the end goal. I would also like to thank my parents (and their spouses), my sisters (and their families and partners) in Australia, as well as my extended family in Finland and my in-laws in Saint Petersburg, who have all provided support and encouragement in different ways and at different times – a heartfelt thank you to all of you as well.

November, 2025
Peter Launonen

Table of Contents

List of Tables and Figures

Tables

Figures

List of Original Publications

This dissertation is based on three original publications which are referenced in the text as follows:

Article I Launonen, P., Roiha, A., & Maijala, M. (2024). Exploring the Relationship between CLIL and L1 Ability in Finland: Analyzing Written and Oral Production. *Latin American Journal of Content & Language Integrated Learning, 15*(2), e1528. https://doi.org/10.5294/laclil.2022.15.2.8

Article II Launonen, P. (2025). Examining L1 ability in a Finnish secondary education context: a comparison of CLIL and non-CLIL students' oral production. *Journal of Immersion and Content-Based Language Education. Advance online publication*. https://doi.org/10.1075/jicb.24016.lau

Article III Launonen, P., Roiha, A., & Maijala, M. (2025, under review). The effect of CLIL on L1 ability: Evidence from a Finnish secondary education context.

1 Introduction

The ability to use a language effectively enables people to function in various aspects of their daily lives. In the European Union (EU), having proficiency in several languages has become increasingly important considering the emphasis given by the EU to multilingualism as an aspect of European identity (European Commission, 1995). Although the EU is linguistically diverse in light of its 24 official languages (European Union, n.d.), some European countries continue to be influenced by monolingual nationalism (Guglielmi, 2020), whereby the existence of a predominant national language is considered important. However, the linguistic context in other countries, such as Luxembourg, is considered to reflect multilingualism (de Bres et al., 2019). The EU's posture on multilingualism also alludes to the role played by education systems, asserting that citizens should be able to communicate in at least two languages as well as their first language (L1) by the end of their compulsory education (European Parliament & Council of the European Union, 1998). This focus on multilingualism is not only relevant at a political and societal level, but it also reflects a changing trend in applied linguistics research and education, whereby language users and students have begun to be viewed through a multilingual rather than a monolingual perspective (Conteh & Meier, 2014; May, 2014).

Although proficiency in another language can be facilitated by various factors (e.g., family language policies, immigration, etc.), language skills are commonly developed through education via the use of different pedagogical approaches. The implementation of such an approach can generally be considered a success if it has achieved its stated educational aims (e.g., the learning outcomes have been met). This is also true for implementations of content and language integrated learning (CLIL), where the learning outcomes reflect students' development of content knowledge as well as proficiency in a language that is typically not the learners' L1. In CLIL, content subjects (e.g., science, geography) are taught via the medium of a target language (e.g., English), meaning that CLIL students invariably participate in less learning via their L1 than non-CLIL students. This has led to calls for further research on the possible impact that CLIL may have on L1 ability (e.g., Cenoz et al., 2014; Dalton-Puffer, 2011). Taking a societal perspective,

understanding what factors may affect L1 ability is important in the EU in light of the existing challenges related to, for example, linguistic diversity (e.g., Grindheim & Lohndal, 2008; Soldat-Jaffe, 2014), social inclusion (e.g., Bian, 2017; Yagmur, 2019) and the spread of English (e.g., Gnutzmann et al., 2015). In Finland, specifically, balancing conflicting linguistic priorities is also a challenge because the country has two official languages (i.e., Finnish and Swedish) – both of which are also official languages of the EU – and because there is already a widespread use of English in Finland (e.g., Jódar Sánchez & Tuomainen, 2014; Peterson, 2022). Therefore, seeking to understand the influence that a pedagogical approach such as CLIL may have on students' L1 ability is a relevant field of inquiry that contributes to a broader discussion about language policy and education at both the national and EU levels.

1.1 Background of the study

The present study is focused on an educational context in Finland where CLIL is considered to play a role in the development of students' ability in their second language (L2), typically English, as well as their L1 Finnish. This is reflected in the term *bilingual education*, which is the English translation of the Finnish term (*kaksikielinen opetus*) used to refer to CLIL in Finland (e.g., Nikula et al., 2022; Wewer, 2020). Given that CLIL emphasises a dual focus (i.e., content and language), it is unsurprising that a considerable amount of CLIL studies have focused on exploring the effects of this approach on L2 learning (e.g., Goris et al., 2019; Pérez Cañado, 2011; Sato & Hemmi, 2022) and content learning (e.g., Fernández-Sanjurjo et al., 2017; Meyerhöffer & Dreesmann, 2019; Pérez Cañado, 2018). However, as mentioned previously, more research is needed on the possible effects of CLIL on L1 ability.

Before outlining the focus of the present study in detail, it is worth acknowledging previous research and findings concerning L1 ability in other similar pedagogical approaches, such as immersion. The effect of immersion education on L1 ability has been studied in, for example, Canada and Finland. In research focused on early Canadian immersion programmes in which English-speaking students participated in French-medium education, students in immersion were not found to be disadvantaged in L1 ability compared to students in mainstream education (see Cummins, 1998, for a summary). Similarly, Swedish-medium immersion programmes in Finland have also been found not to have a negative effect on L1 ability (Elomaa, 2000). Although CLIL and immersion undoubtedly share many similarities (Brown & Bradford, 2017; Cenoz, 2015), they can also be distinguished as having emerged at different times and in considerably different overarching contexts (Nikula & Mård-Miettinen, 2014). Additionally, CLIL is, particularly in the

European context, largely associated with implementations in which the target language is English (Lasagabaster & Sierra, 2010), contrary to the aforementioned immersion programmes. For these reasons, additional research is needed specifically in connection with CLIL implementations, which intersects with the overarching aim of the present study. Moreover, to the researcher's knowledge, this is the first study concerning this area in which the relationship between CLIL and L1 ability is contextualised within a broader discussion about the advantages and costs of bi- or multilingualism (see Section 2.4).

1.2 Aims of the study

This dissertation aims to shed light on the relationship between CLIL and L1 ability in a Finnish education context. The overarching aim and sub-aims of the study are illustrated in Figure 1. The purpose of sub-aim 1 is to explore and test methods for addressing the overarching research aim of the study. The second and third sub-aims directly address the relationship between CLIL and different aspects of students' L1 ability. While sub-aim 2 is focused on students' oral production as measured by their ability to access and retrieve (and produce) lexis, sub-aim 3 targets students' spoken and written ability in response to elicitation tasks. As shown in Figure 1 (on the next page), sub-aim 1 is addressed in Article I, whereas sub-aim 2 is addressed in Articles I and II, and sub-aim 3 in Articles I and III (see also Section 4, Table 8). Additionally, the empirical findings of all the articles address the overarching aim of the study, which is focused on addressing a gap regarding the relationship between CLIL and L1 ability, specifically in a Finnish education context. The study also contributes to a broader discussion about the role that bilingual education (i.e., CLIL) plays in the development of bi- or multilingualism in a multilingual context (i.e., the EU).

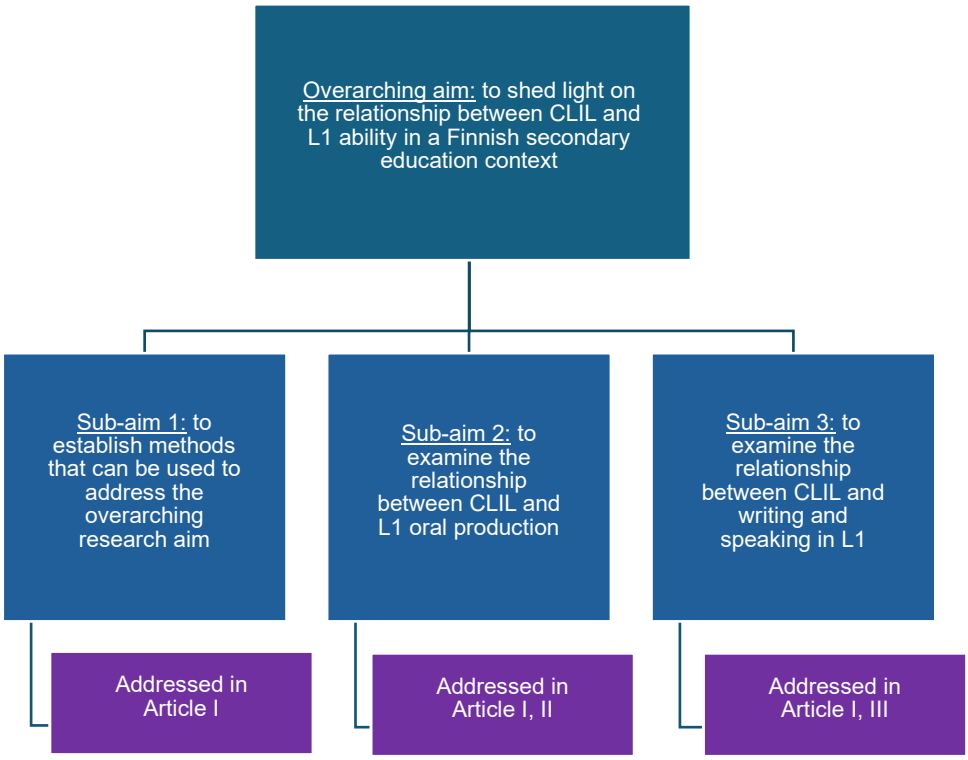

Figure 1. Relationship between overarching aim, sub-aims and articles in this study.

The compilation part of this dissertation comprises four chapters. Firstly, the theoretical and conceptual framework for the study is presented in Chapter 2. This chapter details the theoretical underpinnings of CLIL, critically reviews previous research concerning the relationship between CLIL and L1 ability, and contextualises the present study within a broader discussion about bi- and multilingualism. Secondly, the processes related to data collection and analysis are outlined in Chapter 3. This includes a detailed description of the modifications made in data collection and analysis methods in light of insights gained by addressing the first sub-aim (see Figure 1). Thirdly, the empirical findings emanating from the sub-studies are presented succinctly in Chapter 4 and then examined in connection with previous research in Chapter 5, along with implications for affected stakeholders. Limitations of the study are identified and raised in this chapter, while possible avenues for future research are also explored and conclusions provided.

2 Theoretical and Conceptual Framework

The key concepts related to this research, such as CLIL and bi- or multilingualism, are explored in this chapter. In addition, the previous research in this area is analysed critically, forming the foundation for addressing the aims and choosing the methods for use in the present study.

2.1 Content and language integrated learning

The pedagogical approach known as content and language integrated learning (CLIL) emerged in Europe in the 1990s, although it is now widely employed in various contexts around the world (e.g., Banegas, 2019; Itoi, 2024; Turner, 2013). The fundamental premise of CLIL is the teaching of a subject (e.g., science) in a target language with the aim of students learning the content while also improving their proficiency in the target language. That said, this principle of integrating language with content is not a new phenomenon. Since 1965, French immersion programmes in Canada have aimed to promote the development of anglophone students' French language skills via French-medium instruction of curriculum content (Cummins & Swain, 1986). In light of the success of these programmes, Swedish-language immersion programmes were introduced in Finland in 1987 (Björklund, 2019), and then CLIL education also commenced in Finland after the Basic Education Act was amended in 1991 (Laki peruskoululain muuttamisesta 261/1991).

Although CLIL and immersion are often considered distinct from one another[1], the approaches share considerable conceptual overlap and finding a universally accepted distinction between them has proved challenging (Brown & Bradford, 2017; Cenoz, 2015; Dalton-Puffer et al., 2014; Nikula & Mård-Miettinen, 2014). Consequently, it is not within the scope of the current study to attempt to distinguish

[1] In Finland, immersion education is often referred to as *kielikylpy* (literally *language bath*), whereas CLIL is typically known as *kaksikielinen opetus* (literally *bilingual teaching*).

CLIL from immersion and related concepts such as English-medium instruction (EMI) and content-based instruction (CBI). Instead, any specific instances of implementations that are discussed reflect the labels they have been given either in research or by the corresponding educational authority.

On a conceptual level, one of the pillars of CLIL education is a recognition of the role of the 4Cs: content, communication, cognition and culture. Rather than existing in isolation, in CLIL there is considered to be an "interrelationship between content (subject matter), communication (language), cognition (learning and thinking) and culture (social awareness of self and 'otherness')" (Coyle, 2007, p. 550). As depicted in Figure 2, CLIL is taken to function on the basis of a symbiosis of these four elements, while also acknowledging the importance of *context* in learning (Coyle et al., 2010). As discussed by Llinares (2024), the communicative context in CLIL promotes the emergence of functional language choices being made by students (and teachers), facilitating the creation of meaning in the classroom.

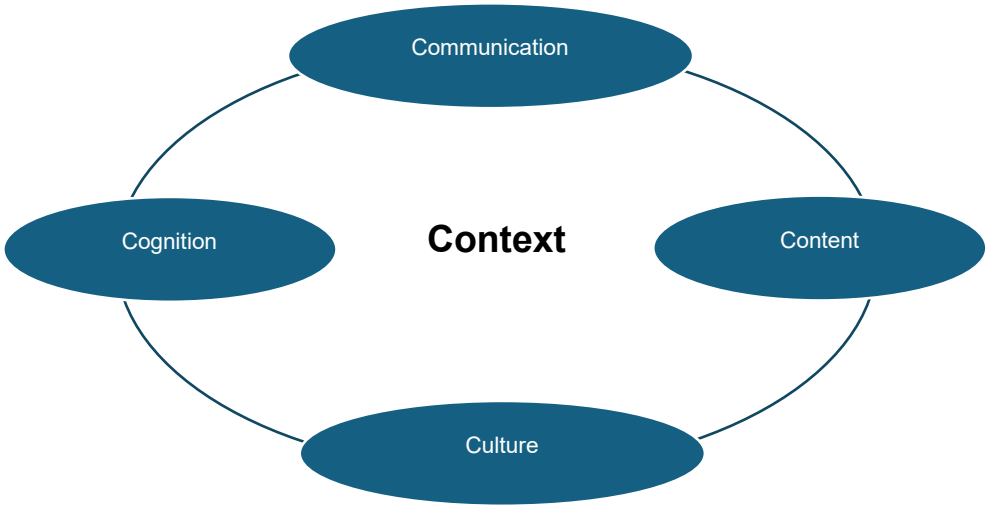

Figure 2. The 4Cs framework of CLIL (adapted from Coyle et al., 2010, p. 41).

The intertwined roles of language and context are reflected in the language triptych (Coyle et al., 2010), which was later expanded by Banegas and Mearns (2023) into the language quadriptych. The language quadriptych distinguishes students' language needs across different learning contexts that are relevant to CLIL education; namely, language *of* learning (i.e., subject-specific language needed to communicate about the content and topics), language *for* learning (i.e., classroom language learners need to negotiate meaning and complete tasks), language *through* learning (i.e., language that arises from students' use of the vehicular language for

learning), and language *about* learning (i.e., language that learners use to reflect on and discuss their learning). The type of language that is developed in (and needed for) CLIL can also be conceptualised in accordance with the distinction made between *basic interpersonal communication skills* (BICS) and *cognitive academic language proficiency* (CALP) (see Cummins, 2000). The use of BICS is considered to comprise informal, contextualised language that can be learned relatively quickly (e.g., *Write your name at the top of the page*), whereas CALP refers to the language required to understand and communicate abstract ideas where the context is not immediately clear (e.g. *Evaluate the issue and justify your perspective*). Another important dimension related to language in CLIL classes concerns cognitive discourse functions (CDFs). As outlined by Dalton-Puffer (2016), the CDF construct encompasses seven CDFs that are represented by the following verbs: classify, define, describe, evaluate, explain, explore and report. Representing communicative intentions that concern both learners and teachers, CDFs emerge in order to help humans "deal with cognitive content for the purpose of learning, representing and exchanging knowledge" (Dalton-Puffer, 2016, p. 31). Recent studies have highlighted the pedagogical benefits of integrating CDFs into CLIL teaching (Gerns, 2023a; Gerns, 2023b), while also identifying differences in students' use of CDFs across subjects (Evnitskaya & Dalton-Puffer, 2020; Llinares & Nikula, 2023).

On the surface, one of the main differences between CLIL and mainstream education is the use of a target language (e.g., an L2) for instruction and as part of the aforementioned language quadriptych. From a cognitive or psycholinguistic perspective, this naturally leads to CLIL students having a higher amount of target-language *input* (see Krashen, 1982; VanPatten & Oikkenon, 1996) and *output* (see Swain, 1985) than students in mainstream education. However, given the dialogic nature of the relationship between the teacher and the students and amongst the students themselves, *interaction* (see Long, 1996) is also considered to play an important role in second language acquisition (SLA) in CLIL (Dalton-Puffer, 2011; Klewitz, 2021). Adopting a sociocultural lens, this process of acquisition (or learning) can also be conceptualised as arising through *social interaction*, *mediation* and *internalisation* in the learner's Zone of Proximal Development (ZPD) (see Vygotsky, 1978). In terms of foreign language pedagogy, CLIL shares principles with some other approaches within the field of teaching English to speakers of other languages (TESOL). For instance, the focus on communication and context also exists in communicative language teaching (CLT) (e.g., Chunliu & Guangsheng, 2025) and task-based learning (TBL) (e.g., Alasal, 2025). Nevertheless, as a form of bilingual education comprising a dual focus on content and language, CLIL can be largely distinguished from other approaches in TESOL in which the focus is almost exclusively on language and communication, with minimal emphasis placed on content learning.

An implementation of CLIL education can be described in terms of its pedagogical focus on a continuum with *content-driven* on one end and *language-driven* on the other (Met, 1999), terms which have also been conceptualised as *hard CLIL* and *soft CLIL*, respectively (e.g., Ikeda, 2022). A content-driven implementation (i.e., hard CLIL) is one in which the outcomes, teaching and assessment are oriented more towards content (e.g., geography) than language (e.g., English), whereas in a language-driven CLIL class (i.e., soft CLIL) the overarching pedagogical focus is on the development of students' target language skills, for which content is used as a tool. Consequently, soft CLIL can be offered as part (or instead) of language lessons, unlike hard CLIL. In addition to the content-language continuum, the present study proposes that a continuum can also be considered to exist with respect to the amount of L1 use in a CLIL implementation. Given that one of the central ideas underpinning CLIL is the use of another language (e.g., an L2), and that the purpose for doing so is to promote the development of students' language skills, the amount of L1 use that teachers (and students) can or should employ in a given CLIL implementation merits consideration. On the one hand, having no L1 use in teaching allows for total immersion and extensive target language exposure for students, a practice which is consistent with some approaches in TESOL such as the direct method (Benati, 2018). On the other hand, there may also be pedagogical benefits associated with using the students' existing language skills (e.g., L1) to facilitate learning, as has been discussed extensively in research on translanguaging (e.g., García, 2011; Itoi, 2024; Nikula & Moore, 2016; Tsuchiya, 2019). In brief, translanguaging refers to "the complex language practices of plurilingual individuals and communities, as well as the pedagogical approaches that use those complex practices" (García & Li, 2014, p. 20). As alluded to in Figure 3 (on the next page), the decisions taken in relation to the emphasis on content or language, as well as L1 use in class, can have considerable pedagogical implications for a given CLIL implementation.

Although all CLIL implementations reflect some shared principles, such as the use of another language as the medium of instruction and an emphasis on the 4Cs, there is also considerable variation in the overarching contexts within which CLIL education takes place, even within the EU alone (see, e.g., Gülle & Nikula, 2024). For instance, there are CLIL implementations in regions of Spain that have been categorised as either monolingual (Pérez Cañado, 2018) or multilingual (San Isidro & Lasagabaster, 2018). Differences also exist in the amount of target language use in class (or in a curriculum) across implementations or education systems. There are CLIL contexts in which the percentage of target-language instruction has been deemed to be, for example, 20% (Merisuo-Storm, 2007), 50% (Admiraal et al., 2006) or even close to 100% (Lim Falk, 2019).

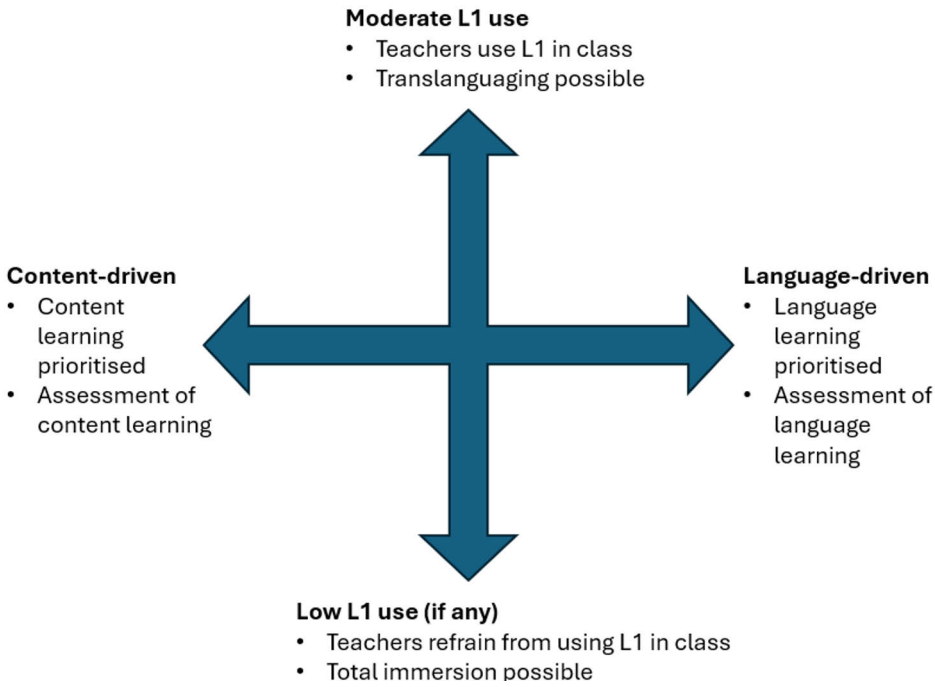

Figure 3. The content-language and L1 use continua.

It is worth noting that while the vehicular language in European CLIL contexts is typically English, there are also CLIL implementations in which the target language is not English (e.g., Bulté et al., 2022; Serra, 2007). However, particularly in those cases where English is the vehicular language, there are additional contextual considerations that may affect the development of language skills differently in some implementations compared to others. Firstly, the level of extramural exposure to English is not comparable between, for example, Spain and Finland. In Spain, media such as television and cinema have traditionally been dubbed into Spanish (Matamala et al., 2017) instead of opting for subtitles, which has been the main approach in Finland (Schurz & Sundqvist, 2022). Apart from such audiovisual input, extramural exposure includes other formats, such as video games and social media (Sundqvist, 2024). Although studies comparing Finland and Spain directly are scarce, teenage Spanish students have been found to have less extramural exposure to English than teenage students in Denmark (Muñoz & Cadierno, 2021), another Nordic country. It is worth considering the influence of such differences in exposure to English in different contexts on students' overall English-learning journey, which takes place partly via formal education and partly outside the classroom. Secondly, the use of English as a *lingua franca* (ELF) varies considerably across countries even within the EU. While this practice has been largely adopted in,

for example, Finland and Sweden (Peterson & Beers, 2023), the process of adopting English for such purposes in Spain has been slow (Tabuenca-Cuevas, 2016). To sum up, despite the theoretical and methodological similarities across CLIL implementations, the inherent contextual differences that exist render direct comparisons challenging, even among implementations from within a single region (e.g., the EU). For this reason, it is crucial to consider the contextual factors of each implementation carefully when aiming to draw conclusions about the role of CLIL in, for example, the development of students' language skills.

2.2 Research on the relationship between CLIL and L1 ability

Since the emergence of CLIL in the 1990s, researchers have expressed an interest in understanding the impact that CLIL education may have on the development of students' L1 ability. For instance, Airey (2004) discussed the increasing tendency to utilise English instead of Swedish for content teaching in Sweden, thereby raising concerns about the possible impact of CLIL on both L1 ability and content learning. Although the need for further research on the relationship between CLIL and L1 ability has been reiterated by various researchers (e.g., Cenoz et al., 2014; Dalton-Puffer, 2011; Pérez Cañado, 2011; San Isidro & Lasagabaster, 2018), the accumulated research in this area remains meagre and has not led to a clear understanding of the effect of CLIL education on L1 ability. This discrepancy in results is somewhat unsurprising, given the aforementioned contextual variation that exists among CLIL implementations.

Early studies aiming to shed light on the relationship between CLIL and L1 ability were conducted in a variety of European contexts. For instance, Seikkula-Leino (2007) investigated the effect of CLIL on students' L1 ability (i.e., Finnish) in a primary education context in Finland by evaluating school success using a test covering L1 writing and reading, together with school grades, while also considering verbal and non-verbal intelligence. In that CLIL implementation, 40-70% of instruction was delivered in an L2 (i.e., English). The study found that CLIL students were not disadvantaged in terms of content learning or L1 ability, despite indicating that the CLIL group had fewer overachievers (i.e., students whose performance exceeded expectations based on their level of intelligence) than the non-CLIL group. In another study in Finland, Merisuo-Storm (2007) explored the impact of CLIL on L1 literacy skills, among other phenomena, in a primary education setting. In that study, the development of students' literacy skills was analysed at three intervals using tests aimed at highlighting differences in reading accuracy, speed and comprehension. In comparing a CLIL group that received 20% of its instruction in English with a control group only taught in Finnish, she found that the CLIL students

performed at least as well as the non-CLIL students in both L1 reading speed and comprehension.

In the Netherlands, Admiraal et al. (2006) investigated L1 ability in a CLIL context involving Dutch students in the first four years of secondary education. In that CLIL provision, 50% of the lessons were taught in an L2 (i.e., English). In order to gauge L1 ability, students' performance in end-of-year exams in their L1 (i.e., Dutch) as a subject was analysed in connection with their participation in CLIL. The results suggested that CLIL education did not have a negative effect on students' performance in their exams in Dutch. Although these early studies all suggested that CLIL students are not at a disadvantage in terms of L1 ability compared to non-CLIL students, it is important to note that there were differences in the approaches used to assess L1 (e.g., productive skills, receptive skills, performance in school exams, school grades), not all key variables, such as socioeconomic status (SES), were considered in all the studies, and there was considerable contextual variation among the CLIL implementations (e.g., primary or secondary education, the amount of L2 use).

As the accumulated CLIL research has continued to expand, a number of criticisms have emerged in terms of research aims and the choice of variables included in studies. For instance, Pérez Cañado (2011) raised various shortcomings of previous CLIL research, calling for future studies to guarantee homogeneity of samples (i.e., comparability of CLIL and non-CLIL groups on the basis of individual differences) and explore additional variables, such as SES, which has often been measured in terms of parents' education level (e.g., Anghel et al., 2015; Pérez Cañado, 2018, 2019; Rascón Moreno & Bretones Callejas, 2018). Parents' education level has also emerged as an important factor influencing selection into CLIL in the Finnish context, as discussed by Aro and Mikkilä-Erdmann (2014) and Merisuo-Storm (2007). Moreover, concerns have previously been raised about the selection of students in CLIL programmes (Bruton, 2011, 2013), arguing that such programmes may be implicitly restrictive and elitist. Thus, accounting for SES, ensuring homogeneity, and exploring the impact of selection in CLIL programmes, have become important considerations in subsequent CLIL research, including, but not limited to, the research that relates to the impact of CLIL on L1 ability (e.g., Pérez Cañado, 2018; San Isidro & Lasagabaster, 2018).

In research conducted in northwestern Spain, San Isidro and Lasagabaster (2018) focused on the impact of a small-scale CLIL implementation (i.e., only one subject was taught via English) on students' ability in their L1/L2 (i.e., Spanish/Galician) and their English as a third language (L3) in a multilingual region of Spain. In order to control for selection bias due to SES, the study focused on a school in a rural region of northwestern Spain, where SES is considered to vary less than in urban areas. The authors found that students in the CLIL group outperformed students in

the non-CLIL group in L1 ability. Similarly, in a study conducted in another multilingual region of Spain (i.e., Basque Country), Merino and Lasagabaster (2015) found CLIL students not to be disadvantaged in terms of L1 ability in a CLIL implementation in which the target language (i.e., English) was used 22.5% of the time. That said, SES was not analysed in connection with the results of that study (Merino & Lasagabaster, 2015). Moreover, the use of a regional school rather than accounting for SES in all students individually renders the conclusions found by San Isidro and Lasagabaster (2018) less compelling than those of other studies that have used, for example, parents' education level as a proxy for SES (Pérez Cañado, 2018, 2019; Rascón Moreno & Bretones Callejas, 2018). Students' SES was also not accounted for in the results of a study by Navarro-Pablo and López Gándara (2019), who studied the impact of a large-scale CLIL implementation (i.e., at least 50% of instruction was in the target language) on L1 in a mix of primary and secondary education settings in southern Spain. Verbal intelligence and motivation were used to homogenise the sample, and the results indicated that CLIL students outperformed non-CLIL students in L1 (i.e., Spanish) as a subject (Navarro-Pablo & López Gándara, 2019). These results seem to contrast with those of Anghel et al. (2015), who found a clear negative effect on scores in a standardised end-of-year test held in the students' L1 Spanish (but not specifically focused on language skills) for primary school CLIL students whose parents had less than upper secondary education. That study comprised a CLIL implementation in which around 40% of the instruction was in the target language (i.e., English). Together, these studies illustrate the importance of considering SES in future studies concerning the relationship between CLIL and L1 ability.

In a comprehensive study comprising large-scale implementations of CLIL (i.e., at least 50% of instruction was in the target language) from southern Spain, Pérez Cañado (2018) sought to guarantee the homogeneity of participant groups in terms of motivation and verbal intelligence, while also factoring in SES. The multivariate analysis revealed that higher L1 ability was linked to higher SES in both the CLIL and non-CLIL groups at primary and secondary education levels. Importantly, CLIL participation was not found to affect L1 competence in that study, a finding which is consistent with a later study by Barrios (2021) in a comparable context. That said, the finding by Pérez Cañado (2018) concerning SES and L1 ability contrasts with the results from another study from Spain (Lorenzo et al., 2021) in which higher SES was found to correspond to higher L1 test scores among non-CLIL students, but not among CLIL students. The contrast in results among the aforementioned studies could be attributable to differences in data collection, whereby the studies by Pérez Cañado (2018) and Lorenzo et al. (2021) used different tests to approximate L1 ability, while Barrios (2021) examined students' attainment in L1 as a subject. Nevertheless, the results obtained by Lorenzo et al. (2021) suggest that CLIL may

mitigate the impact of SES on L1 ability[2]. CLIL's role in reducing the impact of SES on performance has been referred to as a potential *levelling effect* of CLIL (Halbach & Iwaniec, 2020; Iwaniec & Halbach, 2021). The fact that CLIL students' L1 results would be less subject to variation on the basis of SES than those of non-CLIL students is also reflective of an earlier study conducted by Rascón Moreno and Bretones Callejas (2018).

Although recent studies concerning the relationship between CLIL and L1 ability have mostly been conducted in Spain, research aiming to shed light on this phenomenon has also been conducted in Sweden. In a study focused on the development of upper secondary level students' academic vocabulary in L1 Swedish, Holmberg (2019) found that students in the CLIL group with virtually no instruction in their L1 demonstrated weaker development in L1 academic vocabulary compared to both the non-CLIL group and another CLIL group in which L1 use was also employed during class. Comparing similar groupings (i.e., a non-CLIL group and two different CLIL groups in upper secondary education), Lim Falk (2019) found that students in the CLIL group with essentially no L1 use produced more errors in L1 academic writing than students in the other two groups. Conversely, in a further study conducted in Sweden (Ohlsson, 2021), no such differences were found to exist among similar groups to those used in the aforementioned studies. This discrepancy in results is complicated by the fact that variables such as SES were not consistently included or accounted for. That said, these studies were all conducted in an overarching linguistic context that is comparable to the one used in the present study. For instance, the use of ELF is common in both Sweden and Finland (Peterson & Beers, 2023), which means that Finns and Swedes are quite likely to have experience using English as the primary language of communication with speakers of other languages. Additionally, Finland and Sweden are countries in which media such as television programmes and cinema have traditionally been displayed in the original version without resorting to dubbing (Schurz & Sundqvist, 2022), constituting a source of considerable extramural exposure. When combining a CLIL implementation with limited L1 exposure (or use) and a strong practice of using ELF, along with a high level of extramural exposure to English, the distinction between L1 and L2 may become unclear. Such cases highlight the need to consider the relationship between CLIL and L1 ability as part of a broader discussion about bi- and multilingualism.

[2] It is also worth noting that in the study by Lorenzo et al. (2021), the cohort comprising high-SES non-CLIL students had higher scores in the L1 test than all other cohorts from either CLIL or non-CLIL.

2.3 CLIL and plurilingualism in Finland's basic education

Finland is a multilingual country with two official languages (Finnish and Swedish), in addition to the widespread use of English, particularly as a *lingua franca* (see, e.g., Laitinen et al., 2023; Leppänen et al., 2011). Despite the rising demand for schools to provide education via the medium of English, school principals and education specialists have also emphasised the importance of proficiency in the Finnish language for participation in Finnish society (Peltoniemi et al., 2018; Pöyhönen et al., 2024). As a form of bilingual education in Finland, CLIL has been utilised to promote the development of target language skills and content knowledge via implementations with varying degrees of target-language use. All CLIL implementations in Finland are defined by the national core curriculum as either large scale, where at least 25% of instruction is in the target language, or small scale, where less than 25% of instruction is in the target language (Finnish National Agency for Education, 2014). The target language in CLIL implementations in Finland is most commonly English (Pöyhönen et al., 2024). In basic education in Finland, CLIL is offered from grade one through nine, covering primary and lower-secondary education, in diverse regions of Finland (Pöyhönen et al., 2024). During the 2021-2022 school year, 4,906 students participated in English-medium bilingual education (i.e., CLIL) in basic education in Finland, with the majority being in the country's main cities (Pöyhönen et al., 2024).

Language education in Finland is reflective of guidelines and definitions outlined in the Common European Framework of Reference for languages (CEFR). The CEFR (Council of Europe, 2020) outlines six main levels of language competence grouped into three categories: basic user (A1 and A2), independent user (B1 and B2), and proficient user (C1 and C2). Moreover, multilingualism is defined as the "coexistence of different languages at the social or individual level" (Council of Europe, 2020, p. 30), whereas plurilingualism denotes the "dynamic and developing linguistic repertoire of an individual user/learner" (Council of Europe, 2020, p. 30). Given that this repertoire encompasses uneven levels of competence in different languages, a speaker's plurilingual competence, therefore, refers to their ability to switch between and utilise all their linguistic resources for various communicative aims (Council of Europe, 2020). As a consequence, plurilingualism has considerable implications for CLIL education in Finland, where learners have various levels of proficiency in not only the target language (e.g., English), but also in national languages (i.e., Finnish, Swedish) and any other languages a student may speak or study. In terms of pedagogy, while practices such as differentiation can be used to help CLIL students with specific individual learning needs (e.g., Roiha, 2012), approaches such as translanguaging can be used to harness a student's full linguistic repertoire to promote learning in CLIL (e.g., Amondarain-Garrido & Ruiz de Zarobe,

2024; Nikula & Moore, 2016). To sum up, as a form of bilingual education in Finland, CLIL education embraces the framework provided by the CEFR to support the development of learners' plurilingual competence in an overarching multilingual context.

2.4 Bilingualism, multilingualism, and translanguaging

Broadly speaking, child bilingualism can generally be categorised as either simultaneous or sequential. In simultaneous bilingualism, a child acquires two (or even three) languages at the same time (Grosjean, 2010), starting from the moment when acquisition is deemed to begin. Conversely, sequential (or successive) bilingualism refers to those cases in which a child learns a second language after having started learning their first (Paradis et al., 2011). This could include cases where a child learns different languages at home and in society. Sequential bilingualism can also refer to a situation in which a language is acquired through immersion or CLIL education. That said, an instance of sequential bilingualism may also be defined as additive or subtractive, depending on whether the addition of a new language is considered to complement or hinder the child's first language (Enstice, 2017; Pliiddemann, 2013). Even if a child is initially considered bilingual, their linguistic profile could later be affected by other factors, such as immigration, modifications in language practices by parents, educational opportunities, and personal choices concerning language use. Moreover, further possible categories of bilingualism exist, including heritage language speakers (e.g., Hernandez Santacruz et al., 2025) and late bilingualism in adults (e.g., Vega-Mendoza et al., 2015).

Alongside the aforementioned types of bilingualism, an alternative framework for conceptualising bi- and multilingualism has emerged, aiming to treat each of a speaker's languages not as a separate system but rather as integrated parts of a whole. This change in perspective reflects the advent of the *multilingual turn* in research, symbolising a shift away from applied linguistics research couched in a monolingual perspective of language learning (Conteh & Meier, 2014; May, 2014). This perspective aims to characterise languages as part of a symbiotic and dynamic system, not existing independently from each other or from the people who speak them (Meier, 2017). Importantly, in light of the multilingual turn, learners are considered to be multilinguals (i.e., multilingual practitioners) rather than only becoming bi- or multilinguals in the future after having achieved a certain level of proficiency (Meier, 2017). This shift towards a multilingual perspective aligns with assumptions in both psycholinguistics, where bilinguals are considered to operate within an integrated system (e.g., Grosjean, 1989), and socioconstructivist

approaches, which emphasise the socially mediated nature of language learning (e.g., Lantolf, 2011).

However, adopting a multilingual perspective is not necessarily applicable in all cases. Given that the present study is carried out in the context of an implementation of bilingual education (i.e., CLIL) in an overarching multilingual context (i.e., Finland), it is worth scrunitising the notion that everyone is multilingual by way of logic and examples in order to establish a baseline theoretical understanding for use in this dissertation. To begin with, if everyone is to be considered multilingual, then the terms *monolingual, bilingual* and *multilingual* could all be considered redundant and fall into disuse (e.g., *a multilingual individual* would simply be *an individual*). If this were to occur, it would complicate the process of understanding and discussing differences in speakers on the basis of language proficiency. This can be illustrated by way of an example comprising two speakers: Student A and Student B. Students A and B both participate in a Chinese language class in Spain; Student A is a balanced bilingual speaker of Spanish and English, whereas Student B speaks Spanish as their native language, and English only at an elementary (i.e., A1-A2) level according to the CEFR. Both students have now begun learning Chinese. It might be misleading and unhelpful to label both students as *multilingual*, completely disregarding the enormous gap in language skills between the two speakers. Using the term *bilingual* for Student A would, arguably, be much more helpful in understanding the differences between the two students, who may have considerably different cognitive (Białecka et al., 2023) and language learning abilities (Antoniou et al., 2015). As a consequence, it would seem practical and helpful to use, for example, language proficiency to define where monolingualism ends and bi- or multilingualism begins (Edwards, 2012). This means that low-level learners or speakers who cannot, for example, function in a given linguistic environment would not be considered bi- or multilingual.

At this stage, it is also worthwhile acknowledging the concept of translanguaging and its relationship to bi- or multilingualism. As summarised by Li (2018), translanguaging can be conceptualised as a dynamic, integrated practice of using different languages, and as a process for building knowledge. One of its implications for education is that it challenges classroom practices such as the strict use of the target language and the strict separation of languages (Li, 2018). Moreover, translanguaging scholars highlight that communication is multimodal and semiotic (e.g., García & Li, 2014; Li, 2018), which challenges the practice of viewing named languages (e.g., Spanish, Chinese) as fixed, separate entities, thus calling into question earlier interpretations of bi- and multilingualism (e.g., Li & García, 2017). Addressing this claim would require an extensive discussion about the meaning (or meanings) of *language*, which is not within the scope of the present study. Moreover, this dissertation aims to address an existing gap in research by focussing on exploring

questions that presuppose clear boundaries between languages (e.g., L1 and L2), which is why it is not useful to consider a perspective of language that does not support such an understanding. That said, it is worth considering the pedagogical role that translanguaging may play in a discussion about L1 ability in CLIL contexts. In translanguaging research, bi- and multilinguals are considered to possess a single linguistic system, rather than separate systems for, for example, L1 and L2 (García & Li, 2014). This perspective has implications for CLIL classrooms, where teachers can, with the use of translanguaging as a tool, aim to harness the entirety of a student's linguistic repertoire simultaneously, rather than strictly adhering to the use of an L1 or L2 at a given time. This can allow teachers to clarify meaning and emphasise communication in ways that may not be facilitated by a strict use of the target language.

However, even though translanguaging has clear pedagogical applications for CLIL education (e.g., Amondarain-Garrido & Ruiz de Zarobe, 2024; Nikula & Moore, 2016; Wong & Tian, 2025) and is considered reflective of bi- and multilinguals' natural communicative tendencies (Ai et al., 2022; Suarez, 2023), not emphasising the separation between languages may actually conflict with the present social reality, leading to the type of "cognitive dissonance" alluded to by Nikula and Moore (2016, p. 239). For instance, while Finnish and Swedish are both official languages in Finland, these two languages are clearly separated in government-related contexts (e.g., forms, documents, interviews), in the media (e.g., Finnish or Swedish-language programmes and newspapers), and, to a lesser extent, in education (e.g., Finnish or Swedish language matriculation exams). Similarly, in professional contexts, if someone is hired to provide services in, for example, Finnish, translanguaging practices such as borrowing a term from Swedish may not be an appropriate or helpful solution in practice. This means that the bi- or multilingual speaker would need to control their (natural) urge to *translanguage* in such circumstances. Additionally, interpretations of linguistic competence in the EU are reflective of the CEFR (European Parliament & Council of the European Union, 1998), which distinguishes between separate languages (e.g., English, French) and levels (e.g., B2 or C1). Admittedly, a distinction has been drawn between strong and weak versions of translanguaging, whereby the former considers speakers to contain one all-encompassing linguistic repertoire and the latter acknowledges boundaries between named languages (García & Lin, 2017). In sum, while this study recognises the contributions of translanguaging research in terms of its implications for classroom practice, its relevance in shedding light on the overall relationship between CLIL and ability in a named language (i.e., Finnish) is limited, particularly insofar as the strong version of translanguaging is concerned.

Another approach to understanding what is meant by bi- or multilingualism includes considering the possible effects of bi- or multilingualism on the speakers

themselves. Having been explored extensively in research, bilingualism has been associated with benefits in various domains, such as cognition (Białecka et al., 2023), adaptability in different situations (Bialystok et al., 2012), aptitude in language learning (Antoniou et al., 2015) and narrative skills (Haman et al., 2017). However, when isolating and focusing on the L1 of bilinguals, researchers have identified linguistic costs associated with bilingualism in aspects of L1 such as oral production (Baus et al., 2013), vocabulary size (Siow et al., 2023), access to low-frequency words (Gollan et al., 2008; Ivanova & Costa, 2008), and lexical processing (Bylund et al., 2019). As discussed by Jarvis and Pavlenko (2008), the L1 ability of bilinguals may also be influenced by additional phenomena such as reverse transfer, L1 attrition or, for example, an incomplete acquisition of L1 in some cases (e.g., heritage speakers). Bilinguals also appear to be more subject to tip-of-the-tongue moments (TOTs) (Gollan et al., 2013; Kreiner & Degani, 2015) – a TOT can be defined as the experience of struggling to access a word despite having it in one's memory (Cleary, 2017). As CLIL can be considered to make a contribution to students' bi- or multilingualism, the relationship between CLIL and L1 ability can be placed within a broader discussion about the possible bilingual costs (to L1) of participating in bilingual education (e.g., CLIL). This study establishes this connection between CLIL and bi- or multilingualism and furthers the discussion about the benefits and costs of bilingualism and bilingual education.

3 Data and Methods

The data and methods of the present study are presented in this chapter. Firstly, the philosophical orientation and methodological approach employed in this study are discussed in Section 3.1, together with a summary of the ethical considerations concerning data collection. An overview of the data and methodological framework is provided in Table 1. Then, the participants are described in Section 3.2, after which the procedures for data collection and analysis are outlined in Sections 3.3 and 3.4, respectively.

3.1 Philosophy, methodology, and ethical review

Before proceeding to outline the data and methods in detail, it is worth briefly discussing the philosophical orientation that has informed this study. Firstly, it should be noted that academic research conducted by humans does not emerge of its own accord. At all stages of a study, the researcher, guided by their existing knowledge and interests, is forced to make decisions concerning the choice of topic, the selection of literature to review, the generation of research questions, and, for example, the selection of methods. This suggests that the researcher's own subjectivity, or reflexivity (see, e.g., Consoli & Ganassin, 2023), is ultimately inseparable from the research process. This is further complicated in applied linguistics research (e.g., the present study) where abstract notions such as language and learning are scrutinised.

From an ontological perspective, which is concerned with understanding both concrete and abstract conceptualisations of *reality* (Effingham, 2013), the present study aims to elucidate a phenomenon that is partly shaped by human interpretations. The process of seeking to understand reality also concerns the field of epistemology, which focuses on answering questions related to what is meant by *knowledge* (Audi, 2010). In brief, the philosophical orientation of the present study can be said to be reflective of a kind of *philosophical materialism*, according to which existence (including the existence of abstract notions) is intricately connected to matter (Pérez-Jara et al., 2022), and *realism*, which assumes the objective existence of the world regardless of our thoughts about it (Glanzberg, 2025). The relevance of these notions in the present study can be illustrated via the following logical deduction: even if

human languages can be deemed to be socially constructed (e.g., Chen, 2015; Li & García, 2017, 2022), this process of construction would ultimately be carried out by humans operating within the parameters established by physics and biology; as a corollary, any named language (e.g., English, Finnish, Spanish) emerging within those material (i.e., physical and biological) parameters can, despite being shaped by immaterial (e.g., social) factors, be considered to exist in the material world.

Thus, the overarching perspective adopted in this study stands in contrast to views on language that are reflective of *postmodernism*, which can be considered not necessarily a theory itself but rather an amalgamation of ideas that "calls attention to different ways of knowing and thinking and to new forms of representation and communication" (Arai, 2023, p. 5343). The contrast between the orientation guiding the present study and an orientation grounded in postmodernism is illustrated clearly by Pennycook, (2006) who referred to "precisely such materialism and realism that … postmodernism seeks to challenge, attempting either to invert this relationship or collapse it all together (making language, discourse, ideology, and culture primary sites of how the world is organized and understood)" (p. 61). Rather than aligning with a perspective that seeks to bring about such a change, the present study aims to understand and depict reality as it exists within the confines of the material world. Consequently, although this study is deeply couched within a social science tradition where it is often difficult to establish a connection between research findings and a materialist framework, it acknowledges that such a connection is likely to exist irrespective of humans' current ability to perceive it. In doing so, the study invariably alludes to the challenges of using a postmodernist lens to understand the connection between human languages and reality.

In terms of methodology, a considerable challenge of conducting research with students and data from a specific classroom context concerns the selection of appropriate methods that can realistically address the chosen research questions, given the natural limitations that a classroom imposes on sampling possibilities and sample sizes. Moreover, classroom-based research is often focussed on understanding phenomena that do not require strictly quantitative methods, even though such methods can lead to greater systematicity and facilitate data analysis (Plonsky, 2017). As outlined in Table 1, these challenges have been responded to in the present study by using various data collection methods to shed light on L1 ability, including: a verbal fluency task (VFT), a picture-naming task (PNT), a writing elicitation task, a speaking elicitation task, students' self-evaluations and their grades in their L1 as a subject (L1 grades). Consequently, the approach in this study constitutes a form of methodological triangulation, which refers to the use of multiple methods of data collection to elucidate the same phenomenon (Riazi & Candlin, 2014; Sántha & Malomsoki-Sántha, 2023). Given that triangulation reduces a researchers' reliance on any particular method or dataset, adopting such an

approach is considered to have a positive impact on the validity and credibility of a given study (Meydan & Akkaş, 2024).

Table 1. Overview of data and methodological framework.

Aims of the present study	Title of article	Data collection	Data analysis and rationale
Sub-aim 1: to establish methods that can be used to address the overarching research aim in a Finnish context	*Exploring the Relationship between CLIL and L1 Ability in Finland: Analyzing Written and Oral Production*	*n* = 4 (2 CLIL, 2 non-CLIL) Written responses to a prompt question about the environment in L1 and L2. Oral responses to VFTs and PNTs in L1 and L2. Language backgrounds, SES, and self-assessments in L1 and L2.	A small-scale case study approach was used to compare participants' responses, backgrounds and self-assessments, in order to explore research methods (sub-aim 1) while also addressing sub-aim 1, 2 and the overarching research aim (see Section 1.2).
Sub-aim 2: to examine the relationship between CLIL and L1 oral production	*Examining L1 ability in a Finnish secondary education context: a comparison of CLIL and non-CLIL students' oral production*	*n* = 31 (12 CLIL, 19 non-CLIL) Oral responses to a VFT and a PNT in L1. Language backgrounds, SES, L1 grades, and L1 self-assessments.	Independent samples *t* tests with bootstrapping and Mann-Whitney U tests were used to gain insight into group-level differences regarding sub-aim 2. Correlation analyses (Spearman's rank) were utilised to shed light on the role of additional variables.
Sub-aim 3: to examine the relationship between CLIL and writing and speaking in L1	*The effect of CLIL on L1 ability: Evidence from a Finnish secondary education context*	*n* = 30 (11 CLIL, 19 non-CLIL) Written responses to a prompt question about the environment in L1. Spoken responses to an elicitation task in L1. Language backgrounds, SES, L1 grades, and L1 self-assessments.	Mann-Whitney U tests were used to gain insight into group-level differences concerning sub-aim 3. Correlation analyses (Spearman's rank) were utilised to shed light on the role of additional variables.

The data collection and analysis were conducted in line with the requirements stipulated by The Finnish National Board on Research Integrity (TENK). TENK affirms that ethical review is required in research with human participants in various circumstances, such as in cases where there are deviations from typical informed consent principles or if exposing participants to possible mental harm. The present study did not meet any of the conditions outlined by TENK, which is why ethical review was not required. However, approval to collect data was provided by the school, and consent was obtained electronically from all participants in this study before proceeding to data collection. Considering that the students were all minors,

their parents or guardians were notified about the research in advance and given the opportunity to opt out, if desired. The students and their parents or guardians were informed about the aims of the study along with practical considerations regarding data collection and management. All data collected from participants were initially either pseudonymised (Article I) or anonymised (Articles II and III) before proceeding to the data analysis.

3.2 Participants and context

The participants in Article I were students in a secondary education context in southwestern Finland during autumn 2022. At the time of data collection, the students were in the ninth grade and participated in either CLIL or non-CLIL education. Although 24 students completed the initial background survey, four students ultimately participated in this exploratory, small-scale case study comprising oral and written tasks. It was not deemed necessary to include additional participants in order to fulfil the exploratory purpose of Article I, which also served as a pilot study for Articles II and III. Two students were selected from each group (CLIL and non-CLIL) primarily on the basis of Finnish being their L1. Details about the participants (n = 4) in Article I are depicted in Table 2.

In the CLIL context in the present study, the option to participate in CLIL education is available from grade 1 to grade 9, after which students can choose to continue their education in English via an English-medium International Baccalaureate programme. This CLIL programme is categorised as a large-scale implementation of CLIL education which, in the Finnish context, means that at least 25% of instruction is in English in all subjects except for languages other than English. Although all participants were L1 Finnish speakers, there was considerable variation in the participants' previous exposure to other languages, particularly between CLIL and non-CLIL students. In addition to participating in CLIL education in English, Jari and Milja had spent 4.5 and 2 years, respectively, living in English-speaking environments. The non-CLIL students had previously neither lived in a non-Finnish environment nor been educated via a language other than Finnish (e.g., EMI), excluding their foreign language classes. Additionally, the education level of the students' parents varied considerably – ranging from compulsory education to a master's degree – although the variation is somewhat similar in both pairs (i.e., CLIL and non-CLIL) of students. In the present study, parents' education level was used as a proxy for SES as in previous CLIL research (Anghel et al., 2015; Pérez Cañado, 2018, 2019; Rascón Moreno & Bretones Callejas, 2018). Nevertheless, it should be noted that not all studies have used a single data point (e.g., parents' education level) as a proxy for SES (see, e.g., Fernández-Sanjurjo et al., 2017; Lorenzo et al., 2021),

which is why considerable caution must be exercised when interpreting the results of the present study in connection with SES.

Table 2. Details of participants in Article I.

Pseudonym (class)	Age	Sex	Education level (parent 1)	Education level (parent 2)	Most dominant language	First language learnt
Jari (CLIL)	15	M	High school diploma	Compulsory education	Finnish	Finnish
Milja (CLIL)	15	F	Master's degree	Master's degree	Finnish	Finnish
Jenni (non-CLIL)	14	F	Bachelor's degree	Bachelor's degree	Finnish	Finnish
Sanna (non-CLIL)	15	F	High school diploma	High school diploma	Finnish	Finnish

M = Male, F = female

The participants in Articles II and III (see Table 3) were from a single pool of students from two secondary education contexts: one in southwestern Finland and one in southern Finland. The students were all in the ninth grade and either 14 or 15 years old at the time of the data collection, which was during the autumn 2023 and spring 2024 semesters. Although these data and participants are different from those in Article I, the CLIL students in Articles II and III are from the same school as the CLIL students in Article I. In Article II, 12 CLIL students and 6 non-CLIL students from one school participated along with 13 non-CLIL students from a second school, making a total sample size of 31. The second non-CLIL group was selected and included because of the small size of the first non-CLIL group. The participants in Article III ($n = 30$) were the same as those in Article II ($n = 31$), except for one CLIL student who chose not to participate in the data collection for Article III. The focus in Article II was on students' L1 oral production as gauged using a VFT and PNT, whereas Article III scrutinised students' L1 writing and speaking via tasks aimed at eliciting longer stretches of language.

In both Articles II and III, any students with a diagnosed learning difficulty were not included in the study in order to improve homogeneity. Similarly, students whose L1 was not deemed to be Finnish were also excluded. Given that the students in these classes were, in many cases, multilingual and multicultural, participants were considered to have Finnish as their L1 on the basis of the following rationale. Firstly, students who listed Finnish as their most dominant language were included. Secondly, any students who were Finnish citizens, listed Finnish as the first language they acquired, and had not spent more than two years living in a place (e.g., a

country) where another language was spoken, were also included. The decision to exclude students who had spent more than two years in a non-Finnish-speaking environment was made in light of the results of Article I, in which the student with the weakest L1 (Jari) had spent 4.5 years living in an English country. Finally, even if all of the aforementioned criteria were met, students were still excluded if they were found to be taking the Finnish language subject intended for L2 speakers of Finnish rather than L1 speakers.

Table 3. Details of participants in Articles II and III.

Class	Article II (*n* = 31)			Article III (*n* = 30)		
	SES		Mean age	SES		Mean age
	Mean	SD		Mean	SD	
CLIL	4.08 (*n* = 12)	1.31	14.58	3.95 (n = 11)	1.29	14.54
non-CLIL	3.95 (*n* = 19)	1.31	15.00	3.95 (*n* = 19)	1.31	15.00

As in Article I, parents' education level was used as a proxy for SES in Articles II and III. The SES calculations represent the average figure for each students' parents' education level, which ranged from 1 (compulsory education) to 7 (doctoral degree). As indicated in Table 3, the SES figures for the complete dataset used in Articles II and III were comparable between the CLIL and non-CLIL participants. For the 12 CLIL students, the mean was 4.08 and the standard deviation (SD) was 1.31, whereas for the 19 non-CLIL students, the mean was 3.95 and the SD was 1.31. For the 11 CLIL students who participated in Article III, the mean SES was 3.95 and the SD was 1.29.

3.3 Data collection

In Article I, the data were collected from participants at their school and included an online survey completed via the students' own devices, written tasks in both L1 and L2 completed by hand on paper, as well as verbal fluency tasks (VFTs) and picture-naming tasks (PNTs), for which responses were recorded on a laptop computer. The survey captured parents' education level, which was used as a proxy for SES, students' self-evaluations in L1 and L2, and students' linguistic profiles using questions from the LEAP-Q questionnaire (Marian et al., 2007), which targets language experience and proficiency in bi- and multilinguals. The battery of self-evaluation statements included some items used in previous research (Rumlich, 2016; Seikkula-Leino, 2002) as well as items generated by the authors. The 16 statements concerning students' self-perceptions of both their Finnish and English (i.e., eight statements per language) were presented bilingually in Finnish and

English (see Appendix A in Article I). The questions targeted students' language-specific perceptions of their overall ability, writing ability, grammatical knowledge, vocabulary knowledge and performance in classes. A Likert scale was used whereby participants were asked to choose a number from one to five corresponding to their level of agreement with the statement.

For Article I, participants completed the written task in both languages separately. The evaluation of writing is of interest in the present study as it enables comparisons with previous related studies in which writing has been scrutinised (e.g., Lim Falk, 2019; San Isidro & Lasagabaster, 2018). The task question (see Table 4 on the following page) prompted students to produce a response on a topic (i.e., the environment) that had previously been covered in their curriculum. Students were given the instructions in the same language as the task[3], and the tasks were completed in both languages on the same day due to practical limitations inherent to the data collection[4]. Participants were given a suggested word count, which was 50 words lower in Finnish than in English given the considerable morphosyntactic differences between the two languages. Students were given 20 minutes to complete each written task by hand without access to their phones or contact with other students.

Table 4. Prompts used in written tasks in Article I.

Language	Task question	Suggested word count	Time limit
Finnish	Ihmisten toimet ovat syy luonnon ekosysteemien tuhoutumiselle. Oletko samaa vai eri mieltä? Perustele.	150 – 200	20 minutes
English	Human actions are to blame for the destruction of natural ecosystems. Do you agree or disagree with this statement? Justify.	200 – 250	20 minutes

A verbal fluency task (VFT) was utilised in Article I (and Article II) because it has previously been shown to be effective in assessing the L1 oral production of bi- or multilinguals (Aziez et al., 2020; Rojczyk, 2011; Shao et al., 2014). The VFT in Article I was a letter task, whereby participants were asked to name as many words as they could starting with the letter *t*. This letter was chosen because of the large number of common words starting with *t* in each language, as confirmed upon

[3] For all elicitation tasks in Article I, II and III, students were always given the instructions in the same language as the task itself.

[4] In Article I, all elicitation tasks were completed in both English and Finnish on the same day.

consultation with dictionaries. The letter task was selected as it has been found to help participants suppress semantically related words during the task itself (Marko et al., 2022; Shao et al., 2014), thus ostensibly reducing the risk of inconsistencies emerging in participants' results. Students were given 60 seconds to produce as many exemplars as possible, and they were instructed to avoid repetitions and proper nouns. Their responses were recorded using the Audacity software version 3.3.0 on a laptop computer in a quiet room at the participants' school.

The picture-naming task (PNT) in Article I comprised 12 pictures belonging to three categories: everyday terms, thematically-related terms (i.e., tools and hardware), and geometric shapes. The items in the first two categories came from the picture set created by Snodgrass and Vanderwart (1980) for use in experiments related to lexical processing, which was later updated by Rossion and Pourtois (2004), whereas the geometric shapes were created with Microsoft Powerpoint. The reason for using these three categories was to test the participants' ability to retrieve various types of words from different contexts and of different levels of frequency. The everyday terms (*caterpillar, doorknob, light switch* and *kite*) were included as they were not expected to pose problems in lexical access or retrieval for any of the students. In addition to providing a valuable point of contrast with the everyday terms, the group of thematically-related terms (*pliers, wrench, nut* and *screw*) was included in order to gauge differences in the participants' ability to name several items from one semantic field (i.e., tools and hardware). Finally, the geometric shapes (*cylinder, trapezoid, hexagon* and *pentagon*) were included as they represented a group of subject-specific terminology (i.e., mathematics) as well as possibly constituting a considerably greater challenge than the everyday terms with respect to frequency. Students were instructed to name the items in the language of the test, or to indicate if they did not know or could not remember the word. Time was not strictly restricted *per se* in order to allow students to retrieve words without pressure, thus creating the conditions for possible tip-of-the-tongue moments (TOTs) to emerge. Students' responses were recorded on a laptop computer using screen capture software.

Table 5. Self-evaluation statements used in Articles II and III.

	Self-evaluation statement		Use in Articles	
Construct	**Finnish**	**English**	**II**	**III**
L1 global	Olen hyvä suomen kielessä	I am good at Finnish	✓	✓
	Opin asiat nopeasti suomeksi	I learn things quickly in Finnish		
	Suomen kieli on minulle helppoa	Finnish is easy for me		
L1 speaking	Suomeksi puhuminen on minulle helppoa	Talking in Finnish is easy for me	✓	✓
	Luokassa käydyissä keskusteluissa pystyn aina ilmaisemaan itseäni selkeästi suomen kielellä	During discussions in class, I am always able to express myself comprehensibly in Finnish		
	Löydän aina oikeat sanat, jopa vaikeita aiheita käsittelevissä keskusteluissa suomeksi	I always find the right words even in conversations about difficult topics in Finnish		
L1 writing	Osaan kirjoittaa tekstejä hyvin suomeksi	I am good at writing texts in Finnish	✗	✓
	Osaan ilmaista itseäni kirjallisesti hyvin suomen kielellä	I can express myself well in writing in Finnish		
	Minulla ei ole vaikeuksia kirjoittaa pitkää tekstiä suomeksi	I do not have any difficulties writing a long text in Finnish		
L1 vocabulary	Minulla on hyvä suomen kielen sanavarasto	I have a good vocabulary in Finnish	✓	✗
	Tiedän paljon suomenkielisiä sanoja monista eri aiheista	I know a lot of Finnish words about a range of topics		
	Opin uusia suomenkielisiä sanoja nopeasti	I learn new Finnish words quickly		

Building on findings and observations from Article I, Article II focused on assessing participants' performance in a VFT and a PNT. Given that the inclusion of data on performance in students' L2 English in Article I did not lead to additional insights (e.g., crosslinguistic influence), coupled with the fact that the collection of such data in larger sample sizes poses additional practical challenges (e.g., arranging multiple data collection visits), the decision was made to forego collecting data on L2 ability and focus solely on data concerning L1 ability in Articles II and III. Nevertheless, participants' grades in L2 English were obtained (along with L1 Finnish grades) in order to provide some insight into possible group-level differences in L2 ability. The grades for L1 Finnish (L1 grades) were for the subject called Finnish language and literature (*Suomen kieli ja kirjallisuus* in Finnish). The assessment of this subject in Finland is carried out by teachers on the basis of a wide range of criteria and skills. The L1 grades have been included in the present study in order to shed light on individual differences among students within and across

groups. The self-evaluation statements in Articles II and III (see Table 5 on the previous page) were limited to L1 and reflect constructs (*L1 global*, *L1 speaking* and *L1 writing*) previously used by Arens and Jansen (2016), as well as one novel construct (*L1 vocabulary*) created for this study, each comprising three statements per construct. Reliability calculations were conducted using all responses received for the background survey (n = 50), which included many students who were not selected to participate in all aspects of data collection. Cronbach's alpha (α) scores were .81 for L1 global, .65 for L1 speaking, .78 for L1 writing and .82 for L1 vocabulary. The construct *L1 vocabulary* was included in Article II along with *L1 speaking* given the inherent emphasis on lexis and oral production in the VFT and PNT. In Article III, however, *L1 writing* and *L1 speaking* were included as they mirrored the two elicitation tasks that were used in that sub-study. Additionally, *L1 global* was included in Articles II and III as it is deemed to reflect all L1-related abilities elicited in both sub-studies.

As depicted in Figure 4 (on the following page), conceptual overlap can be considered to exist among the individual constructs of L1 global, L1 speaking, L1 writing and L1 vocabulary. In other words, a speaker's global ability encompasses their speaking and writing ability, as well as their knowledge of vocabulary. On the other hand, vocabulary knowledge partially overlaps with speaking and writing ability, as successful utterances and sentences are invariably dependent on the ability to access and select appropriate lexical items, amongst other factors related to language ability (e.g., syntax, morphology, context, etc.). The figure does not reflect any possible constructs that were not included in the study (e.g., L1 grammar, L1 listening and L1 reading). As in Article I, participants in Articles II and III indicated their level of agreement with the statements using a Likert scale ranging from one to five.

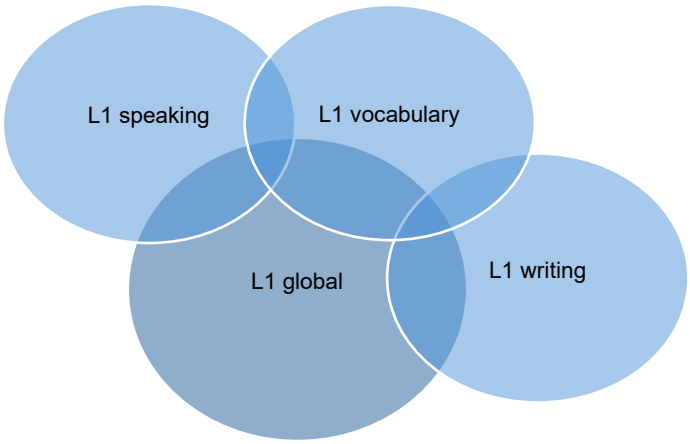

Figure 4. Conceptual overlap in self-evaluation constructs.

In the VFT in Article II, the decision was made to use a semantic task (i.e., participants are asked to produce exemplars from a category, such as *food* or *animals*) instead of the letter task that was used in Article I. It is possible that the letter task more strongly reflects executive ability rather than verbal ability (Shao et al., 2014), which could explain the emergence of the perceived outlier in the VFT results in Article I. Additionally, although semantic and letter tasks both require executive ability (Aita et al., 2019), the semantic task can also serve as a tool to gauge language-related abilities such as vocabulary size (Dubé & Thordardottir, 2024; Wauters & Marquardt, 2018). Thus, a semantic task based on the category *animals* was used in Article II as it was deemed to provide more insight into language ability than a letter task. In the data collection in Article I, the vast majority of all responses for the VFT were produced during the first 30 seconds of the 60 seconds that were allowed. Therefore, in Article II, a time limit of 30 seconds was chosen in order to maximise the amount of data collected within a limited timeframe.

When analysing the PNT data for Article I, it became apparent that both the thematically-grouped items and the geometric shapes pose challenges for students regardless of which class they are in, which complicates the analysis of such data. As a consequence, the PNT in Article II was redesigned to include pictures based on lexis deemed to reflect common, everyday items instead of the pictures used in Article I. Table 6 depicts the subset of 10 pictures considered to represent everyday terms that were chosen from the colourised picture set by Rossion and Pourtois (2004). Moreover, the age of acquisition (AoA) was obtained for all items from a previously published standardised dataset (Nishimoto et al. 2005), and frequency values for all terms were ascertained from the *Suomi24* corpus (City Digital Group, 2021). The fact that there appear to be no outliers in the data concerning AoA and frequency of items (see Figure 5 on the next page) lends weight to the conclusion that all the pictures chosen can indeed be considered to represent common, everyday items. During the PNT, DMDX – a software package designed for presenting stimuli and measuring response times (RTs) – was used for data collection, including capturing the recordings of participants' responses. Using DMDX for the PNT allowed the researcher to calculate participants' RTs with a high degree of precision. In the task, students were instructed to name the images, which were displayed consecutively on a screen for 3000 milliseconds each. Before each picture, there was a pause during which participants were shown an image of a cross in the middle of the screen for 3000 milliseconds in order to focus their attention before the next image appeared.

Table 6. PNT items included in Article II (from Rossion & Pourtois, 2004).

Item	Finnish	English
i	kravatti	(neck)tie
ii	kilpikonna	turtle
iii	ruuvimeisseli	screwdriver
iv	nappi	button
v	joutsen	swan
vi	luuta	broom
vii	pihdit	pliers
viii	pyykkipoika	clothes peg
ix	haitari	accordion
x	rusetti	bow (tie)

In Article III, the participants completed the same L1 writing task as in Article I, along with an L1 speaking elicitation task. Students had 20 minutes to complete the writing task during class time, using a pen and paper. Although the speaking task was not thematically or structurally related to the writing task, it was included in order to complement the writing task, giving Article III an overall focus on productive language skills. Admittedly, had participants completed a similar (or identical) task for both writing and speaking in Article III, it may have given rise to greater objectivity when comparing the results across tasks. That said, such an approach would have required an interval between tasks in order to reduce the risk of the results from the second task being affected by students' having completed a similar task in close temporal proximity (see, e.g., Rogers & Li, 2025). This would not have been feasible in the present study given the limitations (i.e., access, time) inherent to the data collection process at the target school, which is why the decision was made not to use similar (or identical) prompts in both the writing and speaking tasks in Article III.

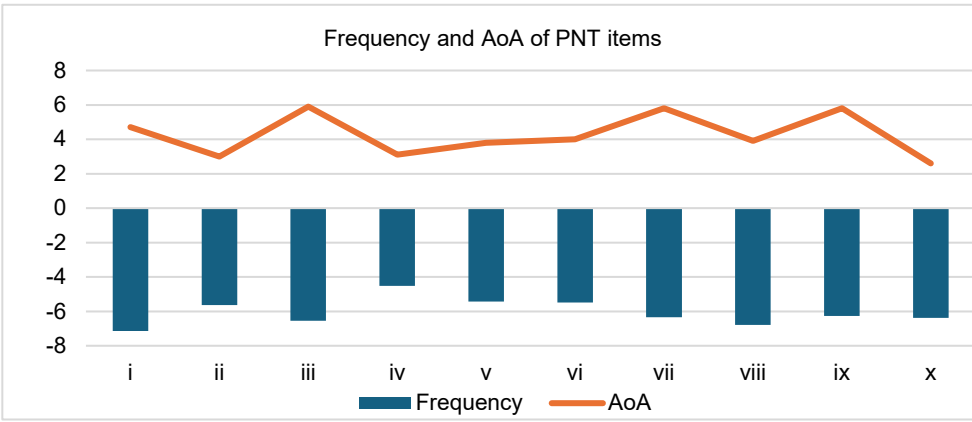

Figure 5. Frequency and AoA of PNT items in Article II.

In the speaking task, participants were first given two minutes to look at a printed cartoon strip containing six frames. After that, they had one minute to prepare to speak, followed by up to two minutes to tell the story depicted in the cartoon strip in their own words using their L1 Finnish. The same cartoon series has been used as an elicitation task in previous research on fluency in multilinguals in Finland (Peltonen & Lintunen, 2022). One minute of preparation time was given to participants as it was deemed a short enough time to elicit spontaneous speech (which is of more interest than planned speech in the present study) but not so short as to risk stressing the participants, which would not only be unethical but could also compromise the data. Instructions were given prior to starting, and students' responses were recorded on a laptop using the Audacity software in a quiet room at their school.

3.4 Data analysis

Before conducting any analysis in Article I, students' names were pseudonymised and responses to the written tasks were scanned and matched to the participants' (n = 4) pseudonyms. The written tasks were evaluated by the authors of Article I using the profile technique (Jacobs et al., 1981), which is used to give feedback on content, organisation, vocabulary, language use and mechanics. The responses were evaluated individually under blind conditions, after which the assessors met to share and discuss their evaluations, and, where agreed, make minor adjustments to their scores. In the end, raters' scores for responses in English and Finnish all had correlation coefficients of at least .74.

For the VFT in Article I, the recordings were analysed in order to identify all valid responses (i.e., words starting with *t*). This required the removal of three types of invalid responses: exact repetitions (e.g., the word *time* produced twice by the same student), repetitions by way of a synonym (e.g., *tiili* and *tiiliskivi*, both of which mean *brick* in English), and responses deemed to belong to a conceptually related category (e.g., *timer*, *timetable* and *timing*). Valid responses for each participant were then totalled and totals compared across individuals and groups. In the task, there appeared to be one clear outlier in L1 Finnish, which complicated the analysis of the results. Therefore, as mentioned earlier, the choice was made to shift from a letter task to a semantic task for Article II.

In the analysis of the PNT data in Article I, any synonyms produced by participants were accepted as valid responses (e.g., *nuppi* and *ovenkahva* for *doorknob*), along with conceptually similar items (e.g. *millipede* instead of *caterpillar*), as it may not be easy to distinguish such items easily from the pictures. Participants generally struggled to retrieve low-frequency items (i.e., those items not deemed to be everyday terms) in their L1 regardless of their teaching group. Admittedly, although the approach used allowed the researcher to collect data on

possible TOTs, it did not facilitate the collection of data on RTs, which could have provided further insight along with participants' total valid responses. For these reasons, the decision was made to focus on high-frequency, everyday items and include RTs in Article II (see Table 7).

Table 7. Differences in PNT between Articles I and II.

Article I	Article II
High- and low-frequency items were included	The focus was on high-frequency items
A distinction was made between not knowing and not remembering items	The focus was only on whether participants could recall the item
There was no time restriction for responses	Time was restricted and RT was analysed

In Article II, the participants' (n = 31) VFT and PNT responses were anonymised, after which the data were analysed quantitatively using SPSS version 28. For the VFT, all valid responses were identified and totalled for each student. Repetitions were excluded, as were superordinate category names (e.g., *bird*) if a subordinate item was also produced (e.g., *hawk*). The frequency for all VFT items was also calculated. However, given that not all words produced in the task were located in a single corpus, frequency was calculated using a makeshift corpus comprising all the responses produced during the VFT. Thus, the frequency of participants' valid responses reflects their frequency in relation to all the valid items produced (n = 453) during the task. Since all participants produced at least ten valid items, the frequency means per teaching group were calculated for the first ten items and all remaining items separately. This facilitated group-level comparisons of students' access to words on the basis of frequency. Differences in VFT totals were analysed using independent samples t tests with bootstrapping because of the small sample sizes. Bootstrapping has been found to be an effective method for improving the validity of results based on small samples (Lindstromberg, 2016; Wilcox, 2012). However, the group-level differences in VFT frequency were analysed using Mann-Whitney U tests because the data were not found to be normally distributed.

The PNT data in Article II included the totals and RTs. The PNT totals comprised all valid items (including synonyms) that corresponded to pictures, whereas the RT represented the moment (to the nearest millisecond) when the participant began responding, excluding hesitations or fillers. Given that the PNT totals were not normally distributed, Mann-Whitney U tests were performed to assess group-level differences for both the totals and RTs. The data were then transformed, making the pictures the focus rather than the participants. This allowed the analysis of any picture-based variations on the basis of AoA, frequency and word length. Finally, the VFT totals, PNT totals and PNT RTs were included in correlation analyses using

Spearman's rho. The analyses also included the variables SES, L1 grades and the self-evaluation constructs of L1 global, L1 speaking and L1 vocabulary. Correlations were analysed between all the results, variables and self-evaluation constructs.

In Article III, participants' (n = 30) spoken and written tasks were anonymised and prepared for evaluation. The written tasks were scanned into PDF format, while the spoken tasks were trimmed to remove unnecessary fragments (e.g., delays before and after speaking). Responses for both tasks were evaluated by two experienced university-level teachers of Finnish communication using rubrics created specifically for this sub-study. The rubrics (see Appendices in Article III), which were loosely based on assessment criteria employed by Tervola et al. (2015), were used to assess the following elements of participants' spoken and written responses on a scale of 0 to 4: *communication*, *lexis*, *grammatical structures* and *coherence*. The average measures intraclass correlation coefficients for speaking and writing were .61 and .78, respectively.

Given that the correlations were not high or very high in both cases, the decision was made to analyse the data for each evaluator separately rather than combining them into a single figure. Using SPSS version 29, the descriptive statistics for each evaluator's scores were calculated and scrutinised for group-level differences. Additionally, similarities between the two raters' findings were identified and discussed. The scores for both evaluators were also included in correlation analyses using Spearman's rho, along with L1 grades, SES and the self-evaluation constructs of L1 global, L1 writing and L1 speaking.

4 Results

In this chapter, the results of each article will be presented in the context of the overarching research aim (i.e., to shed light on the relationship between CLIL and L1 ability in a Finnish secondary education context) and sub-aims of this study. The three sub-aims of the study (see Section 1.2) were

1. to establish methods that can be used to address the overarching research aim in a Finnish context
2. to examine the relationship between CLIL and L1 oral production
3. to examine the relationship between CLIL and writing and speaking in L1.

The connections between the overarching aim and sub-aims of the study, and the focus of each article are depicted in Table 8 below, whereas the results for the original publications are detailed in Sections 4.1–4.3.

Table 8. Connections between article foci and overarching aim and sub-aims of the study.

Focus of the article	Overarching research aim	Sub-aim 1	Sub-aim 2	Sub-aim 3
Article I: Small-scale case study of CLIL and non-CLIL students' performance in L1 writing and L1 oral production	x	x	x	x
Article II: Comparison of CLIL and non-CLIL groups' L1 oral production	x		x	
Article III: Comparison of CLIL and non-CLIL groups' L1 writing and speaking skills	x			x

4.1 Results of Article I

The first article in this dissertation lays the groundwork for subsequent sub-studies by exploring approaches used to analyse the relationship between CLIL and L1 ability in a Finnish CLIL context. Functioning as a pilot study, this small-scale case study enabled the researcher to test various methods of data collection and

analysis in order to refine and modify approaches for use in subsequent articles (sub-aim 1). For instance, in Article I, participants provided responses to all tasks in both English and Finnish separately in order to allow for the potential analysis of students' responses for crosslinguistic influence (e.g., reverse transfer) that may manifest itself in L1 writing (e.g., Rusfandi, 2024). However, upon examining the data, it became clear that focusing only on analysing L1 data would provide sufficient insight into participants' L1 ability (e.g., speaking, writing) to allow for meaningful group-level comparisons with non-CLIL students' data. This realisation, coupled with the practical obstacles of collecting double the amount of data (i.e., in both languages), is why L2 data was not gathered in subsequent articles. The verbal fluency task (VFT) and picture-naming task (PNT) results in Article I also informed the approaches that would be employed in Article II, such as opting for a semantic rather than a phonetic VFT in order to reduce the likelihood of outliers, and replacing those PNT items that were deemed too challenging for most students in both groups.

The results in Article I revealed that the non-CLIL students outperformed the CLIL students in the L1 PNT (i.e., a greater number of valid responses produced among non-CLIL students) and the L1 writing task (i.e., non-CLIL students' short essays about the environment were evaluated more highly by the authors). However, in the L1 VFT, both the strongest and weakest performers were non-CLIL students on the basis of valid responses produced. The first article also outlined the importance of considering students' L1 ability in light of their SES, which was determined by using parents' educational level as a proxy. One task where the role of SES emerged prominently was the L1 VFT, where the two students with the lowest scores were also those who had the lowest SES. Moreover, the CLIL student with low SES performed the worst of all participants in two tasks, while also performing worse than the other CLIL student and the non-CLIL student with higher SES in all three tasks. A summary of the results of this small-scale case study is provided in Table 9 (on the following page).

Table 9. Summary of results in Article I.

		Task performance (participant rankings)		
Student	**SES**	**L1 VFT**	**L1 PNT**	**L1 essay**
Jenni (non-CLIL)	Middle	1	1	2
Sanna (non-CLIL)	Low	4	1	1
Milja (CLIL)	High	2	3	3
Jari (CLIL)	Low	3	4	4

Article I also highlighted the need to consider students' linguistic profiles carefully when selecting participants for future studies. Although students may self-report that a given language is their L1 in terms of order of acquisition and overall dominance, it cannot be assumed that their self-perception provides an entirely accurate picture of their language skills. For instance, Jari, who performed weakest in two of the L1 tasks, had, in addition to participating in CLIL education in English, actually spent 4.5 years living in an English-speaking country. Such extensive exposure to English by the age of 15 raises doubts as to whether this student's (L1) Finnish can be expected to be on par with that of his peers. Given these insights gained from Article I, the criteria for selecting participants for subsequent studies were made more stringent in order to improve the effectiveness of the research methods utilised (sub-aim 1).

Overall, the small-scale case study approach adopted in this first (exploratory) article was successful in that it provided direct insight into how any differences between CLIL and non-CLIL students' L1 ability could be further analysed using variations of the methods employed in Article I. In addition, the article not only highlighted the importance of accounting for SES and carefully scrutinising participants' language backgrounds, but it also underscored the difficulty of trying to draw a clear line between those participants who should or should not be counted as L1 speakers for the purpose of researching the relationship between CLIL and L1 ability in a Finnish context.

4.2 Results of Article II

The second article focused on analysing CLIL and non-CLIL students' L1 oral production (sub-aim 2) in light of their SES, L1 self-evaluations, and L1 grades. By analysing group-level differences in L1 ability between CLIL and non-CLIL students, Article II also sheds light on the relationship between CLIL and L1 ability in a Finnish context (overarching research aim). Additionally, Article II contextualises bilingual education (i.e., CLIL) within bi- or multilingualism in a broad sense, thus establishing a connection between the effects of bilingualism on L1 ability on the one hand and the relationship between CLIL and L1 ability on the other.

In Article II, a semantic VFT (i.e., participants produced exemplars of the category *animals*) and a revised PNT (i.e., only everyday items were included) were administered to measure students' L1 oral production – specifically, their ability to access and retrieve lexis in their L1. As outlined in Table 10, the mean number of valid responses in the VFT (VFT totals) for the CLIL group was higher (16.42) than the mean for the two non-CLIL groups: NC1 (12.67) and NC2 (13.85). Bootstrapped independent samples *t* tests between the CLIL and both non-CLIL groups' VFT

totals resulted in *p* values of .032 (CLIL & NC1) and .033 (CLIL & NC2). Regarding the frequency data, Mann-Whitney U tests indicated that the group-level differences in the VFT frequencies were not significant for the first ten responses, but significant for all remaining responses with *p* values of .012 (CLIL & NC1) and .022 (CLIL & NC2). The results for the VFT totals are suggestive of a difference in favour of the CLIL students, although this cannot be assumed with great confidence because, after applying the Bonferroni correction to the bootstrapped means, the *p* values were slightly above the new significance level of .025. Nevertheless, the results certainly indicate that the CLIL students were not outperformed in this task by the non-CLIL students. This finding is supported by the fact that CLIL students' ability to access and retrieve even low-frequency words in L1 did not appear to be hindered.

Table 10. Group-level means and *t* test results for VFT totals.

Relationship	CLIL group (M ± SD)	Non-CLIL group (M ± SD)	Two-sided *p*
CLIL & NC1	16.42 ± 2.94	12.67 ± 3.08	.023
CLIL & NC1*	N/A	N/A	.032*
CLIL & NC2	16.42 ± 2.94	13.85 ± 2.27	.022
CLIL & NC2*	N/A	N/A	.033*

* Bootstrapped results

The group-level descriptive statistics depicting participants' mean valid responses (PNT totals) and RT for the PNT are shown in Table 11 (on the following page). The SD figures related to the PNT totals suggest that CLIL students' scores contained less within-group variation than those of the non-CLIL students. Group-level differences in the PNT totals and RT were evaluated using Mann-Whitney U tests; however, no significant differences were found in the results. Additionally, the results for both the VFT and PNT were included in correlation analyses with SES, L1 grades and L1 self-evaluations in order to provide greater insight into any group-level differences emerging from the elicitation tasks. While the CLIL group's results for the VFT and PNT were not found to have correlations with significant *p* values for any of the aforementioned variables, an analysis of the results for the combined non-CLIL cohort[5] revealed correlations between SES and VFT totals (r_s = .518, p = .023), SES and PNT totals (r_s = .657, p = .002), L1 grades and PNT totals (r_s = .654, p = .002), as well as L1 grades and PNT RT (r_s = .670, p = .002).

[5] In light of the lack of variation and the small sample in NC1, the two non-CLIL groups were combined in order to enable the correlation analyses in Article II.

Table 11. Group-level differences in PNT results (Table 9 in Article II).

	PNT totals				PNT RT			
Group	**Mean**	**Min**	**Max**	**SD**	**Mean**	**Min**	**Max**	**SD**
CLIL	7.5 (*n* = 12)	5	9	1.168	1.222 (*n* = 90)	0.727	2.555	0.392
NC1	7.67 (*n* = 6)	5	10	1.862	1.230 (*n* = 46)	0.710	2.758	0.474
NC2	7.46 (*n* = 13)	2	10	1.984	1.260 (*n* = 97)	0.743	2.723	0.449

The second article provides direct insight into the relationship between CLIL and L1 ability (overarching research aim) via a novel approach that incorporated psycholinguistic elicitation tests. The findings suggest that, in comparison with the non-CLIL group, the CLIL students were not disadvantaged in terms of oral production in their L1 (sub-aim 2). Aside from addressing sub-aim 2, Article II also revealed that the positive relationship SES tends to have with L1 ability might, in the present context, only apply to the non-CLIL group and not to the CLIL group.

4.3 Results of Article III

In Article III, both CLIL and non-CLIL students' L1 speaking and writing were scrutinised using tasks in which participants were invited to speak and write in response to the prompts that were given (sub-aim 3). As the results of the CLIL students were compared with those of the non-CLIL students, this sub-study also addresses the overarching research aim of this dissertation regarding the relationship between CLIL and L1 ability. Participants' spoken and written responses were evaluated by two raters (i.e., two university teachers of Finnish communication) using rubrics specifically designed for each task. The criteria evaluated in the speaking and writing tasks were communication, lexis, grammatical structures and coherence. Group-level differences were explored using Mann-Whitney U tests, after which the scores were analysed for correlations with SES, L1 grades and L1 self-evaluations, as in Article II.

Table 12. Differences in mean evaluations by raters (synthesis of Tables 3 and 4 in Article III).

	Rubric item	Rater 1			Rater 2		
		CLIL (x)	Non-CLIL (y)	Difference (y – x)	CLIL (x)	Non-CLIL (y)	Difference (y – x)
Writing	Communication	2.27	2.37	0.10	1.36	1.63	0.27
	Lexis	2.36	2.47	0.11	1.27	1.58	0.31
	Grammatical structures	2.00	2.05	0.05	1.18	1.53	0.35
	Coherence	1.82	2.16	0.34	1.73	1.68	-0.05
Speaking	Communication	2.36	2.37	0.01	2.09	2.53	0.44
	Lexis	2.64	2.84	0.20	2.55	2.32	-0.23
	Grammatical structures	2.82	2.84	0.02	2.18	2.21	0.03
	Coherence	2.82	2.47	-0.35	2.09	2.42	0.33

Although the raters helped develop and pilot the rubrics together with the researcher, the raters' scores were not combined together because the evaluations for both tasks did not reflect a high correlation coefficient. When analysed separately, the data indicated that both raters' mean scores for non-CLIL students were higher than those of CLIL students in all rubric items in both skills except coherence (in both writing and speaking) and lexis (in speaking only). Nevertheless, the absolute differences, depicted in Table 12, were not found to be significant as per Mann-Whitney U tests that were conducted. Therefore, the data suggest that CLIL students in this context may not be disadvantaged in terms of L1 speaking and writing compared to non-CLIL students.

Correlation analyses were used to gain insight into individual differences within both the CLIL and non-CLIL groups. As shown in Table 13, several instances of correlations with significant *p* values were found concerning SES and L1 grades in the non-CLIL group. Specifically, SES was found to correlate with lexis in the first rater's scores for writing, whereas SES correlated with communication and coherence in the second rater's scores for writing. There were correlations between L1 grades and both lexis and grammatical structures in the first rater's scores for writing, as well as between L1 grades and all four items in the second rater's scores for writing. Only one item (grammatical structures) was found to correlate with L1 grades for speaking (rater 1 only). Among the CLIL students, no correlations were observed between either SES or L1 grades and any of the four criteria in either speaking or writing. The students' L1 self-evaluations scores did not correlate with any of the assessed criteria for either rater. Finally, there was a strong correlation

between SES and L1 grades in the non-CLIL group (r_s = .592, p = .008), but no such relationship was observed in the CLIL group.

Table 13. Summary of correlations identified in Article III.

		Rater 1		Rater 2	
Group	**L1 skill**	**SES**	**L1 grades**	**SES**	**L1 grades**
CLIL	Writing	-	-	-	-
	Speaking	-	-	-	-
non-CLIL	Writing	L	L, G	C_1, C_2	C_1, L, G, C_2
	Speaking	-	G	-	-

L = lexis, G = grammatical structures, C_1 = communication, C_2 = coherence

The third article provides further insight into the relationship between CLIL and L1 ability in Finland (overarching research aim). While Article II sheds light on the participants' L1 oral production (sub-aim 2), Article III focuses on students' L1 writing and speaking skills (sub-aim 3). The results did not reveal any significant differences between CLIL and non-CLIL groups' writing and speaking in L1. Additionally, the CLIL students' results were less reflective of individual differences in SES and L1 grades than those of the non-CLIL group. Overall, Article III adds weight to the view that CLIL students are not disadvantaged to any considerable degree in their L1 ability compared to non-CLIL students in a Finnish secondary education context.

5 Discussion and Implications

In this chapter, the findings of the present study are explored in depth, giving rise to empirical and methodological implications as well as points to consider for educators and parents or guardians of students participating in CLIL education. The limitations of the study are also discussed, while possible directions for future research regarding the relationship between CLIL and L1 ability are proposed on the basis of insights and analyses from the present study. The chapter ends with some concluding remarks that reflect on this dissertation as a whole.

5.1 Empirical and methodological considerations

In this section, the results of the sub-studies are discussed and analysed in the context of the research aims of the present study (see Section 1.2) concerning the relationship between CLIL and L1 ability in a Finnish secondary education context. Figure 6 (on the following page) illustrates the connection between the overarching aim of the dissertation and the findings that emerged from the underlying articles. Methodological implications emanating from the present study are also identified and explored in the present section, while the results are examined in light of previous research and theoretical considerations relevant to CLIL education.

5.1.1 Methodological insights and reflections

Although Article I provides insight into the relationship between CLIL and L1 ability, the main contribution of this small-scale case study is in helping to establish suitable methods that could be used to address the overarching research aim. The verbal fluency task (VFT) used in Article I was a letter task, meaning students had to produce words starting with a given letter (the letter *t* was chosen). The VFT was a suitable choice because it not only sheds light on speakers' L1 oral production (e.g., Aziez et al., 2020; Rojczyk, 2011; Shao et al., 2014), but also because conducting the corresponding data collection does not constitute considerable practical or logistical obstacles. In Article I, the results of the VFT highlighted the possible role that SES might play in performance in the task, but the analysis of the role of CLIL was complicated due to the existence of a probable outlier in the data. The emergence

of the outlier was possibly facilitated by the parameters of the letter VFT itself, which allows participants to produce semantically and thematically unrelated words with only the starting letter in common (e.g., Marko et al., 2022; Shao et al., 2014). In light of this, together with the fact that it is unclear whether the letter VFT is more reflective of executive or verbal ability (Shao et al., 2014), the choice was made to opt for the semantic VFT in Article II. Another factor for this decision is the semantic VFT's potential in reflecting, for example, vocabulary size (Dubé & Thordardottir, 2024; Wauters & Marquardt, 2018). Combined with a larger dataset, the semantic VFT (i.e., participants were asked to produce exemplars for the category *animals*) in Article II allowed the researcher to gain insight into group-level differences between CLIL and non-CLIL students' L1 oral production without any outliers in the data.

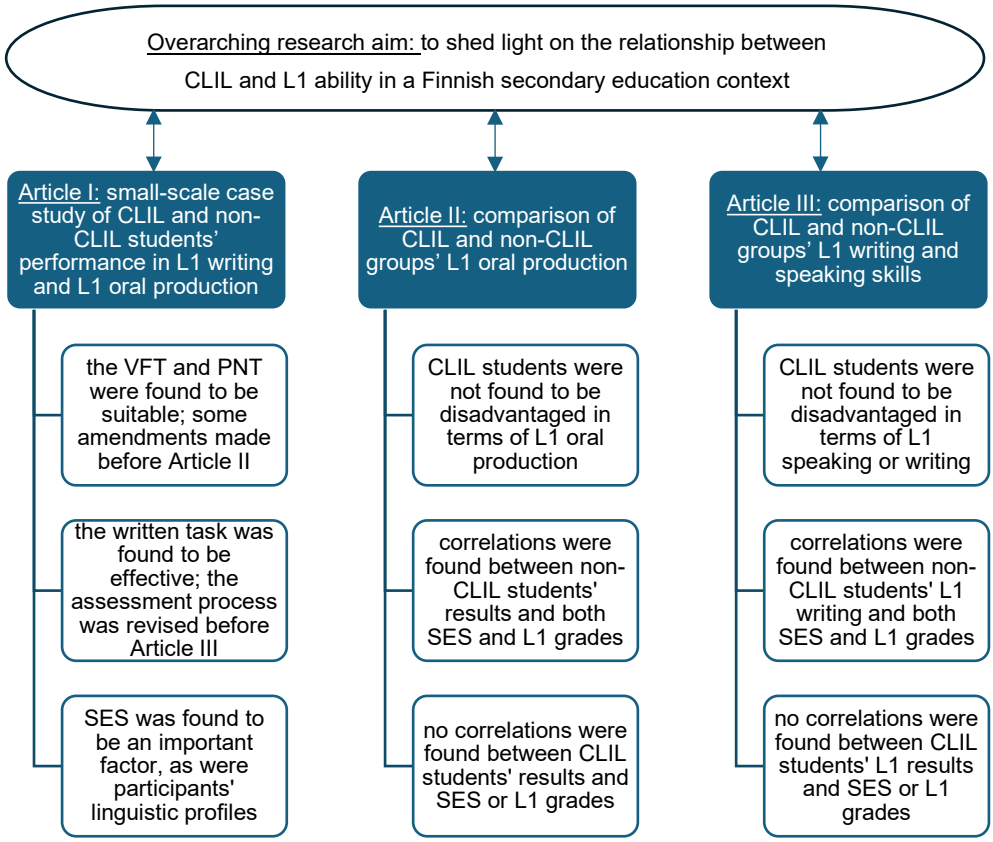

Figure 6. Connections between overarching aim, article foci, and results.

The picture-naming task (PNT) results in Article I also served a dual purpose related to both testing research methods and addressing the overarching aim of the

dissertation. Regarding the overarching aim, the results highlighted the combined role that CLIL, additional extensive exposure to English (e.g., time spent living abroad), and SES may play in the ability of students to access and retrieve lexis in their L1. In terms of the PNT as a data collection method, although the approach used in Article I allowed the researcher to establish the total number of valid responses, as well as to capture possible tip-of-the-tongue moments (TOTs), which appear to be more common in bilinguals (Gollan et al., 2013; Kreiner & Degani, 2015), it was not conducive to calculating RTs because participants were also given time to state whether they did not know or did not remember a word. In addition, this format (i.e., longer recordings using screen capture software) would not have facilitated the (straightforward) calculation of RTs even in those cases where participants successfully named the picture. Given the limited access the researcher had to the participants in the present context, it became apparent that the approach used in Article I would have made both data collection and data analysis considerably more complicated with a larger sample size (i.e., in Articles II and III). Regarding the pictures themselves, the PNT contained four everyday items, four items that were thematically related (i.e., tools and hardware), and four items that were all geometric shapes (see Section 3.3). The participants' ability to name the thematically-related terms and geometric shapes was somewhat inconsistent regardless of their group, language background or SES. Consequently, in light of the insights gained from Article I, the PNT was modified prior to the study conducted for Article II to comprise only everyday images from a single source (Rossion & Pourtois, 2004). The decision was also made to use DMDX software for data collection, thus enabling the calculation of RTs rather than aiming to capture TOTs. These changes allowed the researcher to focus on both PNT totals and RTs in Article II while also collecting data from a larger number of participants within a limited timeframe.

The results of the writing task in Article I were also suggestive of differences between students on the basis of their teaching group, language background and SES. Moreover, given that the task was simple to administer and that the task prompt reflected content that students had previously covered (i.e., the environment), it was considered a suitable elicitation method to use in future studies. Nevertheless, as with the PNT and VFT, some changes were identified with respect to the data collection and analysis. Firstly, the idea of also administering the task in the L2 (i.e., English) was abandoned after Article I because of the time that would be required to collect the additional data and the need to include a delay between task completion in each language. Although having the L2 data could potentially provide more insights (e.g., reverse transfer), it was also not deemed necessary for addressing the aims of the present study. Secondly, the assessment approach utilised in Article I (i.e., the profile technique) was originally created for the assessment of English (Jacobs et al., 1981)

and, in this study, modified for use with Finnish. Given that no new English data was going to be collected, the decision was made to create a new rubric specifically for the evaluation of the Finnish written task in future studies. The rubric was modelled on an approach previously used by Tervola et al. (2015) but modified considerably in consultation with the two raters (i.e., two university teachers of Finnish communication) who were recruited for Article III. A similar rubric was also created for the purpose of assessing the spoken tasks as part of Article III. The rubrics allowed the researcher to gain insight into four key domains related to participants' language use: *communication*, *lexis*, *grammatical structures* and *coherence*.

5.1.2 The relationship between CLIL and L1 oral production

Article II focused on addressing the overarching research aim by analysing CLIL (n = 12) and non-CLIL (n = 19) students' L1 oral production via a VFT and a PNT along with their SES, L1 grades, and L1 self-evaluations. Group level-differences were analysed using bootstrapped independent samples t tests and, where applicable, Mann-Whitney U tests, whereas the relationships concerning additional variables were assessed using Spearman's rank correlation analyses. The VFT results indicated that the CLIL students' performance was either equal to or better than that of the non-CLIL students in terms of both total valid responses and the relative frequency of items produced. This finding is consistent with previous research in which CLIL students were found to perform equally well or better than non-CLIL students in L1-related tasks or skills (e.g., Merisuo-Storm, 2007; San Isidro & Lasagabaster, 2018). In terms of bi- or multilingualism more generally, the VFT results in Article II are more suggestive of a bilingual advantage rather than a cost because the CLIL students had access to an equal or greater amount of lexis compared to the non-CLIL students. This can be contrasted with CLIL research conducted in Sweden in which students in the strong implementation of CLIL were found to be disadvantaged in terms of L1 (Holmberg, 2019; Lim Falk, 2019). Although Finland and Sweden share some overarching contextual similarities, such as the use of ELF (Peterson & Beers, 2023) and the lack of dubbing in media content (Schurz & Sundqvist, 2022), the two contrasting results reflect important differences in the amount of L1 (and L2) use in CLIL classes. While CLIL students in the strong implementations referred to by Holmberg (2019) and Lim Falk (2019) were deemed to have virtually all their instruction in English (which could be classified as *immersion*), CLIL students in the present study only had 25% of their instruction in English. The considerable difference in the amount of exposure to L1 that the students had in these two contexts cannot be ignored when comparing students' ability to perform in L1. Additionally, it is difficult to establish the extent to which other pedagogical practices may differ between the aforementioned CLIL and non-

CLIL implementations. For instance, it is possible that differences in teachers' integration of CDFs (Dalton-Puffer, 2016), use of the language quadriptych (Banegas & Mearns, 2023), and provision of scaffolding within a ZPD (Vygotsky, 1978) could, together with variations in target language use, lead to differences in language proficiencies of students within different groups (e.g., CLIL, non-CLIL, immersion). Therefore, although group-level differences in L1 ability partly reflect the amount of L1/L2 exposure received both inside and outside the classroom, it would be difficult to draw strong conclusions about any such insights without also analysing differences in pedagogical practices across those implementations.

It is also worth remembering that not all CLIL implementations reflect the same dynamic between a local language (e.g., Finnish, Swedish) and a global language (e.g., English) used as the target language in teaching. For instance, in the study by San Isidro and Lasagabaster (2018), students were already bilingual (i.e., Spanish and Galician) prior to participating in CLIL, with English constituting an L3. Additionally, given that Spanish itself can be considered an international language owing to its status as a *de jure*[6] or *de facto*[7] official language across most of Latin America (Lewis, 2009) as well as one of the six official languages of the United Nations (United Nations, n.d.), the dynamic between the CLIL language (i.e., English) and a local language such as Spanish contrasts considerably with the situation in, for example, Finland and Sweden, where the local languages do not hold such status internationally.

The results of the PNT in Article II suggested that CLIL students were not disadvantaged in terms of their L1 oral production as measured via this task. This finding reflects previous research in different contexts in which students' L1 ability was not found to be negatively affected by CLIL (e.g., Ohlsson, 2021; Pérez Cañado, 2018; Seikkula-Leino, 2007). However, this result contrasts with previous research in which bilingualism has been associated with weaker performance in such tasks (e.g., Baus et al., 2013; Faroqi-Shah et al., 2021), thereby suggesting that CLIL may, in the present context, promote a type of bi- or multilingualism that does not engender a linguistic cost in terms of L1 ability. The SD figures of the PNT results in Article II (Section 4.2, Table 10) also indicated that there was less variance among the CLIL students' performance than among that of the non-CLIL students. Additionally, contrary to the results of the non-CLIL group, the CLIL students' ability to produce valid items did not correlate with the speed at which the items were produced. This finding suggests that the CLIL students might have possessed a greater ability to navigate linguistic ambiguity in this task than the non-CLIL students, whose PNT totals correlated strongly with their RTs. The picture-based

6 Latin for *of law*; legally recognised or established.
7 Latin for *of fact*; existing in reality even if not legally recognized.

analyses conducted in Article II did not provide strong evidence of group-level differences, although they did suggest that the CLIL students may have experienced more difficulties in producing long or low-frequency words than the non-CLIL students. Overall, the PNT results from Article II suggest that CLIL students in the present context are not disadvantaged in terms of L1 ability. It is also worth noting that the students' own self-evaluations of L1 ability did not reveal significant group-level differences.

While noteworthy differences in L1 ability did not emerge from the group-level comparisons, the correlation analyses conducted using SES and L1 grades were suggestive of differences at an individual level between the CLIL and non-CLIL groups. For the non-CLIL group only, there were significant correlations between SES and both VFT and PNT totals, as well as between L1 grades and both PNT totals and RT. Firstly, the fact that SES featured prominently in correlations with the results of the non-CLIL group and not those of the CLIL group suggests that the CLIL students' L1 ability may be less strongly influenced by students' socioeconomic background than that of the non-CLIL students. This finding, together with previous research (e.g., Lorenzo et al., 2021; Rascón Moreno & Bretones Callejas, 2018), supports the notion that CLIL may mitigate the impact of SES on L1 ability, thereby promoting equality. This is further supported by the strong correlation found between SES and L1 grades in the non-CLIL group only. Suggested explanations for CLIL's possible role in reducing the impact of SES on performance include differences in CLIL and non-CLIL teachers' profiles and practices (Iwaniec & Halbach, 2021), and the creation, in CLIL classrooms, of a motivating learning environment where students' needs are met in terms of autonomy, competence and relatedness (Halbach & Iwaniec, 2020). Secondly, the non-CLIL students' performance in the PNT was found to be more strongly associated with their L1 grades than that of the CLIL group. This suggests that CLIL students' ability to perform in the task was less shaped by their ability to perform well in the school subject dedicated to their L1. Although not conclusive, this adds further weight to the notion that CLIL students' L1 ability is not determined to the same extent by the same factors as that of non-CLIL students. To sum up, these insights indicate that CLIL has the potential to reduce the extent to which factors such as SES and academic ability may otherwise lead to disparate performance in tasks related to L1 oral production.

5.1.3 The relationship between CLIL and L1 writing and speaking

In Article III, the focus was shifted to CLIL (n = 11) and non-CLIL (n = 19) students' ability to produce extended spoken and written language in response to elicitation

prompts. The data comprised raters' scores of students' language samples, leading to descriptive statistics and the analysis of group-level differences, as well as the use of Spearman's rank correlations to enable the inclusion of SES, L1 ability, and L1 self-evaluations in analyses of variation based on individual differences. Although the CLIL students' scores were lower than those of the non-CLIL students on several rubric items in both tasks, no significant differences were found between groups in L1 ability in either speaking or writing. While the results in Article III support previous research in which CLIL students were not found to be disadvantaged in terms of L1 (e.g., Ohlsson, 2021; Pérez Cañado, 2018), they contrast with those of other research conducted in different contexts where CLIL students have been found to perform better (San Isidro & Lasagabaster, 2018) and worse[8] (Holmberg, 2019; Lim Falk, 2019) than non-CLIL students in L1-related tasks.

It is important to consider the role that contextual differences may play in explaining such variation in results. For instance, the use of ELF is an established practice in both Sweden and Finland (Peterson & Beers, 2023) as is the use of subtitles instead of dubbing in media (Schurz & Sundqvist, 2022), which can lead to considerable extramural exposure to English. The combined effect of these factors is, arguably, similar for both L1 Swedish speakers in Sweden and L1 Finnish speakers in Finland[9], meaning the amount of exposure to both L1 and L2 would be comparable for such speakers. It is, therefore, unsurprising that the results in Article III reflect the results of CLIL students in the weak implementation from CLIL studies conducted in Sweden (Holmberg, 2019; Lim Falk, 2019). On the other hand, the students in the strong implementation of CLIL in those same studies from Sweden were deemed to receive virtually no instruction in their L1, which contrasts sharply with the present context in which students received 25% of their instruction in their L2 and the rest in their L1.

The contextual differences are even more salient when comparing the CLIL implementation from the present study with those used in research conducted in Spain. For instance, implementations analysed in multilingual regions of Spain (e.g., Merino & Lasagabaster, 2015; San Isidro & Lasagabaster, 2018) comprised students who were already bilingual prior to CLIL education (i.e., Spanish and one of the other official languages of Spain), with the CLIL language constituting an L3 (i.e., English). This contrasts with research conducted in other regions of Spain where the overarching context is considered to be monolingual (e.g., Navarro-Pablo & López Gándara, 2019; Pérez Cañado, 2018). Apart from differences within Spain itself, it is important to consider other sources of contextual variation between Spain and

[8] Only the results of the strong CLIL implementation were worse than those of the non-CLIL group.

[9] It is also worth noting that Swedish is typologically closer to English than Finnish.

Nordic countries such as Finland and Sweden. For example, Spain has typically been considered a country in which media is dubbed into Spanish (Matamala et al., 2017), whereas Finland and Sweden are subtitling countries (Schurz & Sundqvist, 2022), as mentioned previously. Given the prominence of English-language media both on television and on the Internet, the contrast in the amount of exposure to English in such contexts may play a considerable role in the development of L1 (e.g., Finnish, Spanish) and L2/L3 (i.e., English) skills. Moreover, the possible role played by other types of extramural exposure, such as video games and social media (Sundqvist, 2024), should not be ignored. Additionally, it is worth considering the role that ELF plays in different contexts, as alluded to previously. While the use of ELF has long been established in Sweden and Finland (Peterson & Beers, 2023), Spain has traditionally been slower than other European countries in the adoption of this practice (Tabuenca-Cuevas, 2016).

In the present context, the CLIL students were not found to be disadvantaged in terms of L1 ability as per the results of Articles II and III, despite the fact that these students live in a context where extramural exposure to English is high and the use of ELF is common. To synthesise findings from the present study and previous research, it can be concluded that the risk of CLIL students being disadvantaged in L1 ability appears to be higher if the following conditions exist simultaneously:

- students participate in a strong CLIL implementation with minimal L1 use
- the country or region has an established practice of ELF
- the country or region has a high level of extramural exposure to English

As with Article II, the correlation analyses conducted in Article III also revealed variations in individual differences between the CLIL and non-CLIL groups, particularly regarding L1 writing. There was evidence supporting the existence of a relationship between SES and L1 writing in the non-CLIL group but not in the CLIL group. There was also a strong correlation between SES and L1 grades in the non-CLIL group only. This reflects previous research in which the effect of SES on L1 ability has been found to be greater in the non-CLIL group than in the CLIL group (Lorenzo et al., 2021; Rascón Moreno & Bretones Callejas, 2018). In Article III, the results were suggestive of a relationship between L1 grades and L1 writing, but this was only in the non-CLIL group. Together with the findings from Article II, these results add weight to the notion that CLIL may promote equality by mitigating the impact of SES on L1 ability. The relationship between L1 ability and L1 grades can also be added to this equation given that the results of the present study suggest that CLIL students' L1 ability is less likely to reflect their academic ability in L1 as a subject than that of the non-CLIL students. Although the nature of the relationship between SES and both L1 writing and L1 grades seems to be clear (i.e., SES affects these two variables positively), the nature of the relationship between L1 grades and

L1 writing is not entirely clear. In other words, while it is quite possible that non-CLIL students' academic ability in L1 influences their L1 ability as measured via elicitation tasks, it is also plausible that their existing L1 ability promotes greater academic achievement in L1 as a subject. A third possibility is that a bidirectional relationship could exist between these two variables. Unfortunately, exploring this topic in depth would not be within the scope of the present study. Nevertheless, it is important to note that this multivariable interaction was only observed in the non-CLIL students' results and only for L1 writing (not L1 speaking) in Article III. In Article II, however, the relationship applied to L1 oral production as measured via both the VFT and PNT. Taken together, Articles II and III are suggestive of a relationship between SES, L1 ability (i.e., L1 oral production and writing), and L1 grades in non-CLIL students. This relationship is depicted in Figure 7, whereby SES in non-CLIL students is considered to affect L1 ability (i.e., L1 oral production and writing) and L1 grades (i.e., academic ability in L1 as a subject). However, it is unclear what the relationships is between L1 ability and L1 grades.

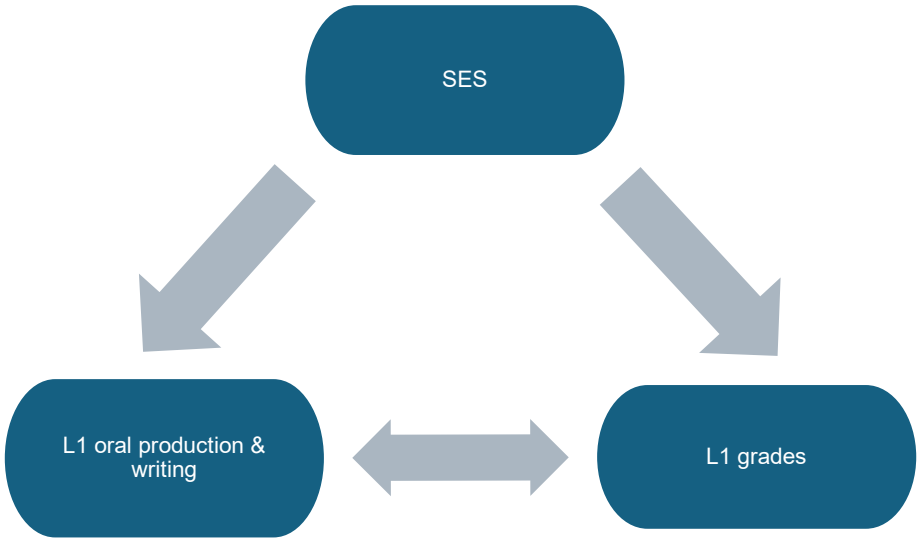

Figure 7. Non-CLIL students' SES, L1 ability and L1 grades.

5.2 Considerations for affected stakeholders

The findings from the present research underscore the need to analyse students' socioeconomic and linguistic backgrounds when considering CLIL as an alternative to mainstream education. While CLIL seems to have a positive effect on the L1 ability of students from low-SES backgrounds, it might not be the best choice if a

student is, for example, already considerably behind their peers in terms of L1 ability (e.g., because of previous expatriation outside of Finland). Therefore, key factors to consider at the admissions stage include students' SES, linguistic background (including time spent in different environments or countries), and current L1 and L2 ability. Additional variables that would provide further insight could include students' verbal intelligence, motivation and academic ability. Moreover, the decision to admit a child into a CLIL programme should, arguably, align with students' and parents'/guardians' own wishes or goals. That said, perceptions of parents/guardians have been found to vary, ranging from positive perceptions about CLIL's pedagogical benefits on the one hand (Ráez-Padilla, 2018; San Isidro & Lasagabaster, 2020) to concerns about its effect on L1 and content learning on the other (Pladevall-Ballester, 2014). Differences have also been found in perceptions among three key stakeholder groups: teachers, students and parents (Barrios & Milla Lara, 2020). As a result, it is important to engage all stakeholders in decisions concerning CLIL participation through clear communication and information-sharing with the help of research-informed perspectives.

Although students, teachers and parents/guardians could all benefit from being more research-informed, in practice it is unrealistic to place such an expectation on minors who are still developing, which is why parents/guardians and teachers have an important role to play. On the one hand, parents/guardians could, arguably, combine a basic understanding of research findings with an understanding of their child's linguistic development to make more informed decisions than those parents/guardians who are unaware of the significance of factors such as SES and time spent in different linguistic environments. Additionally, considering that CLIL has previously been considered selective and elitist (e.g., Dallinger et al., 2016) and not always suitable for all students (Lo & Lin., 2019), parents could benefit from considering the implications for a child who is, for example, already bilingual but not verbally strong or academically oriented, as CLIL may not necessarily be the best option even if it seems logical to choose CLIL to capitalise on a child's existing bilingualism. On the other hand, teachers who are both research-informed and aware of their students' backgrounds should be more able, where necessary, to take effective pedagogical decisions, such as providing additional scaffolding, ensuring differentiation, and allowing L1 and translanguaging in the classroom. In addition to teachers, the role that school administrations play is also worth emphasising, as they are invariably involved in the admissions process and in important interactions with parents/guardians (and students). Consequently, it is necessary to ensure that discussions that take place involve students, parents/guardians, teachers and the school administration responsible for admissions. In light of the results of the present study, a possible sequence of interactions between stakeholders in CLIL is proposed in Figure 8.

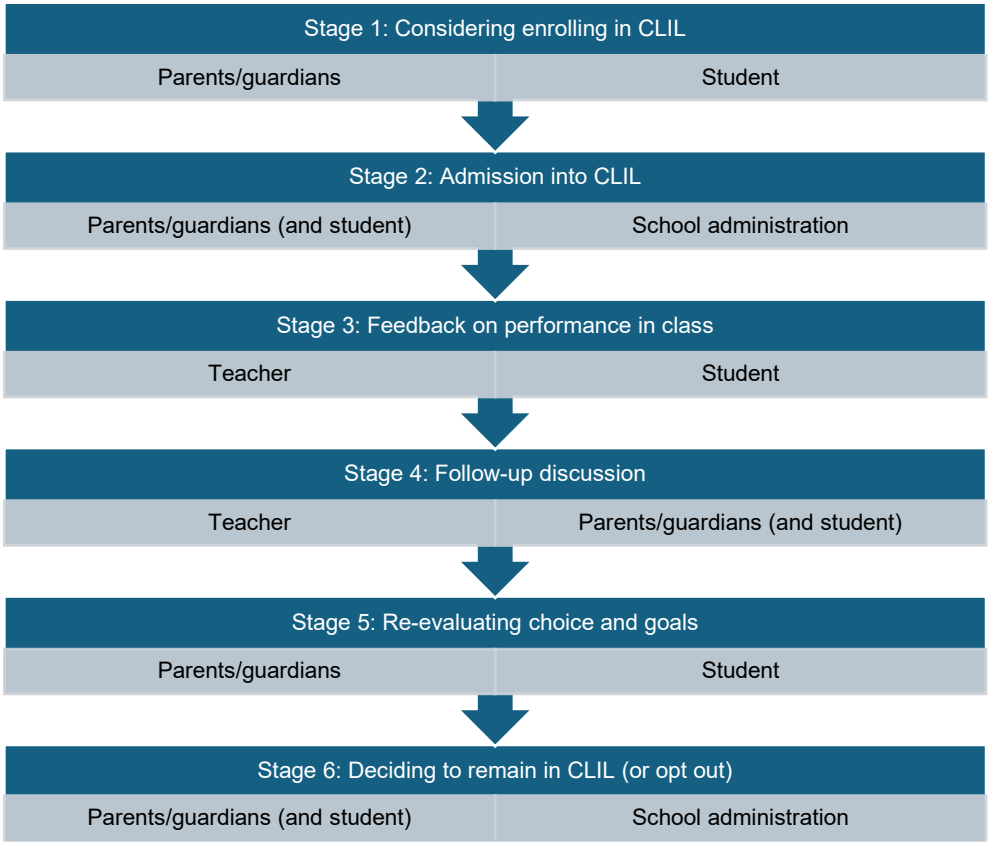

Figure 8. Sequence of stakeholder interactions concerning CLIL participation.

If admission policies and procedures of a given CLIL school are reflective of developments in, for example, research concerning the relationship between CLIL and L1 ability, discussions between the school administration and parents/guardians are likely to cover not only L2 but also L1-related goals and challenges, driven by an understanding of the child's linguistic and educational needs. All stakeholders are then likely to have an enhanced capacity to identify obstacles, communicate concerns and reassess decisions in the future. This awareness could also help to address divergences in beliefs and priorities that can emerge between different stakeholders in CLIL (e.g. see Moreno de Diezmas, 2018). As illustrated in Figure 8, the different stages of stakeholders' interactions could include, for example, family-level discussions about the possibility of participating in CLIL (Stage 1), a discussion with the school administration about the child's suitability for CLIL during the admission stage (Stage 2), interactions between the teacher and student comprising feedback about performance in class (Stage 3), follow-up discussions between the teacher and the parents or guardians (Stage 4), a family-level re-

evaluation of the suitability of CLIL (Stage 5), and a further discussion with the school administration about whether to remain in or opt out of CLIL (Stage 6). Although less likely to occur in Stage 1, discussions at all other stages could benefit from stakeholders being research informed. In practice, this could mean that schools might provide evidence-based fact sheets about CLIL participation, covering not only the benefits but also any possible challenges, for example in connection with any specific factors related to the learner or the context. Similarly, a survey could be set up and provided to families interested in a particular CLIL programme. The questions could be selected and the survey designed in such a way that it captures the information needed to identify both opportunities and possible challenges a given child may encounter in CLIL education.

5.3 Limitations

At this stage it is worth addressing the limitations of the present study in light of the scope and aims established in Section 1. Firstly, although the methods chosen in Article I and subsequently refined before the main data collection conducted for Articles II and III are oriented towards quantitative analysis, the sample sizes ended up being smaller than expected due to unforeseeable factors related to data collection, including small class sizes, restricted participant eligibility (e.g., L1s other than Finnish), limited time available with participants, as well as *force majeure* events (i.e., industrial action) on two of the scheduled days for data collection. Therefore, the generalisability of the results is naturally somewhat limited; however, the impact of this drawback is mitigated by the fact that several sources of data were gathered, constituting an approach that is reflective of methodological triangulation. Additionally, bootstrapping was used in Article II in order to increase the robustness of the results, whereas nonparametric tests were used where needed in both Articles II and III in order to overcome limitations related to the sample size and distribution. In all articles, the conclusions are drawn with considerable caution, always leaving room for alternative perspectives to be explored.

Secondly, while the study includes and accounts for SES – albeit elicited from teenage participants, who may not know their parents' precise education level – and students' linguistic backgrounds, the possible impact of variables such as verbal intelligence, working memory, motivation or years of participation in CLIL has not been explored in this cross-sectional study. Such variables could provide greater insight into how individual differences among students lead to variation in L1 ability as measured by performance on elicitation tasks. Given that CLIL education has previously been criticised as being elitist and selective (e.g., Bruton, 2011, 2013), it is important to uncover and identify how factors related to individual differences can impact language ability. That said, understanding the nature of the relationship

between bi- or multilingualism, verbal intelligence and working memory is complicated (e.g., Sarani, 2018); hence, it could not have been comprehensively addressed in the present study. To mitigate this shortcoming, the present study has incorporated students' academic achievement in L1 as a subject together with L1-related elicitation tasks in order to shed light on students' abilities from two contrasting yet somewhat converging perspectives related to L1.

Thirdly, although the present study has shed light on L1 ability through a variety of approaches (i.e., speaking, writing, lexical access and retrieval, academic achievement and self-evaluations), L1 ability remains a rather broad concept whose accurate representation may not be feasible with one (or even several) measures in a single study. In previous studies on the relationship between CLIL and L1 ability, researchers have focused on school exams or grades (e.g., Navarro-Pablo & López Gándara, 2019; Pérez Cañado, 2018; Rascón Moreno & Bretones Callejas, 2018), students' written texts (e.g., Holmberg, 2019; Lim Falk, 2019; Ohlsson, 2021), several productive and receptive skills (e.g., Merino & Lasagabaster, 2015; San Isidro & Lasagabaster, 2018), and literacy (e.g., Merisuo-Storm, 2007; Merisuo-Storm & Soininen, 2014). In light of such differences in approaches previously taken to target L1 ability, it can be concluded that a single reliable method for measuring L1 ability has not yet emerged in this area of research. Nevertheless, the present study's contribution is both consistent with previous research in which speaking and writing were scrutinised (e.g., Merino & Lasagabaster, 2015; San Isidro & Lasagabaster, 2018) while also constituting a novel contribution by incorporating psycholinguistically-oriented tests (i.e., a VFT and a PNT) used in research into the effects of bilingualism (e.g., Baus et al., 2013).

Finally, perhaps the most important aspect that has not been explored in this study is the broad set of pedagogical factors encompassed within the classroom context. Although CLIL methodology provides the overarching pedagogical framework for teachers (see Section 2.1), the impact of decisions and actions taken by individual teachers on the developing L1 ability of CLIL students should not be underestimated. For instance, teachers may employ translanguaging-related practices such as L1 use in order to promote learning in CLIL (García, 2011; Itoi, 2024; Nikula & Moore, 2016; Tsuchiya, 2019). Such practices could be utilised effectively in, for example, explanations and clarifications of concepts, translations of terminology, the provision of instructions, and bi- or multilingually-oriented classroom tasks. Additionally, teachers' purposeful integration of CDFs in teaching (e.g., encouraging students to reformulate content knowledge in their own words) together with an awareness of the language quadriptych (Banegas & Mearns, 2023) can help optimise the simultaneous learning of language and content, thereby leaving time and space for any further clarification and reflection as needed. Alternatively, teachers may utilise other pedagogical tools such as differentiation in order to cater

to students with varying needs and abilities (Roiha, 2012). This could apply with respect to students' needs and challenges in L1 even in a CLIL class, where the focus is generally on the target language and content. To sum up, a number of classroom-related factors could influence students' L1 ability differently in a CLIL class compared to a non-CLIL class, but exploring such phenomena in depth was not within the scope of the present study.

5.4 Directions for future research

In this section, possible avenues for future research are explored in light of the findings and analyses emerging from the present study. The accumulated research on the relationship between CLIL and L1 ability has covered a variety of contexts while employing different approaches to gauge L1 ability and account for individual differences among students. Although it now seems increasingly clear that CLIL participation alone does not hinder L1 ability, the contribution it makes, along with other factors, to students' developing bi- or multilingualism invariably has an impact on the students' ability to use their L1. This understanding has implications for future research on this relationship.

One area that merits additional attention in research is the impact of CLIL on the L1 ability of students in educational contexts in which classroom-based L1 use is limited and the vehicular language is English. This is of particular interest in, for example, Nordic contexts, where the use of ELF is common and students are likely to receive high levels of extramural exposure to English via, for example, video games, social media, audiovisual media, and other online content. As depicted in Table 14, in order to address key questions concerning CLIL and L1, such studies should carefully analyse the contextual considerations before proceeding to choose variables related to individual differences and establish measures to represent L1 ability.

Table 14. Considerations for research on CLIL's impact on L1 ability.

Contextual analysis	Amount of L2 and L1 use across curriculum
	Extent of extramural exposure (e.g., to media, online content)
	Extent of use of ELF
	Stage of implementation (e.g., primary education)
Individual differences	Academic ability/achievement (e.g., L1 grades)
	Existing bi- or multilingualism (and linguistic background)
	Language-learning aptitude
	Motivation (e.g., L1 learning, L2 learning, academic motivation)
	SES (e.g., parents' educational level or other metrics)
	Time spent in CLIL
	Verbal intelligence
	Working memory
L1 ability	Lexical access and retrieval
	Literacy skills
	Productive skills (i.e., speaking, writing)
	Receptive skills (i.e., listening, reading)
	Self-concept
	Vocabulary size

If there were scope to conduct longitudinal studies, then measuring L1 ability at several stages would also provide insight into the impact of CLIL on L1 development over time. Although exploring the relationship between CLIL and L1 ability is not limited to any particular research method, it is probably suited to quantitative or mixed-methods studies with sufficiently large sample sizes to enable the analysis of group-level differences while also considering the impact of several variables related to individual differences. Such methods would help in addressing the somewhat experimentally-oriented question concerning the *impact* of CLIL on L1 ability.

On the other hand, in order to explore the relationship between CLIL and L1 ability while also addressing one of the main limitations of the present study, future researchers could seek to understand what factors within the classroom context promote the development of CLIL students' L1 ability, and how. This field of inquiry is of interest because it could shed light on any unexpected, undocumented or misunderstood implications of CLIL pedagogical practices with respect to L1 ability. Although the focus of CLIL is generally on achieving its overarching learning objectives related to content and L2 learning, ensuring that L1 ability develops accordingly is consistent with the notion of *bilingual education*, which is the label used for CLIL in Finland (i.e., *kaksikielinen opetus*). Three areas in which the

relationship between CLIL classroom practices and L1 ability could be explored are detailed in Figure 9.

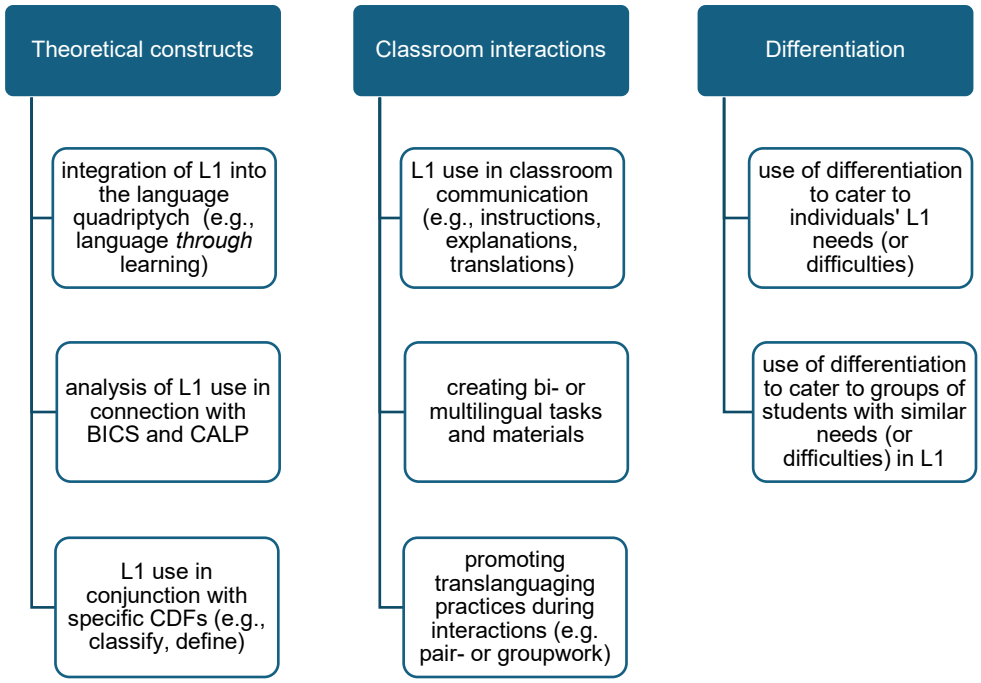

Figure 9. Perspectives for exploring the role of classroom factors in L1 development.

Firstly, researchers could focus on analysing how existing theoretical constructs such as the language quadriptych (Banegas & Mearns, 2023), BICS and CALP (Cummins, 2000), and CDFs (Dalton-Puffer, 2016) may be harnessed to promote the development of L1 ability in CLIL teaching. These theoretical considerations would naturally have implications in the classroom and could be explored via, for example, surveys and interviews conducted with CLIL teachers in order to understand their decision-making processes when planning curricula, syllabi or individual classes. Secondly, analysing the specific approaches teachers adopt for classroom interactions related to L1 use and other translanguaging methods (García & Li, 2014) may provide insight into pedagogical practices that promote L1 ability in classrooms that are typically oriented towards L2 (and content) development. One suitable method for exploring this question would be to apply discourse or conversation analysis techniques to the analysis of classroom interactions. This approach would shed light on students' immediate responses to, for example, L1 use during class. Thirdly, research focused on understanding how differentiation in CLIL (e.g., see Roiha, 2012) can be used to promote L1 development would also provide valuable

insights for both researchers and CLIL teachers. Although different research methods are possible, exploring pedagogical decisions and classroom interactions in depth is particularly conducive to using qualitative rather than quantitative methods. If there is scope and sufficient resources, research on any of these areas could be conducted in line with a grounded theory approach comprising several rounds of simultaneous data collection and analysis (Flick, 2018; Morse et al., 2021). Starting with the generation of descriptions from the analysis of saturated data, such an approach could then lead to conceptualisations or categorisations and, ultimately, a practical theory related to the development of L1 ability in CLIL.

Finally, given CLIL's inherent dual focus on language and content, future studies focusing on the relationship between CLIL and L1 ability could also benefit from scrutinising content knowledge (e.g., science) and language ability together rather than in isolation. For instance, research aiming to compare CLIL and non-CLIL students' L1 ability could not only incorporate considerations outlined in Table 14, but also use an integrated approach aimed at evaluating both content knowledge and language skills simultaneously (e.g., del Pozo & Llinares, 2021). Similarly, research exploring classroom-related phenomena, such as those outlined in Figure 9, could also incorporate a greater emphasis on the integration of content and language. For example, if teachers were interviewed about their integration of CDFs into particular aspects of a syllabus, the researcher could aim to understand teachers' perceptions concerning the role of CDFs in the development of students' content knowledge and L1 ability. Such research may shed light on CLIL's capacity to fulfil its function as a form of *bilingual education* in addition to its stated pedagogical aim of providing a dual focus on content and language.

5.5 Concluding remarks

This study sought to investigate the relationship between CLIL and L1 ability in a Finnish secondary education context. This overarching aim was approached using three sub-studies (Articles I-III) that reflected three sub-aims, as outlined in the introduction section of this dissertation. The sub-aims related to establishing methods that could be used to address the overarching research question in the given context (sub-aim 1), examining the relationship between CLIL and L1 oral production (sub-aim 2), and examining the relationship between CLIL and L1 writing and speaking (sub-aim 3).

The results for Article I illustrated some important considerations inherent to addressing the overarching research aim in the present context, such as accounting for socioeconomic status (SES) and differences in students' linguistic profiles. Additionally, the first sub-study highlighted some methodological insights that needed to be considered prior to subsequent sub-studies being conducted. For

instance, while gathering L2 data enables the comparison of proficiency in L1 and L2, using this data to address the overarching research aim would have required a different focus (e.g., crosslinguistic influence), which was, ultimately, not deemed necessary or practical in the present study. This underscores both the importance of pilot testing and the methodological benefits of undertaking a small-scale exploratory study as part of a larger research project. The second sub-study (Article II) harnessed instruments sometimes used in research in psycholinguistics and bilingualism in order to analyse speakers' abilities in lexical access and retrieval – construed as *oral production* in the present study – thus shedding light on sub-aim 2 and, subsequently, the overarching research question. To the researcher's knowledge, this is the first study in which such an approach has been taken to scrutinise the L1 ability of CLIL students in a secondary education context. The third sub-study (Article III) addressed the overarching research question by using instruments designed to elicit the production of spoken and written language that is relatively spontaneous and cohesive. In Articles II and III, including students' SES and grades in their L1 as a subject in the analyses provided additional insight into how background variables may impact CLIL students' L1 ability differently from that of non-CLIL students.

The study fills a gap in research concerning the relationship between CLIL and L1 ability, specifically in a Finnish education context. The findings suggest that CLIL students' L1 ability is not adversely affected by their participation in a CLIL implementation in the present context. This finding is reflective of results obtained in previous research in which CLIL students were not found to be disadvantaged in terms of L1 ability (e.g., Ohlsson, 2021; Pérez Cañado, 2018; San Isidro & Lasagabaster, 2018). That said, the study highlights the importance of considering contextual factors, such as extramural exposure and the use of English as a *lingua franca* (ELF), as well as pedagogical differences among implementations or groups (e.g., the amount of L2 use in teaching), all of which may contribute to considerable variation in results across different CLIL contexts (see Section 5.1.2 and 5.1.3). Additionally, the results from Articles II and III indicated that the CLIL students' L1 ability appeared not to be influenced by their SES to the same extent as that of the non-CLIL students. This finding aligns with previous research in which CLIL has been found to reduce the impact of SES on L1 ability (Lorenzo et al., 2021; Rascón Moreno & Bretones Callejas, 2018). Such findings are suggestive of CLIL's potential in promoting equality via a type of *levelling effect* (Halbach & Iwaniec, 2020; Iwaniec & Halbach, 2021), which may boost the performance of low-SES students while, admittedly, seeming to have a restrictive effect on the outcomes of high-SES students.

In light of the results of this study, possible areas of future research emerged along with methodological considerations that could help provide greater insight into

the relationship between CLIL and L1 ability. Firstly, the existing body of research could benefit from additional studies that seeks to understand the *impact* of CLIL on L1 ability. Although addressing this specific question was not within the aims or scope of the present study, considerations related to this question are outlined in Section 5.4 (see Table 14). Future researchers seeking to determine the impact of CLIL on L1 ability may wish to employ quantitative or mixed-methods approaches in order to measure the impact of the independent or grouping variable (CLIL or non-CLIL) on the dependent variable (L1 ability) while considering possible covariates (e.g., verbal intelligence, motivation). Secondly, the study highlights that further research could be conducted to gain insight into how classroom-related factors may influence L1 development in CLIL (see Section 5.4, Figure 9). The use of qualitative methods in particular could facilitate an understanding of how teachers harness, for example, cognitive discourse functions (CDFs), translanguaging or differentiation in order to promote students' development not only in L2 and content knowledge, but also in L1. To sum up, the present study outlines several possible avenues of exploration that would build on current knowledge concerning the relationship between CLIL and L1 ability.

Lastly, in a broad sense, this study has contributed to the discussion about the role of *bilingual education* in the context of an overarching *multilingual* European Union (EU). Despite multilingualism already being clearly established in some countries, such as Luxembourg (de Bres et al., 2019), the EU's complex political, cultural and linguistic climate forces it to grapple with the challenge of simultaneously maintaining linguistic diversity (e.g., Grindheim & Lohndal, 2008; Soldat-Jaffe, 2014) and accepting the increasingly dominant role of English for intercultural communication (Gnutzmann et al., 2015). It is against the backdrop of this linguistic and political tension that the present study is situated. Moreover, in scrutinising the relationship between CLIL and L1 ability in Finland, the study implicitly assumes that named languages (e.g., Finnish) not only exist but are worth analysing despite not explicitly forming part of CLIL's dual pedagogical aim of promoting the learning of content and the target language (e.g., English in Finland). Thus, the perspective adopted in the present study contrasts somewhat with the view that named languages are a social construct (e.g., Chen, 2015; Li & García, 2017, 2022). Moreover, although the framework adopted in the present study could be construed as representing a *monolingual* rather than a *multilingual* perspective on learners and speakers of languages (see Conteh & Meier, 2014; May, 2014), it could equally be claimed that the existence of named languages is *real* and has *real-world implications* in individual EU countries, many of which still reflect monolingual characteristics (Guglielmi, 2020). It is in line with this kind of *realism* that the present study has been conducted.

Abbreviations

AoA	age of acquisition
CBI	content-based instruction
CDF	cognitive discourse functions
CEFR	Common European Framework of Reference for Languages
CLIL	content and language integrated learning
CLT	communicative language teaching
EMI	English-medium instruction
ELF	English as a *lingua franca*
EU	European Union
L1	first language
L2	second language
L3	third language
PNT	picture-naming task
RT	response time
SD	standard deviation
SES	socioeconomic status
SLA	second language acquisition
TBL	task-based learning
TENK	The Finnish National Board on Research Integrity
TESOL	teaching English to speakers of other languages
TOT	tip-of-the-tongue moment
VFT	verbal fluency task
ZPD	Zone of Proximal Development

List of References

Admiraal, W., Westhoff, G., & de Bot, K. (2006). Evaluation of bilingual secondary education in the Netherlands: Students' language proficiency in English. *Educational Research and Evaluation*. *12*(1), 75-93. https://doi.org/10.1080/13803610500392160

Ai, B., Hao, M., & Qiao, X. (2022). Unpacking Translanguaging Practices in Multilingual Business Communication in China: A Qualitative Phenomenological Approach. *Applied Linguistics*, *43*(6), 1184–1206. https://doi.org/10.1093/applin/amac029

Airey, J. (2004). Can you teach it in English? Aspects of the language choice debate in Swedish higher education. In R. Wilkinson (Ed.), *Integrating content and Language: Meeting the challenge of a multilingual higher education* (pp. 97-109). Maastricht University Press.

Aita, S. L., Beach, J. D., Taylor, S. E., Borgogna, N. C., Harrell, M. N., & Hill, B. D. (2019). Executive, language, or both? An examination of the construct validity of verbal fluency measures. *Applied Neuropsychology. Adult*, *26*(5), 441–451. https://doi.org/10.1080/23279095.2018.1439830

Alasal, M. (2025). Exploring the Role of Task-Based Language Teaching in Enhancing Communicative Competence. *International Journal of Linguistics, Literature and Translation*, *8*(2), 48–59. https://doi.org/10.32996/ijllt.2025.8.2.6

Amondarain-Garrido, M., & Ruiz de Zarobe, Y. (2024). Looking into multimodal translanguaging in CLIL: a case study in the Basque Country. *Journal of Multilingual and Multicultural Development*, 1–19. https://doi.org/10.1080/01434632.2024.2429419

Anghel, B., Cabrales, A., & Carro, J. M. (2015). Evaluating a bilingual education program in Spain: The impact beyond foreign language learning. *Economic Inquiry, 54*(2), 1202-1223. https://doi.org./10.1111/ecin.12305

Antoniou, M., Liang, E., Ettlinger, M., & Wong, P. C. M. (2015). The bilingual advantage in phonetic learning. *Bilingualism (Cambridge, England), 18*(4), 683–695. https://doi.org/10.1017/S1366728914000777

Arai, S. (2023). Postmodernism. In F. Maggino (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. 5343–5346). Springer International Publishing. https://doi.org/10.1007/978-3-031-17299-1_2223

Arens, A.K., & Jansen, M. (2016). Self-concepts in reading, writing, listening, and speaking: A multidimensional and hierarchical structure and its generalizability across native and foreign languages. *Journal of Educational Psychology, 108*(5), 646–664. https://doi.org/10.1037/edu0000081

Aro, S., & Mikkilä-Erdmann, M. (2014). School-external Factors in Finnish Content and Language Integrated Learning (CLIL) Programs. *Scandinavian Journal of Educational Research*, *59*(2), 127–142. https://doi.org/10.1080/00313831.2014.894937

Audi, R. (2010). *Epistemology: A Contemporary Introduction to the Theory of Knowledge* (3rd ed.). Routledge. https://doi.org/10.4324/9780203846469

Aziez, F., Alkhrisheh, H., Aziez, F., & Mualim, M. (2020). Measuring Verbal Fluency Task Performance of Indonesian Bilinguals. *Indonesian Research Journal in Education |IRJE|, 4*(2), 306–323. https://doi.org/10.22437/irje.v4i2.9429

Banegas, D. L. (2019). Teacher professional development in language-driven CLIL: A case study. *Latin American Journal of Content & Language Integrated Learning*, *12*(2), 242-264. https://doi.org/10.5294/laclil.2019.12.2.3

Banegas, D. L., & Mearns, T. (2023). The language quadriptych in content and language integrated learning: Findings from a collaborative action research study. *Journal of Multilingual and Multicultural Development,* 1–16. https://doi.org/10.1080/01434632.2023.2281393

Barrios, E. (2021). CLIL and L1 Competence Development. In: Pérez Cañado, M.L. (eds) *Content and Language Integrated Learning in Monolingual Settings*. Multilingual Education, vol 38. Springer, Cham. https://doi.org/10.1007/978-3-030-68329-0_6

Barrios, E., & Milla Lara, M. D. (2020). CLIL methodology, materials and resources, and assessment in a monolingual context: an analysis of stakeholders' perceptions in Andalusia. *Language Learning Journal*, *48*(1), 60–80. https://doi.org/10.1080/09571736.2018.1544269

Baus, C., Costa, A., & Carreiras, M. (2013). On the effects of second language immersion on first language production. *Acta Psychologica, 142*(3), 402–409. https://doi.org/10.1016/j.actpsy.2013.01.010

Benati, A. (2018). Direct Method. In M. DelliCarpini & J. I. Liontas (Eds.), *The TESOL Encyclopedia of English Language Teaching* (pp. 1–5). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118784235.eelt0154

Białecka, M., Wodniecka, Z., Muszyńska, K., Szpak, M., & Haman, E. (2023). Both L1 and L2 proficiency impact ToM reasoning in children aged 4 to 6. Painting a more nuanced picture of the relation between bilingualism and ToM. *Bilingualism (Cambridge, England)*, 1–19. https://doi.org/10.1017/S1366728923000652

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: consequences for mind and brain. *Trends in Cognitive Sciences*, *16*(4), 240–250. https://doi.org/10.1016/j.tics.2012.03.001

Bian, C. (2017). The potential of transnational language policy to promote social inclusion of immigrants: An analysis and evaluation of the European Union's INCLUDE project. *International Review of Education*, *63*(4), 475–494. https://doi.org/10.1007/s11159-017-9655-0

Björklund, S. (2019). 3. Research trends and future challenges in Swedish immersion. In M. Haneda & H. Nassaji (Ed.), *Perspectives on Language as Action* (pp. 45-60). Bristol, Blue Ridge Summit: Multilingual Matters. https://doi.org/10.21832/9781788922944-006

Brown, H., & Bradford, A. (2017). EMI, CLIL, & CBI: Differing approaches and goals. In P. Clements, A. Krause, & H. Brown (Eds.), *Transformation in language education*. JALT

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

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

Bulté, B., Surmont, J., & Martens, L. (2022). The impact of CLIL on the L2 French and L1 Dutch proficiency of Flemish secondary school pupils. *International Journal of Bilingual Education and Bilingualism*, *25*(9), 3151–3170. https://doi.org/10.1080/13670050.2021.2018400

Bylund, E., Abrahamsson, N., Hyltenstam, K., & Norrman, G. (2019). Revisiting the bilingual lexical deficit: The impact of age of acquisition. *Cognition*, *182*, 45–49. https://doi.org/10.1016/j.cognition.2018.08.020

Cenoz, J. (2015). Content-based instruction and content and language integrated learning: The same or different? Language, *Culture, and Curriculum, 28*(1), 8-24. https://doi.org/10.1080/07908318.2014.1000922

Cenoz, J., Genesee, F., & Gorter, D. (2014) Critical Analysis of CLIL: Taking Stock and Looking Forward, *Applied Linguistics, 35*(3), 243–262, https://doi.org/10.1093/applin/amt011

Chen, B. (2015). Social Constructivism of Language and Meaning. *Croatian Journal of Philosophy*, *XV*(43), 87–113.

Chunliu, X., & Guangsheng, L. (2025). The effectiveness of task-based language teaching in improving communicative competence. *International Journal of Advanced and Applied Sciences*, *12*(1), 263–273. https://doi.org/10.21833/ijaas.2025.01.025

City Digital Group (2021). *Suomi24 virkkeet -korpus 2001-2020, Korp-versio [korpus]*. Kielipankki. Retrieved from: http://urn.fi/urn:nbn:fi:lb-2021101525

Cleary, A. M. (2017). 2.24 - Tip-of-the-Tongue States. In *Learning and Memory: A Comprehensive Reference* (Second Edition, pp. 433–449). Elsevier Ltd. https://doi.org/10.1016/B978-0-12-809324-5.21059-6

Consoli, S., & Ganassin, S. (2023). *Reflexivity in Applied Linguistics: Opportunities, Challenges, and Suggestions* (1st ed.). Routledge. https://doi.org/10.4324/9781003149408

Conteh, J., & Meier, G. (2014). The Multilingual Turn in Languages Education: Opportunities and Challenges. In *Multilingual Matters* (1st ed., Vol. 40). Multilingual Matters.

Council of Europe (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume.* Council of Europe Publishing, Strasbourg, available at www.coe.int/lang-cefr.

Coyle, D. (2007). Content and Language Integrated Learning: Towards a Connected Research Agenda for CLIL Pedagogies. *International Journal of Bilingual Education and Bilingualism*, *10*(5), 543–562. https://doi.org/10.2167/beb459.0

Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and language integrated learning*. Cambridge University Press.

Cummins, J. (1998). Immersion education for the millennium: What have we learned from 30 years of research on second language immersion? In M. R. Childs & R. M. Bostwick (Eds.) *Learning through two languages: Research and practice. Second Katoh Gakuen International Symposium on Immersion and Bilingual Education*. (pp. 34-47). Katoh Gakuen, Japan.

Cummins, J. (2000). Language Proficiency in Academic Contexts. In *Language, Power and Pedagogy*. United Kingdom: Multilingual Matters.

Cummins, J., & Swain, M. (1986). *Bilingualism in education: Aspects of theory, research and practice* (1st ed.). Longman. https://doi.org/10.4324/9781315835877

Dallinger, S., Jonkmann, K., & Hollm, J. (2016). Selectivity of content and language integrated learning programmes in German secondary schools. *International Journal of Bilingual Education and Bilingualism*, *21*(1), 93–104. https://doi.org/10.1080/13670050.2015.1130015

Dalton-Puffer, C. (2011). Content-and-Language Integrated Learning: From Practice to Principles? *Annual Review of Applied Linguistics, 31*, 182–204. https://doi.org/10.1017/S0267190511000092

Dalton-Puffer, C. (2016). Cognitive discourse functions: Specifying an integrative interdisciplinary construct. In T. Nikula, E. Dafouz, P. Moore, & U. Smit (Eds.), *Conceptualising integration in CLIL and multilingual education* (pp. 29–54). Multilingual Matters. https://doi.org/10.21832/9781783096145-005

Dalton-Puffer, C., Llinares, A., Lorenzo, F., & Nikula, T. (2014). "You can stand under my umbrella": Immersion, CLIL and bilingual education. A response to Cenoz, Genesee & Gorter (2013). *Applied Linguistics*, *35*(2), 213-218. https://doi.org/10.1093/applin/amu010

de Bres, J., Rivera Cosme, G., & Remesch, A. (2019). Walking the tightrope of linguistic nationalism in a multilingual state: constructing language in political party programmes in Luxembourg. *Journal of Multilingual and Multicultural Development*, *41*(9), 779–793. https://doi.org/10.1080/01434632.2019.1655567

del Pozo, E., & Llinares, A. (2021). Assessing Students' Learning of History Content in Spanish CLIL Programmes: A Content and Language Integrated Perspective. In D. L. Banegas & C. Hemmi (Eds.), *International Perspectives on CLIL* (pp. 43–61). Springer International Publishing. https://doi.org/10.1007/978-3-030-70095-9_3

Dubé, D., & Thordardottir, E. (2024). Using semantic verbal fluency to estimate the relative and absolute vocabulary size of bilinguals: An exploratory study of children and adolescents. *Journal of Communication Disorders*, *111*, Article 106450. https://doi.org/10.1016/j.jcomdis.2024.106450

Edwards, J. (2012). Bilingualism and Multilingualism: Some Central Concepts. In W. C. Ritchie & T. K. Bhatia (Eds.), *The Handbook of Bilingualism and Multilingualism* (pp. 5–25). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118332382.ch1

Effingham, N. (2013). *An Introduction to Ontology*. Cambridge: Polity.

Elomaa, M. (2000). *Suomen ensimmäisten kielikylpyläisten äidinkieliset kirjoitelmat peruskoulussa*. Universitas Wasaensis.

Enstice, E. M. (2017). Latino Parent Perspectives: How to Promote and Implement Additive Bilingualism. *Journal for Leadership and Instruction*, *16*(1), 33-36.

European Commission (1995). *White paper on education and training: Teaching and learning — towards the learning society*. Retrieved from https://europa.eu/documents/comm/white_papers/pdf/com95_590_en.pdf

European Parliament & Council of the European Union (1998). *Council resolution of 20 November 1997 on a European year of languages (1998)*. Retrieved from https://eurlex.europa.eu/legal-content/EN/TXT/?uri=CELEX:31998Y0103(01)

European Union. (n.d.). *Languages, multilingualism, language rules*. European Union. https://european-union.europa.eu/principles-countries-history/languages_en

Evnitskaya, N., & Dalton-Puffer, C. (2020). Cognitive discourse functions in CLIL classrooms: eliciting and analysing students' oral categorizations in science and history. *International Journal of Bilingual Education and Bilingualism*, *26*(3), 311–330. https://doi.org/10.1080/13670050.2020.1804824

Faroqi-Shah, Y., Kevas, Y., & Li, R. (2021). Lexical category differences in bilingual picture naming: Implications for models of lexical representation. *Bilingualism (Cambridge, England)*, *24*(5), 849–863. https://doi.org/10.1017/S1366728921000213

Fernández-Sanjurjo, J., Fernández-Costales, A., & Arias Blanco, J. M. (2017). Analysing students' content-learning in science in CLIL vs. non-CLIL programmes: empirical evidence from Spain. *International Journal of Bilingual Education and Bilingualism*, *22*(6), 661–674. https://doi.org/10.1080/13670050.2017.1294142

Finnish National Agency for Education (2014). *National core curriculum for basic education 2014*. Finnish National Agency for Education.

Flick, U. (2018). Background: Approaches and philosophies of grounded theory. In *Doing Grounded Theory* (pp. 1–16). SAGE Publications Ltd. https://doi.org/10.4135/9781529716658.n1

García, O. (2011). Educating New York's bilingual children: constructing a future from the past. *International Journal of Bilingual Education and Bilingualism*, *14*(2), 133–153. https://doi.org/10.1080/13670050.2010.539670

García, O., & Li, W. (2014). *Translanguaging : Language, Bilingualism and Education* (1st ed. 2014.). Palgrave Macmillan UK. https://doi.org/10.1057/9781137385765

García, O., & Lin, A. (2017). Translanguaging in Bilingual Education. In: García, O., Lin, A., May, S. (Eds.), *Bilingual and Multilingual Education. Encyclopedia of Language and Education*. Springer, Cham. https://doi.org/10.1007/978-3-319-02258-1_9

Gerns, P. (2023). Building Scientific Knowledge in English: Integrating Content, Cognition and Communication in Secondary School CLIL Biology. *Journal of Language and Education*, *9*(3), 52–78. https://doi.org/10.17323/jle.2023.17569

Gerns, P. (2023). Qualitative insights and a first evaluation tool for teaching with cognitive discourse function: "comparing" in the CLIL science classroom. *Porta Linguarum*, *2023*(40), 161–179. https://doi.org/10.30827/portalin.vi40.26619

Gnutzmann, C., Jakisch, J., & Rabe, F. (2015). Communicating across Europe: What German students think about multilingualism, language norms, and English as a lingua franca. In A. Linn, N. Bermel, & G. Ferguson (Eds.), *Attitudes towards English in Europe* (Vol. 1, pp. 165–192). De Gruyter Mouton. https://doi.org/10.1515/9781614515517-009

Gollan, T. H., Ferreira, V. S., Cera, C., & Flett, S. (2013). Translation-priming effects on tip-of-the-tongue states. *Language and Cognitive Processes*, *29*(3), 274–288. https://doi.org/10.1080/01690965.2012.762457

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*(3), 787–814. https://doi.org/10.1016/j.jml.2007.07.001

Goris, J., Denessen, E., & Verhoeven, L. (2019). Effects of content and language integrated learning in Europe A systematic review of longitudinal experimental studies. *European Educational Research Journal EERJ*, *18*(6), 675–698. https://doi.org/10.1177/1474904119872426

Grindheim, J. E., & Lohndal, T. (2008). Lost in Translation? European Integration and Language Diversity. *Perspectives on European Politics and Society*, *9*(4), 451–465. https://doi.org/10.1080/15705850802416879

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language, 36*(1), 3–15.

Grosjean, F. (2010). *Bilingual: Life and reality*. Harvard University Press.

Guglielmi, S. (2020). Monolingualism and national identity: Lessons from Europe. In K. Strani (Ed.), *Multilingualism and politics: Revisiting multilingual citizenship* (pp. 207–230). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-40701-8_9

Gülle, T., & Nikula, T. (2024). *Overview of CLIL provision in Europe and country-specific insights: A report by CLILNetLe Working Group 1* (Version 1). PHAIDRA Repository, University of Vienna. https://doi.org/10.25365/phaidra.511

Halbach, A., & Iwaniec, J. (2020). Responsible, competent and with a sense of belonging: an explanation for the purported levelling effect of CLIL. *International Journal of Bilingual Education and Bilingualism*, *25*(5), 1609–1623. https://doi.org/10.1080/13670050.2020.1786496

Haman, E., Wodniecka, Z., Marecka, M., Szewczyk, J., Białecka-Pikul, M., Otwinowska, A., … Foryś-Nogala, M. (2017). How does L1 and L2 exposure impact L1 performance in bilingual children? Evidence from Polish-English migrants to the United Kingdom. *Frontiers in Psychology*, *8*, 1444–1444. https://doi.org/10.3389/fpsyg.2017.01444

Hernandez Santacruz, J., Torres, J., & Kroll, J. F. (2025). Heritage speakers reveal the dynamics of bilingual language regulation. *Bilingualism (Cambridge, England)*, 1–14. https://doi.org/10.1017/S1366728925000203

Holmberg, P. (2019). The development of academic vocabulary in Swedish. In L. Sylvén (Ed.), *Investigating content and language integrated learning: Insights from Swedish high schools* (pp. 173-186). Multilingual Matters. https://doi.org/10.21832/9781788922425-016

Ikeda, M. (2022). "How would you like your CLIL?": Conceptualising Soft CLIL. In *Soft CLIL and English Language Teaching* (1st ed., pp. 9–26). Routledge. https://doi.org/10.4324/9780429032332-1-2

Itoi, K. (2024). Fostering Inclusive Learning and 21st-Century Skills: Creating Translanguaging Spaces in University Content and Language Integrated Learning Courses. *International Journal of Applied Linguistics*. https://doi.org/10.1111/ijal.12643

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica*, *127*(2), 277–288. https://doi.org/10.1016/j.actpsy.2007.06.003

Iwaniec, J., & Halbach, A. (2021). Teachers' views on their methodology and their profiles: in search of the possible reasons for the levelling effect of CLIL. *Journal of Multilingual and Multicultural Development*, *45*(4), 1112–1126. https://doi.org/10.1080/01434632.2021.1946548

Jacobs, H. L., S. A. Zingraf, D. R. Wormuth, V. F. Hartfiel, and J. B. Hughey. (1981). *Testing ESL composition: A practical approach*. Newbury House.

Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition*. New York and London: Routledge.

Jódar Sánchez, J. A., & Tuomainen, S. (2014). English in Finland: Opinions and attitudes on the use of anglicisms. *Estudios de Lingüística Inglesa Aplicada (ELIA)*, 14, 97–144. https://doi.org/10.12795/elia.2014.i14.05

Klewitz, B. (2021). *Content and Language Integrated Learning (CLIL): A Methodology of Bilingual Teaching* (Vol. 1). Hannover: ibidem.

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon.

Kreiner, H., & Degani, T. (2015). Tip-of-the-tongue in a second language: The effects of brief first-language exposure and long-term use. *Cognition*, *137*(Apr), 106–114. https://doi.org/10.1016/j.cognition.2014.12.011

Laitinen, M. Leppänen, S, Rautionaho, P. & Backman, S. (2023). Englanti Suomen kansalliskielten rinnalla: Kohti joustavaa monikielisyyttä. Policy brief: valtioneuvoston selvitys ja tutkimustoiminta. Valtioneuvoston kanslia. http://urn.fi/URN:ISBN:978-952-383-055-4

Laki peruskoululain muuttamisesta 8.2.1991/261 [Act amending the Basic Education Act]. Retrieved from http://www.finlex.fi/fi/laki/alkup/1991/19910261

Lantolf, J. P. (2011). The Sociocultural Approach to Second Language Acquisition: Sociocultural theory, second language acquisition, and artificial L2 development. In D. Atkinson (Ed.), *Alternative Approaches to Second Language Acquisition* (1st ed., pp. 24–47). Routledge. https://doi.org/10.4324/9780203830932-2

Lasagabaster, D., & Sierra, J. M. (2010). Immersion and CLIL in English: More differences than similarities. *ELT Journal*, *64*, 367-375. https://do.org/10.1093/elt/ccp082

Leppänen, S., Pitkänen-Huhta, A., Nikula, T., Kytölä, S., Törmäkangas, T., Nissinen, K., Kääntä, L., Räisänen, T., Laitinen, M., Pahta, P., Koskela, H., Lähdesmäki, S., & Jousmäki, H. (2011). *National survey on the English language in Finland: Uses, meanings and attitudes.*

Li, W. (2018). Translanguaging as a Practical Theory of Language. *Applied Linguistics*, *39*(1), 9–30. https://doi.org/10.1093/applin/amx039

Li, W., & García, O. (2017). From Researching Translanguaging to Translanguaging Research. In K. Kendall, Y. J. Lai, & S. May (Eds.), *Research Methods. Encyclopedia of Language and Education* (pp. 227-240.). Springer. https://doi.org/10.1007/978-3-319-02249-9_16

Li, W., & García, O. (2022). Not a First Language but One Repertoire: Translanguaging as a Decolonizing Project. *RELC Journal*, *53*(2), 313–324. https://doi.org/10.1177/00336882221092841

Lim Falk, M. (2019). The development of linguistic correctness in CLIL and non-CLIL students' writing in the L1 at upper secondary school. In L. Sylvén (Ed.), *Investigating content and language integrated learning: Insights from Swedish high schools* (pp. 187-215). Multilingual Matters. https://doi.org/10.21832/9781788922425-017

Lindstromberg, S. (2016). Inferential statistics in Language Teaching Research: A review and ways forward. *Language Teaching Research : LTR*, *20*(6), 741–768. https://doi.org/10.1177/1362168816649979

Llinares, A. (2024). CLIL and linguistics. In D. L. Banegas & S. Zappa-Hollman (Eds.), *The Routledge handbook of content and language integrated learning* (1st ed., Vol. 1, pp. 11–27). Routledge. https://doi.org/10.4324/9781003173151-3

Llinares, A., & Nikula, T. (2023). CLIL students' production of cognitive discourse functions: Comparing Finnish and Spanish contexts. *Language and Education*, *38*(3), 381–400. https://doi.org/10.1080/09500782.2023.2211049

Lo, Y. Y., & Lin, A. M. Y. (2019). Content and language integrated learning in Hong Kong. In X. Gao (Ed.), *Second handbook of English language teaching* (pp. 963–982). Springer International Publishing. https://doi.org/10.1007/978-3-030-02899-2_50

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie, & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). *New York: Academic Press.

Lorenzo, F., Granados, A., & Rico, N. (2021). Equity in bilingual education: Socioeconomic status and content and language integrated learning in monolingual southern Europe. *Applied Linguistics*, *42*(3), 393–413. https://doi.org/10.1093/applin/amaa037

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*(4), 940-967. https://doi.org/10.1044/1092-4388(2007/067)

Marko, M., Michalko, D., Dragašek, J., Vančová, Z., Jarčušková, D., & Riečanský, I. (2022). Assessment of Automatic and Controlled Retrieval Using Verbal Fluency Tasks. *Assessment (Odessa, Fla.)*, *30*(7), 2198–2211. https://doi.org/10.1177/10731911221117512

Matamala, A., Perego, E., & Bottiroli, S. (2017). Dubbing versus subtitling yet again?: An empirical study on user comprehension and preferences in Spain. *Babel (Frankfurt)*, *63*(3), 423–441. https://doi.org/10.1075/babel.63.3.07mat

May, S. (2014). Introducing the "Multilingual Turn." In *The Multilingual Turn* (1st ed., pp. 1–6). Routledge. https://doi.org/10.4324/9780203113493-1

Meier, G. (2017). The multilingual turn as a critical movement in education: assumptions, challenges and a need for reflection. *Applied Linguistics Review*, *8*(1), 131–161. https://doi.org/10.1515/applirev-2016-2010

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

Merisuo-Storm, T. (2007). Pupils' attitudes towards foreign-language learning and the development of literacy skills in bilingual education. *Teaching and Teacher Education, 23*(2), 226–235. https://doi.org/10.1016/j.tate.2006.04.024

Merisuo-Storm, T., & Soininen, M. (2014). Students' first language skills after six years in bilingual education. *Mediterranean Journal of Social Sciences*, *5*(22), 72–81. https://doi.org/10.5901/mjss.2014.v5n22p72

Met. M. (1999, January). Content-based instruction: Defining terms, making decisions. *NFLC Reports.* Washington, DC: The National Foreign Language Center.

Meydan, C. H., & Akkaş, H. (2024). The role of triangulation in qualitative research: Converging perspectives. In A. Elhami, A. Roshan, & H. Chandan (Eds.), *Principles of conducting qualitative research in multicultural settings* (pp. 101–132). IGI Global. https://doi.org/10.4018/979-8-3693-3306-8.ch006

Meyerhöffer, N., & Dreesmann, D. C. (2019). 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*, *41*(1), 1–20. https://doi.org/10.1080/09500693.2018.1529446

Moreno de Diezmas, E. (2018). Students, teachers, and management teams in bilingual programmes: Shared perceptions and areas for improvement. *Journal of English Studies*, *17*(17), 277–297. https://doi.org/10.18172/jes.3564

Morse, J. M., Bowers, B. J., Clarke, A. E., Charmaz, K., Corbin, J., & Porr, C. J. (2021). The maturation of grounded theory. In J. M. Morse, P. N. Stern, J. Corbin, B. J. Bowers, K. Charmaz, & A. E. Clarke (Eds.), *Developing grounded theory: The second generation revisited* (pp. 3–22). Routledge. https://doi.org/10.4324/9781315169170

Muñoz, C., & Cadierno, T. (2021). How do differences in exposure affect English language learning? A comparison of teenagers in two learning environments. *Studies in Second Language Learning and Teaching*, *11*(2), 185–212. https://doi.org/10.14746/ssllt.2021.11.2.2

Navarro-Pablo, M., & López Gándara, Y. (2019). The effects of CLIL on L1 competence development in monolingual contexts. *The Language Learning Journal: Bilingual Education in Monolingual Contexts, 48*(1), 18–35. https://doi.org/10.1080/09571736.2019.1656764

Nikula, T., & Mård-Miettinen, K. (2014). Language learning in immersion and CLIL classrooms. In J.-O. Östman, & J. Verschueren (Eds.), *Handbook of pragmatics: 2014 Installment*. John Benjamins Publishing Company. https://doi.org/10.1075/hop.18.lan10

Nikula, T., & Moore, P. (2016). Exploring translanguaging in CLIL. *International Journal of Bilingual Education and Bilingualism*, *22*(2), 237–249. https://doi.org/10.1080/13670050.2016.1254151

Nikula, T., Skinnari, K., & Mård-Miettinen, K. (2022). Diversity in CLIL as experienced by Finnish CLIL teachers and students: Matters of equality and equity. *International Journal of Bilingual Education and Bilingualism, 26*(9), 1066–1079. https://doi.org/10.1080/13670050.2022.2028125

Nishimoto, T., Miyawaki, K., Ueda, T., Une, Y., & Takahashi, M. (2005). Japanese normative set of 359 pictures. *Behavior Research Methods, 37*(3), 398–416. https://doi.org/10.3758/BF03192709

Ohlsson, E. (2021). Perspectives on written L1 in Swedish CLIL education. *Apples* (Jyväskylä, Finland), *15*(2), 19-41. https://doi.org/10.47862/apples.98178

Paradis, J., Genesee, F., & Crago, M. B. (2011). *Dual language development and disorders: A handbook on bilingualism and second language learning* (2nd ed.). Brookes Publishing.

Peltonen, P., & Lintunen, P. (2022). Multilingual speakers' L1, L2, and L3 fluency across languages: A study of Finnish, Swedish, and English. *Nordand*, *17*(1), 48–63. https://doi.org/10.18261/nordand.17.1.4

Peltoniemi, A., Skinnari, K., Mård-Miettinen, K., & Sjöberg, S. (2018). *Monella kielellä Suomen kunnissa 2017. Selvitys muun laajamittaisen ja suppeamman kaksikielisen varhaiskasvatuksen, esiopetuksen ja perusopetuksen tilanteesta*. Jyväskylä: Jyväskylän yliopisto. https://urn.fi/URN:ISBN:978-951-39-7391-9

Pennycook, A. (2006). Postmodernism in language policy. In T. Ricento (Ed.), *An introduction to language policy: Theory and method* (pp. 60–76). Blackwell Publishing.

Pérez Cañado, M. (2011). 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

Pérez Cañado, M. (2018). The effects of CLIL on L1 and content learning: Updated empirical evidence from monolingual contexts. *Learning and Instruction*, *57*, 18-33. https://doi.org/10.1016/j.learninstruc.2017.12.002

Pérez Cañado, M. (2019). CLIL and elitism: myth or reality? *The Language Learning Journal: Bilingual Education in Monolingual Contexts, 48*(1), 4–17. https://doi.org/10.1080/09571736.2019.1645872

Pérez-Jara, J., Romero, G. E., Camprubí, L. (2022). What is Materialism? History and Concepts. In: Romero, G.E., Pérez-Jara, J., Camprubí, L. (Eds.) *Contemporary Materialism: Its Ontology and Epistemology*. Synthese Library, vol 447. Springer, Cham. https://doi.org/10.1007/978-3-030-89488-7_1

Peterson, E. (2022). The English language in Finland: Tool of modernity or tool of coloniality? In J. Hoegaerts, T. Liimatainen, L. Hekanaho, & E. Peterson (Eds.), *Finnishness, whiteness and coloniality* (pp. 216–234). Helsinki University Press. https://doi.org/10.33134/HUP-17-11

Peterson, E., & Beers K. (Eds.). (2023). *English in the Nordic countries : Connections, tensions, and everyday realities*. Taylor & Francis Group. https://doi.org/10.4324/9781003272687

Pladevall-Ballester, E. (2014). Exploring primary school CLIL perceptions in Catalonia: students', teachers' and parents' opinions and expectations. *International Journal of Bilingual Education and Bilingualism*, *18*(1), 45–59. https://doi.org/10.1080/13670050.2013.874972

Pliiddemann, P. (2013). "Additive" and "subtractive": Challenges in education for multilingualism. *Per Linguam : A Journal of Language Learning*, *13*(1). https://doi.org/10.5785/13-1-197

Plonsky, L. (2017). Quantitative research methods. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 505–521). Routledge. https://doi.org/10.4324/9781315676968-28

Ráez Padilla, J. (2018). Parent Perspectives on CLIL Implementation: Which Variables Make a Difference? *Porta Linguarum Revista Interuniversitaria de Didáctica de Las Lenguas Extranjeras, 29*, 181–196. https://doi.org/10.30827/Digibug.54033

Rascón Moreno, D., & Bretones Callejas, C. (2018). Socioeconomic Status and its Impact on Language and Content Attainment in CLIL Contexts. *Porta Linguarum Revista Interuniversitaria de Didáctica de Las Lenguas Extranjeras, 29*, 115–135. https://doi.org/10.30827/Digibug.54025

Riazi, A. M., & Candlin, C. N. (2014). Mixed-methods research in language teaching and learning: Opportunities, issues and challenges. *Language Teaching*, *47*(2), 135–173. https://doi.org/10.1017/S0261444813000505

Rogers, J., & Li, P. (2025). The time between tasks in task repetition research: A systematic review. *Language Teaching Research : LTR*. https://doi.org/10.1177/13621688251323047

Roiha, A. S. (2012). Teachers' views on differentiation in content and language integrated learning (CLIL): Perceptions, practices and challenges. *Language and Education*, *28*(1), 1–18. https://doi.org/10.1080/09500782.2012.748061

Rojczyk, A. (2011). Chapter 13. Time-Limited Verbal Fluency Task with Polish−English Unbalanced Bilinguals. In J. Arabski & A. Wojtaszek (Ed.), *Individual Learner Differences in SLA* (pp. 210-225). Bristol, Blue Ridge Summit: Multilingual Matters. https://doi.org/10.21832/9781847694355-015

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's Object Pictorial Set: The Role of Surface Detail in Basic-Level Object Recognition. *Perception (London), 33*(2), 217-236. https://doi.org/10.1068/p5117

Rumlich, D. (2016). *Evaluating bilingual education in Germany: CLIL students' general English proficiency, EFL self-concept and interest*. Peter Lang Edition. https://doi.org/10.3726/978-3-653-06460-5

Rusfandi, R. (2024). Reverse Transfer of Writing Rhetorical Structures by Indonesian EFL Learners. *Celt*, *24*(2), 275–296. https://doi.org/10.24167/celt.v24i2.12263

San Isidro, X., & Lasagabaster, D. (2018). The impact of CLIL on pluriliteracy development and content learning in a rural multilingual setting: A longitudinal study. *Language Teaching Research, 23*(5), 584–602. https://doi.org/10.1177/1362168817754103

San Isidro, X., & Lasagabaster, D. (2020). Students' and families' attitudes and motivations to language learning and CLIL: a longitudinal study. *Language Learning Journal*, *50*(1), 119–134. https://doi.org/10.1080/09571736.2020.1724185

Sántha, K., & Malomsoki-Sántha, Á. (2023). Methodological Aspects for Applied Linguistic Research – Mixed Methods, Triangulation. *Acta Universitatis Sapientiae, Philologica*, *15*(2), 1–13. https://doi.org/10.2478/ausp-2023-0013

Sarani, A. (2018). The Effect of Bilingualism/ Monolinguals on L2 Working Memory Capacity and Verbal Intelligence. *Iranian Journal of Applied Language Studies*, *10*(1), 205–231. https://doi.org/10.22111/ijals.2018.4271

Sato, T., & Hemmi, C. (2022). Development of second language productive skills through CLIL in a Japanese university: a pre-experimental longitudinal study. *Language Learning in Higher Education (Berlin, Germany)*, *12*(1), 309–326. https://doi.org/10.1515/cercles-2022-2040

Schurz, A., & Sundqvist, P. (2022). Connecting Extramural English with ELT: Teacher Reports from Austria, Finland, France, and Sweden. *Applied Linguistics* 43(5). 934–957. https://doi.org/10.1093/applin/amac013

Seikkula-Leino, J. (2002). *Miten oppilaat oppivat vieraskielisessä opetuksessa? oppilaiden suoriutumistasot, itsetunto ja motivaatio vieraskielisessä opetuksessa*. Turun yliopisto

Seikkula-Leino, J. (2007). CLIL Learning: Achievement Levels and Affective Factors. *Language and Education, 21*(4), 328–341. https://doi.org/10.2167/le635.0

Serra, C. (2007). Assessing CLIL at Primary School: A Longitudinal Study. *International Journal of Bilingual Education and Bilingualism*, *10*(5), 582–602. https://doi.org/10.2167/beb461.0

Shao, Z., Janse, E., Visser, K., & Meyer, A. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, *5*, 772. https://doi.org/10.3389/fpsyg.2014.00772

Siow, S., Gillen, N. A., Lepădatu, I., & Plunkett, K. (2023). Double it up: Vocabulary size comparisons between UK bilingual and monolingual toddlers. *Infancy*, *28*(6), 1030–1051. https://doi.org/10.1111/infa.12562

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of experimental psychology: Human learning and memory, 6*(2), 174. https://doi.org/10.1037/0278-7393.6.2.174

Soldat-Jaffe, T. (2014). The European Charter for regional or minority languages: a magnum opus or an incomplete modus vivendi? *Journal of Multilingual and Multicultural Development*, *36*(4), 372–385. https://doi.org/10.1080/01434632.2014.926910

Suarez, C.A. (2023). The Role of Local Languages in Public Spaces: Investigating the Functions of Translanguaging in Homilies. In: Canilao, M.L.E.N., De Los Reyes, R.A. (Eds.) *Translanguaging for Empowerment and Equity*. Springer, Singapore. https://doi-org.libproxy.tuni.fi/10.1007/978-981-99-8589-0_9

Sundqvist, P. (2024). Extramural English as an individual difference variable in L2 research: Methodology matters. *Annual Review of Applied Linguistics*, *44*, 79–91. https://doi.org/10.1017/S0267190524000072

Swain, M. (1985). "Communicative Competence: Some Roles of Comprehensible Input and Comprehensible Output in Its Development." In S. Gass & C. Madden (Eds.), *Input in Second Language Acquisition* (pp. 235-253). Rowley, MA: Newbury House

Tabuenca-Cuevas, M. (2016). Language Policies and Language Certificates in Spain-What's the Real Cost? *Educational Studies*, *52*(5), 438–451. https://doi.org/10.1080/00131946.2016.1214914

Tervola, M., Pajunen, A., Vainio, S., Honko, M., & Mattila, K. (2015). Maahanmuuttajataustaisten lääkärien suomen kielen taito laillistamiskuulustelussa. *Duodecim, 131*(4), 339–346.

Tsuchiya, K. (2019). Translanguaging Performances in a CLIL Classroom at a Japanese University. In: Tsuchiya, K., Pérez Murillo, M.D. (eds) *Content and Language Integrated Learning in Spanish and Japanese Contexts*. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-030-27443-6_11

Turner, M. (2013). Content-based Japanese Language Teaching in Australian Schools: Is CLIL a Good Fit? *Japanese Studies*, *33*(3), 315–330. https://doi.org/10.1080/10371397.2013.846211

United Nations. (n.d.). *Official languages*. Retrieved October 8, 2025, from https://www.un.org/en/our-work/official-languages

VanPatten, B., & Oikkenon, S. (1996). Explanation versus Structured Input in Processing Instruction. *Studies in Second Language Acquisition*, *18*(4), 495–510. https://doi.org/10.1017/S0272263100015394

Vega-Mendoza, M., West, H., Sorace, A., & Bak, T. H. (2015). The impact of late, non-balanced bilingualism on cognitive performance. *Cognition*, *137*(Apr), 40–46. https://doi.org/10.1016/j.cognition.2014.12.008

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Trans. & Eds.). Cambridge, MA: Harvard University Press.

Wauters, L., & Marquardt, T. P. (2018). Category, Letter, and Emotional Verbal Fluency in Spanish–English Bilingual Speakers: A Preliminary Report. *Archives of Clinical Neuropsychology, 33*(4), 444-457. https://doi.org/10.1093/arclin/acx063

Wewer, T. (2020). Assessment of young learners' English proficiency in bilingual content instruction (CLIL) in Finland: Practices, challenges, and points for development. In M. DeBoer & D. Leontjev (Eds.), *Assessment and learning in content and language integrated learning (CLIL) classrooms* (pp. 137–166). Springer. https://doi.org/10.1007/978-3-030-54128-6_6

Wilcox, R.R. (2012). *Introduction to Robust Estimation and Hypothesis Testing* (3rd ed.). St. Louis: Elsevier Science & Technology.

Wong, C.-Y. (Cathy), & Tian, Z. (2025). Maximizing students' content and language development: The pedagogical potential of translanguaging in a Chinese immersion setting. *Learning and Instruction*, *95*, 102023-. https://doi.org/10.1016/j.learninstruc.2024.102023

Yagmur, K. (2019). The concept of minority/minorities in the European national and supranational EU discourse. *Multilingua*, *38*(2), 213–229. https://doi.org/10.1515/multi-2018-0063

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

Painosalama, Turku, Finland 2025

ISBN 978-952-02-0376-4 (PRINT)
ISBN 978-952-02-0377-1 (PDF)
ISSN 0082-6987 (Print)
ISSN 2343-3191 (Online)